

(51) International Patent Classification:
G06F 12/02 (2006.01)(21) International Application Number:
PCT/US2017/036680(22) International Filing Date:
09 June 2017 (09.06.2017)

(25) Filing Language: English

(26) Publication Language: English

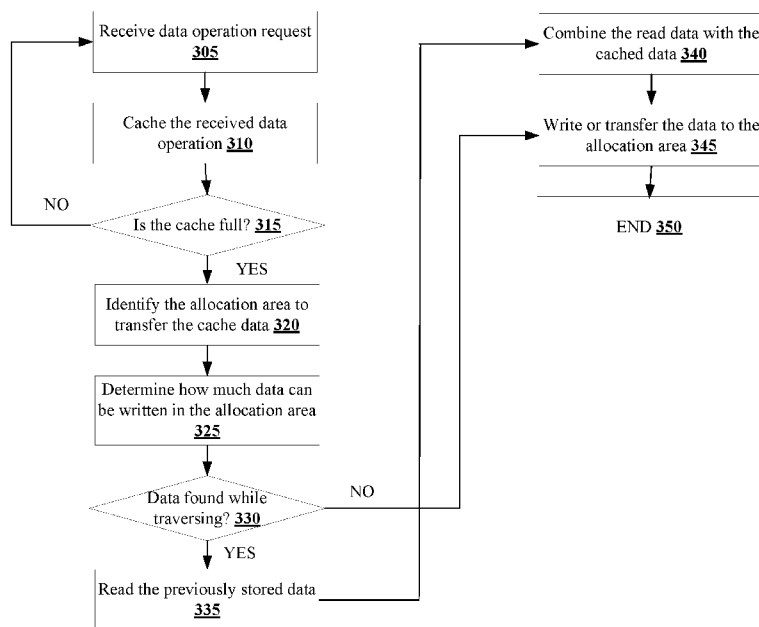
(30) Priority Data:
15/195,093 28 June 2016 (28.06.2016) US(71) Applicant: NETAPP INC. [US/US]; 495 East Java Drive,
Sunnyvale, CA 94089 (US).(72) Inventors: DRONAMRAJU, Ravikanth; 7740 Creekside
Drive, Pleasanton, CA 94588 (US). GUPTA, Shivali; Sy-
camore B24, tat Sherwood Apartment, Basavanagar, Ban-
galore, Karnataka 560037 (IN). STERLING, Kyle; 1550Eddy Street, Apt. 506, San Francisco, CA 94115 (US).
GOEL, Atul; 1574 Taipei Drive, San Jose, CA 95131 (US).(74) Agent: GALLO, Nicholas J. et al.; LeClairRyan, A Profes-
sional Corporation, 70 Linden Oaks, Suite 210, Rochester,
NY 14625 (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, DJ, DK, DM, DO,
DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,
HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,
KR, KW, 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,(54) Title: METHODS FOR MINIMIZING FRAGMENTATION IN SSD WITHIN A STORAGE SYSTEM AND DEVICES THERE-
OF

FIG. 3

(57) Abstract: A method, non-transitory computer readable medium, and device that assists with reducing memory fragmentation in solid state devices includes identifying an allocation area within an address range to write data from a cache. Next, the identified allocation area is determined for including previously stored data. The previously stored data is read from the identified allocation area when it is determined that the identified allocation area comprises previously stored data. Next, both the write data from the cache and the read previously stored data are written back into the identified allocation area sequentially through the address range.



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).

Published:

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

-1-

METHODS FOR MINIMIZING FRAGMENTATION IN SSD WITHIN A STORAGE SYSTEM AND DEVICES THEREOF

FIELD

[0001] This application claims priority benefit of U.S. Utility Patent Application Serial
5 No. 15/195,093, filed June 28, 2016 and is hereby incorporated by reference in its entirety.

[0002] This technology generally relates to data storage management and, more particularly, methods for minimizing fragmentation in solid state devices (SSD) within a storage system and devices thereof.

BACKGROUND

10 **[0003]** Repeated writing data sets to, and erasing data sets from, a data storage medium causes storage space to be used inefficiently and this phenomenon is called fragmentation within the data storage medium. Fragmentation within the data storage mediums phenomenon reduces storage capacity, and increases access and read times. Fragmentation occurs when a computing device does not allocate enough contiguous storage space to encode a complete data set as a
15 single sequence, and instead divides the data set into portions which are encoded in gaps between other data sets. Typically, these data storage medium gaps arise because an earlier data set was deleted, or excess space was allocated to another stored data set. The degree to which a data storage medium is fragmented heavily dependent upon the size and number of data sets stored on it.

20 **[0004]** To address the fragmentation issue, prior technologies uses the technique of defragmentation where the datasets written to a data storage medium is rearranged such that segments of each data set are written contiguously. Unfortunately, the technique used in the prior technologies fails to significantly reduce fragmentation and are time consuming, memory intensive operations that can significantly reduce the performance of a computer system while
25 being executed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] FIG. 1 is a block diagram of an environment with a storage management computing device that minimizes fragmentation in SSD within a storage system;

-2-

[0006] FIG. 2 is a block diagram of the exemplary storage management computing device shown in FIG. 1; and

[0007] FIG. 3 is an exemplary flow chart of an example of a method for minimizing fragmentation in SSD within a storage system.

5

DETAILED DESCRIPTION

[0008] An environment 10 with a plurality of client computing devices 12(1)-12(n), an exemplary storage management computing device 14, a plurality of storage devices 16(1)-16(n) is illustrated in FIG. 1. In this particular example, the environment 10 in FIG. 1 includes the plurality of client computing devices 12(1)-12(n), the storage management computing device 14 and a plurality of storage devices 16(1)-16(n) coupled via one or more communication networks 30, although the environment could include other types and numbers of systems, devices, components, and/or other elements. The example of a method for minimizing fragmentation in SSD within a storage system is executed by the storage management computing device 14, although the approaches illustrated and described herein could be executed by other types and numbers of other computing systems and devices. The environment 10 may include other types and numbers of other network elements and devices, as is generally known in the art and will not be illustrated or described herein. This technology provides a number of advantages including providing methods, non-transitory computer readable media and devices for minimizing fragmentation in SSD within a storage system.

20 [0009] Referring to FIG. 2, in this example the storage management computing device 14 includes a processor 18, a memory 20, and a communication interface 24 which are coupled together by a bus 26, although the storage management computing device 14 may include other types and numbers of elements in other configurations.

[0010] The processor 18 of the storage management computing device 14 may execute one or more programmed instructions stored in the memory 20 for dynamic resource reservation based on classified input/output requests as illustrated and described in the examples herein, although other types and numbers of functions and/or other operation can be performed. The processor 18 of the storage management computing device 14 may include one or more central processing units (“CPUs”) or general purpose processors with one or more processing cores, such as AMD® processor(s), although other types of processor(s) could be used (e.g., Intel®).

30

[0011] The memory 20 of the storage management computing device 14 stores the programmed instructions and other data for one or more aspects of the present technology as described and illustrated herein, although some or all of the programmed instructions could be stored and executed elsewhere. A variety of different types of memory storage devices, such as a non-volatile memory, random access memory (RAM) or a read only memory (ROM) in the system or a floppy disk, hard disk, CD ROM, DVD ROM, or other computer readable medium which is read from and written to by a magnetic, optical, or other reading and writing system that is coupled to the processor 18, can be used for the memory 20.

[0012] The communication interface 24 of the storage management computing device 14 operatively couples and communicates with the plurality of client computing devices 12(1)-12(n) and the plurality of storage devices 16(1)-16(n), which are all coupled together by the communication network 30, although other types and numbers of communication networks or systems with other types and numbers of connections and configurations to other devices and elements. By way of example only, the communication network 30 can use TCP/IP over Ethernet and industry-standard protocols, including NFS, CIFS, SOAP, XML, LDAP, and SNMP, although other types and numbers of communication networks, can be used. The communication networks 30 in this example may employ any suitable interface mechanisms and network communication technologies, including, for example, any local area network, any wide area network (e.g., Internet), teletraffic in any suitable form (e.g., voice, modem, and the like), Public Switched Telephone Network (PSTNs), Ethernet-based Packet Data Networks (PDNs), and any combinations thereof and the like. In this example, the bus 26 is a universal serial bus, although other bus types and links may be used, such as PCI-Express or hyper-transport bus.

[0013] Each of the plurality of client computing devices 12(1)-12(n) includes a central processing unit (CPU) or processor, a memory, and an I/O system, which are coupled together by a bus or other link, although other numbers and types of network devices could be used. The plurality of client computing devices 12(1)-12(n) communicates with the storage management computing device 14 for storage management, although the client computing devices 12(1)-12(n) can interact with the storage management computing device 14 for other purposes. By way of example, the plurality of client computing devices 12(1)-12(n) may run application(s) that may provide an interface to make requests to access, modify, delete, edit, read or write data within storage management computing device 14 or the plurality of storage devices 16(1)-16(n) via the communication network 30.

-4-

[0014] Each of the plurality of storage devices 16(1)-16(n) includes a central processing unit (CPU) or processor, and an I/O system, which are coupled together by a bus or other link, although other numbers and types of network devices could be used. Each plurality of storage devices 16(1)-16(n) assists with storing data, although the plurality of storage devices 16(1)-16(n) can assist with other types of operations such as storing of files or data. Various network processing applications, such as CIFS applications, NFS applications, HTTP Web Data storage device applications, and/or FTP applications, may be operating on the plurality of storage devices 16(1)-16(n) and transmitting data (e.g., files or web pages) in response to requests from the storage management computing device 14 and the plurality of client computing devices 12(1)-12(n). It is to be understood that the plurality of storage devices 16(1)-16(n) may be hardware such as hard disk drives, solid state devices (SSD), or magnetic tapes, or software or may represent a system with multiple external resource servers, which may include internal or external networks. While in this example it is illustrated that the plurality of storage devices 16(1)-16(n) are outside the storage management computing device 14, the plurality of storage devices 16(1)-16(n) can reside within the storage management computing device 14 in other examples.

[0015] Although the exemplary network environment 10 includes the plurality of client computing devices 12(1)-12(n), the storage management computing device 14, and the plurality of storage devices 16(1)-16(n) described and illustrated herein, other types and numbers of systems, devices, components, and/or other elements in other topologies can be used. It is to be understood that the systems of the examples described herein are for exemplary purposes, as many variations of the specific hardware and software used to implement the examples are possible, as will be appreciated by those of ordinary skill in the art.

[0016] In addition, two or more computing systems or devices can be substituted for any one of the systems or devices in any example. Accordingly, principles and advantages of distributed processing, such as redundancy and replication also can be implemented, as desired, to increase the robustness and performance of the devices and systems of the examples. The examples may also be implemented on computer system(s) that extend across any suitable network using any suitable interface mechanisms and traffic technologies, including by way of example only teletraffic in any suitable form (e.g., voice and modem), wireless traffic media, wireless traffic networks, cellular traffic networks, G3 traffic networks, Public Switched

-5-

Telephone Network (PSTNs), Packet Data Networks (PDNs), the Internet, intranets, and combinations thereof.

[0017] The examples also may be embodied as a non-transitory computer readable medium having instructions stored thereon for one or more aspects of the present technology as described and illustrated by way of the examples herein, as described herein, which when executed by the processor, cause the processor to carry out the steps necessary to implement the methods of this technology as described and illustrated with the examples herein.

[0018] An example of a method for minimizing fragmentation in SSD within a storage system will now be described herein with reference to FIGS.1-3. The exemplary method begins at step 305 where the storage management computing device 14 receives a data operation request, such as a read request or a write request by way of example, from one of the plurality of client computing devices 12(1)-12(n), although the storage management computing device 14 can receive other types and/or numbers of requests from other devices. In this particular example, the received data operation request includes a write request on an object, although the storage management computing device 14 can receive other types and/or numbers of data operation requests from other devices.

[0019] Next in step 310, the storage management computing device 14 caches the received write data operation into the cache within memory 20 of the storage management computing device 14, although the storage management computing device 14 can cache or otherwise store the operations at other memory locations.

[0020] Next in step 315, upon caching the received write data operation the storage management computing device 14 may determine when the cache within memory 20 has reached a threshold size. In this particular example, the cache within the memory 20 has a stored or otherwise set threshold size after which the cache cannot receive further data. Accordingly, when the storage management computing device 14 determines that the cache has not yet reached the threshold size, then the No branch is taken back to step 305 where the storage management computing device 14 continues to receive the data operation request from the plurality of client computing devices 12(1)-12(n).

[0021] However, if back in step 315 when the storage management computing device 14 determines that the cache is full, then the Yes branch is taken to step 320. In step 320, the

-6-

storage management computing device 14 identifies an allocation area in one of the plurality of storage devices 16(1)-16(n) which is a solid state drive (SSD), although the storage management computing device 14 can identify the allocation area at other memory locations, such as in memory 20. In this example, allocation area relates to the memory space in the SSD where the contents from the cache could be transferred. Additionally in this example, the storage management computing device 14 includes metadata stored in the memory 20 that includes the address range in the memory strip within the SSD that can stage the data being flushed out of the cache within the memory 20. The storage management computing device 14 may use the metadata to identify the allocation area, although the storage management computing device 14 can use other types and/or amounts of information to identify the allocation area. Further in this example, the size allocation area identified by the storage management computing device 14 is at least equal to the size of the data in the cache that is going to be transferred, although the size of the allocation area can be more than the size of the data being transferred. Furthermore, the address range of the identified allocation area can include a combination of free space and previously stored data (due to fragmented data that was previously stored) in this example.

[0022] Next in step 325, the storage management computing device 14 determines or identifies the amount of data in the cache that can be transferred to the identified allocation area sequentially from the starting of the address range of the identified allocation area to the end of the address range of the identified allocation area.

[0023] In step 330, during the traversing the storage management computing device 14 determines when there is data stored in the address range of the identified allocation area, although the storage management computing device 14 can use other techniques to make the determination. In this example, the allocation area identified by the storage management computing device 14 can include data that was previously transferred from the cache within memory 20, although the identified allocation area can include other types and/or amounts of data. Accordingly, when the storage management computing device 14 determines that the data is present in the address range, then the Yes branch is taken in to step 335.

[0024] In step 335, the storage management computing device 14 reads the previously stored data in the address range of the allocation area. In this example, the storage management computing device 14 is required to read the previously stored data within the address range in the allocation area as the data has to be first read before writing the data back into the SSD.

[0025] Next in step 340, the storage management computing device 14 combines the read data from step 335 with the cache data that is required to be transferred to the allocation area, although the storage management computing device 14 can add other types and/or amounts of information.

5 **[0026]** Next in step 345, the storage management computing device 14 writes or transfers the data present in the cache within the memory 20 and the previously stored data that was read back into the identified allocation area within the SSD. Additionally back in step 330, when the storage management computing device 14 determines that there was no data found during the traversing, then the No branch is taken to this step 345. When the exemplary flow proceeds to
10 step 345 with no data in the address range of the identified allocation area, then the storage management computing device 14 only writes or transfers the data from the cache within memory 20 to the identified allocation area. In this example, the storage management computing device 14 begins the write or transfers the data from the start of the address range of the identified allocation and proceeds to write or transfer the data sequentially in the address range.
15 By writing the data into the allocation area sequentially, the technology is able to reduce the fragmentation in the SSD within the plurality of storage devices 16(1)-16(n) and the exemplary method ends at step 350.

[0027] Accordingly, as illustrated and described by way of the examples herein, this technology provides a number of advantages including providing methods, non-transitory
20 computer readable media and devices for minimizing fragmentation in SSD within an enterprise storage system. By writing the data sequentially into the allocation area in the SSD, the technology able to significantly reduce the fragmentation in SSD thereby increases the memory utilization of the SSD.

[0028] Having thus described the basic concept of the technology, it will be rather
25 apparent to those skilled in the art that the foregoing detailed disclosure is intended to be presented by way of example only, and is not limiting. Various alterations, improvements, and modifications will occur and are intended to those skilled in the art, though not expressly stated herein. These alterations, improvements, and modifications are intended to be suggested hereby, and are within the spirit and scope of the technology. Additionally, the recited order of
30 processing elements or sequences, or the use of numbers, letters, or other designations therefore, is not intended to limit the claimed processes to any order except as may be specified in the

-8-

claims. Accordingly, the invention is limited only by the following claims and equivalents thereto.

CLAIMS

What is claimed is:

1. A method for reducing fragmentation in a solid state device (SSD), the method comprising:

5 identifying, by a storage management computing device, an allocation area within an address range to write data from a cache;

determining, by the storage management computing device, when there is previously stored data in the identified allocation area;

10 reading, by the storage management computing device, the previously stored data in the identified allocation area when the determination indicates that the identified allocation area comprises previously stored data; and

writing, by the storage management computing device, both the write data from the cache and the read previously stored data back into the identified allocation area sequentially through the address range.

15

2. The method as set forth in claim 1 further comprising writing, by the storage management computing device, only write data from the cache to the identified allocation area sequentially through the address range when the previously stored data is not determined to be present in the identified allocation area.

20

3. The method as set forth in claim 1 or 2 further comprising receiving, by the storage management computing device, a write data operation from a client computing device.

4. The method as set forth in claim 3 further comprising:

25 caching, by the storage management computing device, the received write data operation in the cache; and

determining, by the storage management computing device, when the cache has reach a threshold memory size.

30 5. The method as set forth in claim 4 further comprising identifying, by the storage management computing device, the allocation area to transfer data from the cache when the cache is determined to have reached the threshold memory size.

-10-

6. A non-transitory computer readable medium having stored thereon instructions for reducing fragmentation in a solid state device (SSD) comprising executable code which when executed by a processor, causes the processor to perform steps comprising:

identifying an allocation area within an address range to write data from a cache;

determining when there is previously stored data in the identified allocation area;

reading the previously stored data in the identified allocation area when the determination indicates that the identified allocation area comprises previously stored data; and

writing both the write data from the cache and the read previously stored data back into the identified allocation area sequentially through the address range.

7. The medium as set forth in claim 6 further comprising writing only write data from the cache to the identified allocation area sequentially through the address range when the previously stored data is not determined to be present in the identified allocation area.

8. The medium as set forth in claim 6 or 7 further comprising receiving a write data operation from a client computing device.

9. The medium as set forth in claim 8 further comprising:

caching the received write data operation in the cache; and

determining when the cache has reached a threshold memory size.

10. The medium as set forth in claim 9 further comprising identifying the allocation area to transfer data from the cache when the cache is determined to have reached the threshold memory size.

11. A storage management computing device comprising:

a processor;

a memory coupled to the processor, wherein the processor is configured to execute program instructions stored in the memory to:

identify an allocation area within an address range to write data from a cache;

determine when there is previously stored data in the identified allocation area;

read the previously stored data in the identified allocation area when the

determination indicates that the identified allocation area comprises previously stored data; and

-11-

write both the write data from the cache and the read previously stored data back into the identified allocation area sequentially through the address range.

12. The device as set forth in claim 11 wherein the processor coupled to the memory is further configured to execute at least one additional program instruction stored in the memory to write only write data from the cache to the identified allocation area sequentially through the address range when the previously stored data is not determined to be present in the identified allocation area.

13. The device as set forth in claim 11 or 12 wherein the processor coupled to the memory is further configured to execute at least one additional program instruction stored in the memory to receive a write data operation from a client computing device.

14. The device as set forth in claim 13 wherein the processor coupled to the memory is further configured to execute at least one additional program instruction stored in the memory to: cache the received write data operation in the cache; and determine when the cache has reached a threshold memory size.

15. The device as set forth in claim 14 wherein the processor coupled to the memory is further configured to execute at least one additional program instruction stored in the memory to identify the allocation area to transfer data from the cache when the cache is determined to have reached the threshold memory size.

16. The device as set forth in any of claims 11 to 15, wherein the allocation area comprises memory space in one or more solid state devices.

17. A computer program comprising program instructions that when executed by one or more processors in a computing system cause the computing system to perform the method of any of claims 1 to 5.

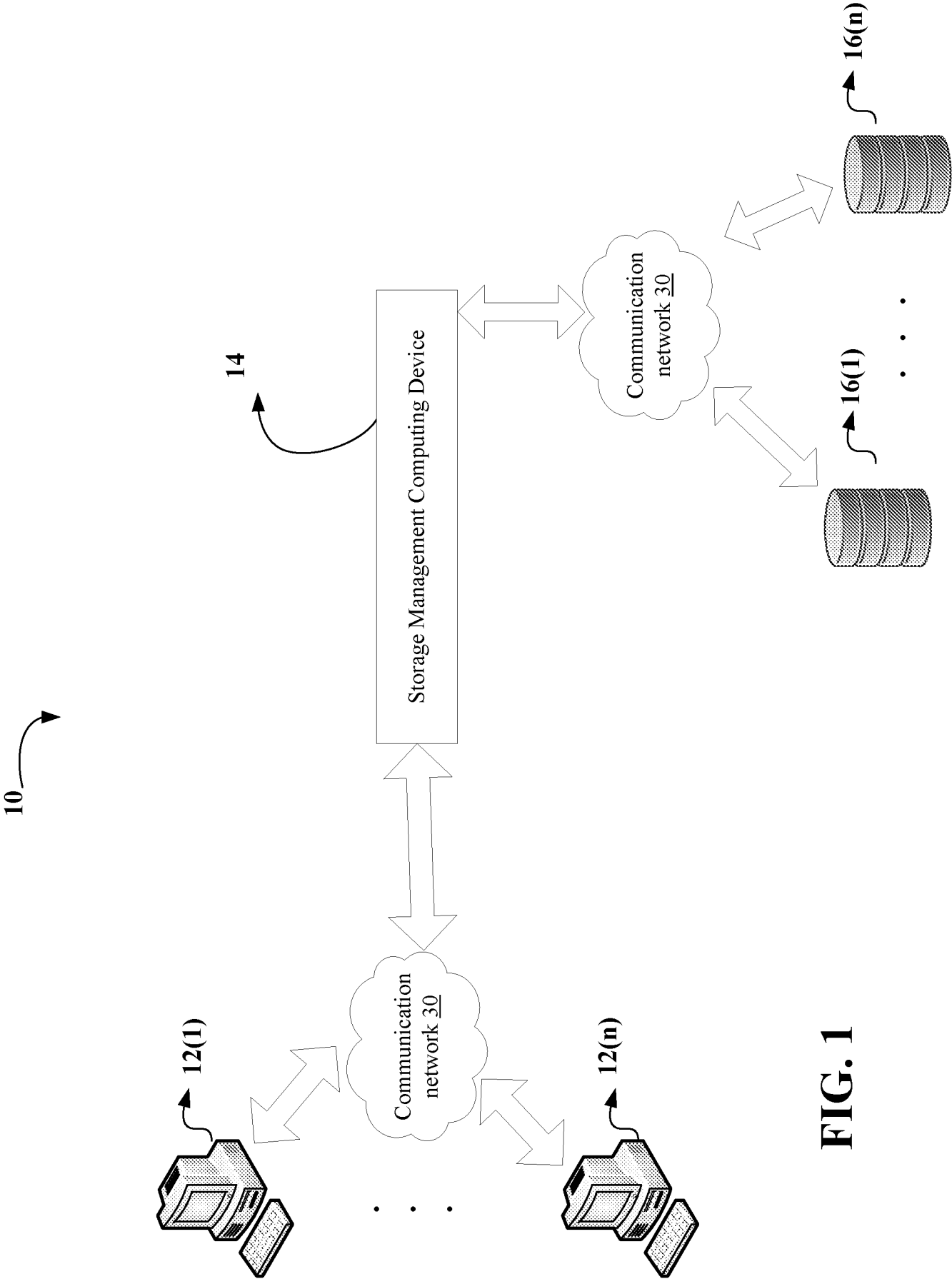


FIG. 1

Storage Management Computing Device 14

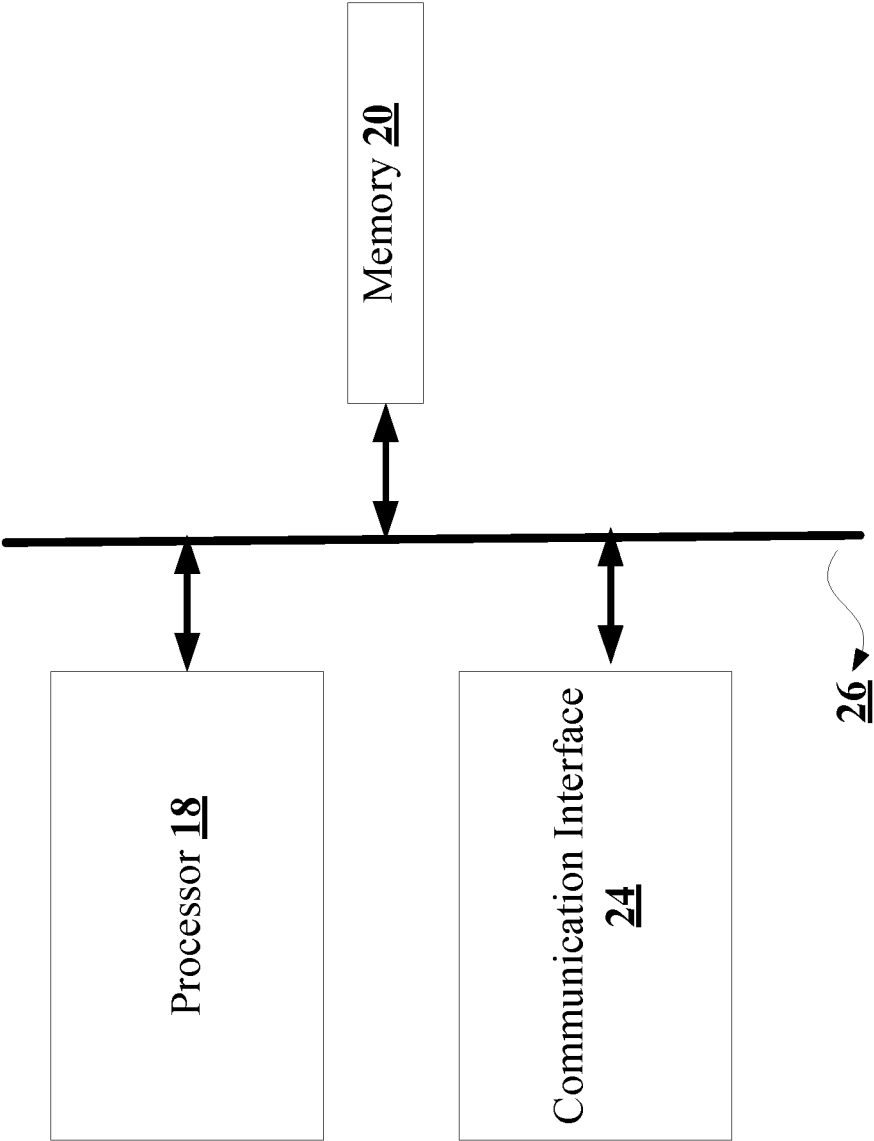


FIG. 2

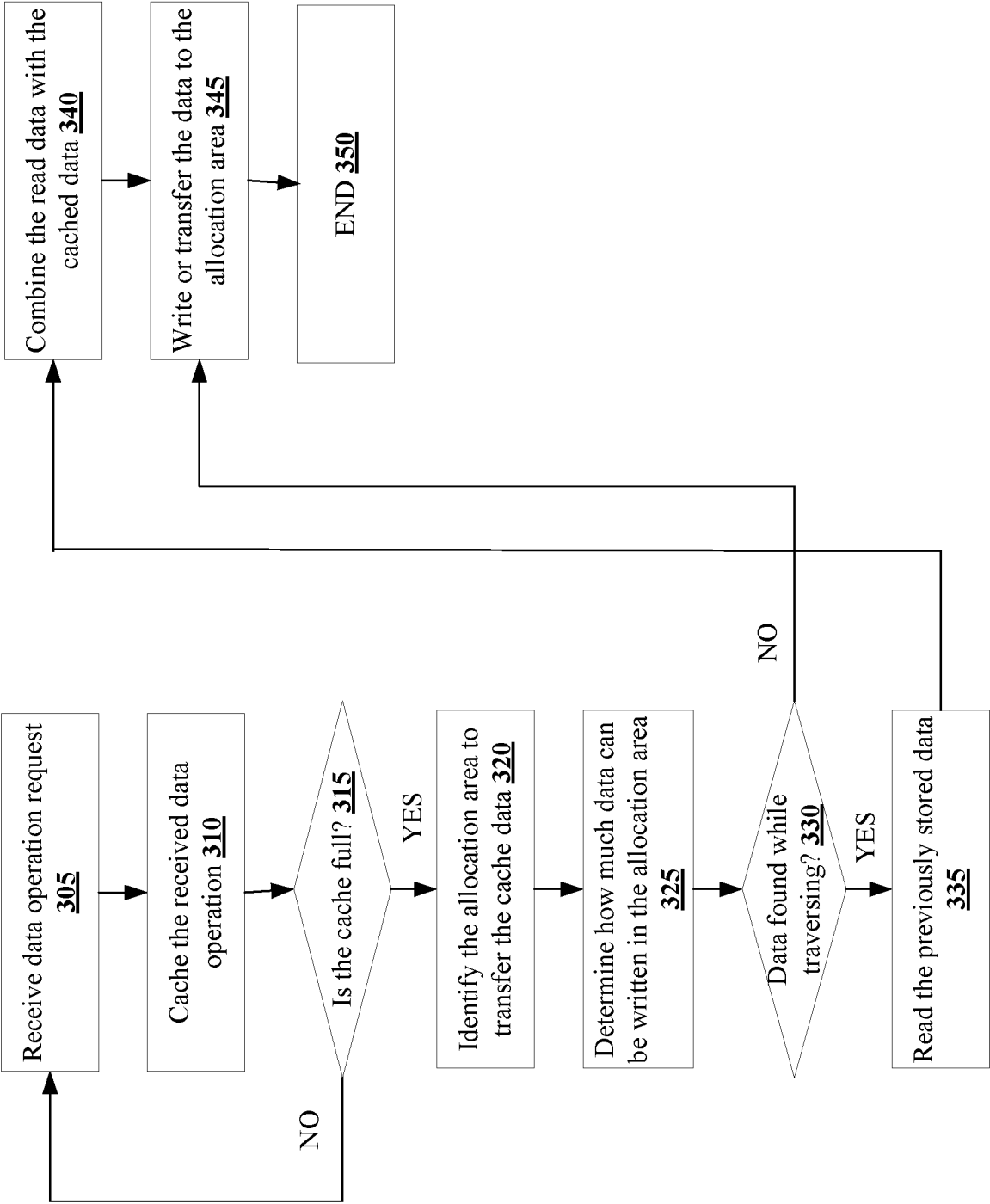


FIG. 3

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/036680

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F12/02
ADD.

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

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2015/106410 A1 (ZALTSMAN ETAI [IL] ET AL) 16 April 2015 (2015-04-16) figures 1-3 paragraph [0038] - paragraph [0060] -----	1-17
A	US 2014/289449 A1 (OGATA KATSUHIKO [JP]) 25 September 2014 (2014-09-25) paragraph [0033] - paragraph [0048] figure 2 -----	1-17
A	US 2011/320733 A1 (SANFORD STEVEN TED [US] ET AL) 29 December 2011 (2011-12-29) paragraph [0033] - paragraph [0048] figure 5 -----	1-17



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

18 August 2017

Date of mailing of the international search report

25/08/2017

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Nguyen Xuan Hiep, C

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2017/036680

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2015106410 A1	16-04-2015	TW 201520793 A US 2015106410 A1 WO 2015053966 A1	01-06-2015 16-04-2015 16-04-2015
US 2014289449 A1	25-09-2014	CN 103827970 A JP 5550741 B1 JP WO2014049678 A1 US 2014289449 A1 WO 2014049678 A1	28-05-2014 16-07-2014 18-08-2016 25-09-2014 03-04-2014
US 2011320733 A1	29-12-2011	EP 2577470 A2 US 2011320733 A1 WO 2011153478 A2	10-04-2013 29-12-2011 08-12-2011