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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 November 2016 (17.11.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/182587 A1

(51) International Patent Classification:
H04L 12/24 (2006.01) *H04L 12/26* (2006.01)

(21) International Application Number:
PCT/US2015/042340

(22) International Filing Date:
28 July 2015 (28.07.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2358/CHE/2015 8 May 2015 (08.05.2015) IN

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

(72) Inventors: **RANJAN, Jyoti**; Sy.No.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN). **ITI, Suprit A.**; Sy.No.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN). **AV, Pradeep Kumar**; Sy.No.192, Whitefield Road, Mahadevapura Post, Bangalore 560048 (IN).

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

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

[Continued on next page]

WO 2016/182587 A1

(54) Title: HETEROGENEOUS HOT SPARE SERVER POOL

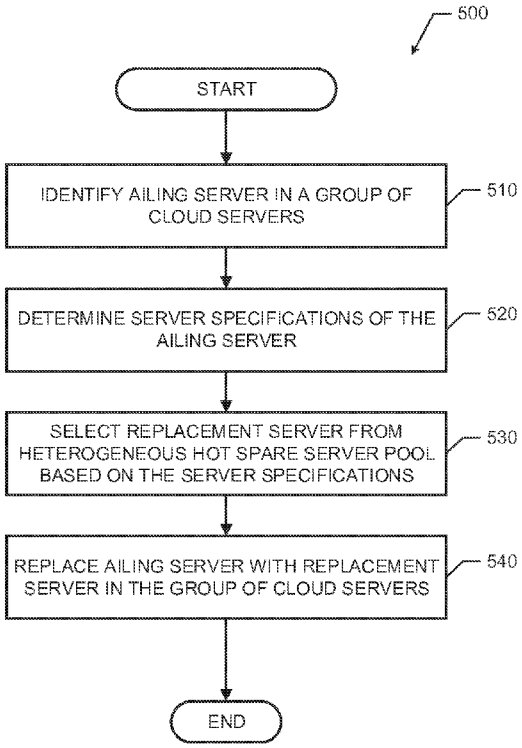


FIG. 5

(57) Abstract: An example involves identifying an ailing server in a group of cloud servers, determining server specifications of the ailing server by analyzing services provided by the ailing server, selecting a replacement server from a heterogeneous hot spare server pool based on the server specifications; and replacing the ailing server with the replacement server in the group of cloud servers.

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

Declarations under Rule 4.17:

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

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

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

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

HETEROGENEOUS HOT SPARE SERVER POOL

BACKGROUND

[0001] Cloud servers provide a diverse set of services such as computing, imaging, storage, identity, etc. A cloud may include of a plurality of servers accessible by a client device to utilize the services of the cloud. The client devices may communicate with the cloud using a variety of communication links (e.g., wired or wireless), communication networks (e.g., the Internet, a local area network (LAN), a wide area network (WAN), etc.), or communication protocols.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] FIG. 1 illustrates a schematic diagram of an example server system including an example hot spare manager implemented in accordance with an aspect of this disclosure.

[0003] FIG. 2 is a block diagram of an example hot spare manager that may implement the hot spare manager of FIG. 1.

[0004] FIG. 3 is an example mapping of servers of a group of cloud servers that may be implemented by the hot spare manager of FIG. 2.

[0005] FIG. 4 is an example implementation of the server system of FIG. 1 that illustrates an example replacement of an ailing server with a replacement server from a heterogeneous hot spare server pool.

[0006] FIG. 5 is a flowchart representative of example machine readable instructions that may be executed to implement the touch detector of FIG. 2.

[0007] FIG. 6 is a block diagram of an example processor platform capable of executing the instructions of FIGS. 5 to implement the hot spare manager of FIG. 2.

[0008] Wherever possible, the same reference numbers will be used throughout the drawing(s) and accompanying written description to refer to the same or like parts.

DETAILED DESCRIPTION

[0009] Examples disclosed herein involve a heterogeneous hot spare server pool including replacement servers. The replacement servers, which may be referred to interchangeably as spare servers or hot spare servers, of the heterogeneous hot spare server pool may replace ailing servers (e.g., servers experiencing errors) in a group of cloud servers. An example hot spare manager identifies the ailing servers and selects an appropriate replacement server from the heterogeneous hot spare server pool based on the type of service provided by the ailing server, the server type of the ailing server, or specifications of the ailing server.

[0010] Examples disclosed herein provide a hot spare manager to analyze a group of cloud servers (e.g., nodes) and replace ailing servers in the group of cloud servers with replacement servers (spare servers) from a heterogeneous hot spare server pool. Examples disclosed herein provide increased service performance by identifying ailing servers and identifying appropriate replacement servers for the ailing servers based on specifications of the ailing servers, services provided by the ailing servers, software executing on the ailing servers, etc. Furthermore the ailing servers may be replaced with the selected replacement servers. In some examples, the hot spare manager deploys appropriate software to the replacement servers for restoring the services of the ailing servers (e.g., by providing the same or similar services). Examples disclosed herein may be automated such that an operator, user, or administrator of a group of servers may not necessarily manually replace ailing servers with replacement servers. In examples disclosed herein, replacement servers may not be dedicated as a backup to a particular server, but may be used as a replacement server for a plurality of servers (e.g., a plurality of servers being a same type as the replacement server).

[0011] An example method includes identifying an ailing server in a group of cloud servers; identifying server specifications of the ailing server by analyzing services provided by the ailing server; selecting a replacement server from a heterogeneous hot spare server pool based on the server specifications;

and replacing the ailing server with the replacement server in the group of cloud servers.

[0012] FIG. 1 is a schematic illustration of an example server system 100 including an example hot spare manager 110 implemented in accordance with the teachings of this disclosure. The example server system 100 of FIG. 1 includes the hot spare manager 110, a group of cloud servers 120 (which may be referred to herein as the cloud servers 120), and a heterogeneous hot spare server pool 130. In the illustrated example of FIG. 1, the hot spare manager 110 is in communication with (e.g., via wired or wireless communication link(s)) the cloud servers 120 and the heterogeneous hot spare server pool 130. In some examples, the hot spare manager 110, group of cloud servers 120, and/or heterogeneous hot spare server pool 130 may be physically located at a same location (e.g., within a same room, same building, etc.) or located at different physical locations.

[0013] The example group of cloud servers 120 of FIG. 1 may include a plurality of servers that provide a plurality of respective services. For example, the cloud servers 120 may be configured as a cloud managed by an entity or a plurality of entities. Accordingly, the cloud servers 120 may be grouped into subgroups to provide respective services or may provide services as a whole. The example cloud servers 120 may be a group of heterogeneous servers of various types, models, etc. that have various specifications, such as various compute speeds, storage, memory capabilities, sizes, etc. Accordingly, one of the cloud servers 120 may be better suited for performing certain services rather than other services. For example, a first cloud server may be better suited for computing services while a second cloud server may be better suited for storage services. In some examples, some or all of the cloud servers 120 may work together to provide a single service (e.g., for a single entity) or some or all of the cloud servers 120 may work separately from one another to provide a plurality of services (e.g., for a plurality of entities).

[0014] In examples disclosed herein, the group of cloud servers 120 of FIG. 1 may include: 1) average compute intensive nodes (e.g., for authentication/authorization, PXE booting services, etc.), 2) high compute

intensive nodes (e.g., for hosting, indexing, etc., etc.), 3) high storage capacity nodes (e.g., for storage services). Other example server types may also be included in the group of cloud servers 120.

[0015] In examples disclosed herein, the cloud servers in the example group of cloud servers 120 may implement a variety of example services, such as an image application programming interface (API) node (e.g., implemented by a high compute intensive node), an image storage node (e.g., implemented by a high storage capacity node), a block API controller node (e.g., implemented by a compute intensive node), a block storage node (e.g., implemented by a high storage capacity node), a compute API controller node (e.g., implemented by a high compute intensive node), a compute hypervisor node (e.g., implemented by a high compute intensive node), an object API controller node (e.g., implemented by a high compute intensive node), an identity service node (e.g., implemented by an average compute intensive node), a user interface node (e.g., implemented by an average compute intensive node), a proxy node (e.g., implemented by an average compute intensive node), etc.

[0016] The heterogeneous hot spare server pool 130 of FIG. 1 includes a plurality of replacement servers (i.e., servers that are not providing any particular service or are not part of a designated cloud (or service) or group of servers that provides services). The example replacement servers of the heterogeneous hot spare server pool 130 of FIG. 1 are on standby for use in the group of cloud servers 120 when the hot spare manager 110 determines that at least one of the cloud servers 120 is/are ailing. As used herein, an ailing server is a server that may be encountering errors, failing, having speed issues (e.g., processing data below a threshold speed), having storage issues (e.g., unable to complete storage of data, unable to securely store data, etc.), having security issues, etc. In some examples, the replacement servers of the heterogeneous hot spare server pool 130 may not necessarily replace an ailing server but can be added to the group of cloud servers 120 to cure deficiencies of an ailing server (e.g., if the ailing server does not have enough power to process an unexpected amount of data).

[0017] In examples disclosed herein, the hot spare manager 110 may identify an ailing server in the group of cloud servers 120 and replace the ailing server with a corresponding server from the heterogeneous hot spare server pool 130, which is referred to herein as a replacement server, to maintain services of the cloud servers 120. In examples disclosed herein, the hot spare manager 110 selects an appropriate replacement server from the heterogeneous hot spare server pool 130 based on specifications of the ailing server (e.g., computing power, storage capacity, memory capabilities, speed, size, etc.), service(s) provided by the ailing server, a software executed by the ailing server, a type (e.g., model, manufacturer, identifier, etc.) of the ailing server, etc. In some examples, the hot spare manager 110 may determine specifications of the ailing server based on services provided by the ailing server or software (or software components) executed by the ailing server. In some examples, the hot spare manager 110 may retrieve and/or upload software corresponding to the ailing server to the replacement server to maintain the service(s) of the cloud servers 120.

[0018] FIG. 2 is a block diagram of an example hot spare manager 110 that may be used to implement the hot spare manager 110 of FIG. 1. The example hot spare manager 110 of FIG. 2 includes a server monitor 210, a server analyzer 220, a server selector 230, and a replacement manager 240. In the illustrated example of FIG. 2, a communication bus 250 facilitates communication between the server monitor 210, the server analyzer 220, the server selector 230, and the replacement manager 240.

[0019] The example server monitor 210 of FIG. 2 monitors the cloud servers 120 of FIG. 1 to identify or detect ailing servers (e.g., servers that are experiencing errors). For example, the server monitor 210 may monitor operating parameters of the cloud servers 120. Such example operating parameters may include processing speed, data storage capabilities, data loss, disk errors, statuses, clogging, etc. of the cloud servers 120 to identify errors, faults, etc. or any other unexpected performance issues. In the event that the server monitor 210 determines that operating parameters of a server in the group of cloud servers 120 are outside of an expected threshold, the server

monitor 210 determines that the server is ailing or is likely to fail. In some examples, the server monitor may monitor the cloud servers 120 and determine that a server is ailing based on a performance parameters of the group of cloud servers 120. In such an example, the server monitor 210 may further investigate the group of cloud server 120 to identify the server(s) responsible for the performance issues of the group. In some examples, the server monitor 210 may receive error notifications (e.g., from server controllers) indicating that a server is experiencing an error. In response to detecting an ailing server in the group of cloud servers 120, the server monitor 210 notifies the server analyzer 220 to analyze the ailing server.

[0020] The server analyzer 220 of FIG. 2 analyzes the ailing server to identify specifications of the ailing server, a type of the ailing server, a type of service provided by the ailing server, or service components of the ailing server. For example, the server analyzer 220 may identify software (e.g., executable instructions that are being executed by the ailing server) deployed on the ailing server. Based on the identified software, the server analyzer 220 may determine specifications (e.g., at least minimum specifications) of the ailing server for executing the software to provide the services of the ailing server. The server analyzer 220 may identify a reference, an identification (e.g., a server ID), installed components, operating requirements, etc. of the software. In some examples, the server analyzer 220 may capture a snapshot of the ailing server in the group of cloud servers to identify software components of the ailing server. The snapshot (e.g., a golden image) may create an image (e.g., a summary) of software components of the ailing server. The example snapshot may be captured in response to identifying there is an ailing server in the group of cloud servers 120. The example server analyzer 220 may use the snapshot image and a server ID to determine the software components for services deployed on the ailing server.

[0021] In examples disclosed herein, based on the software deployed on the ailing server, the server analyzer 220 may identify the type of server (e.g., storage, control, proxy, etc.), software requirements, server specifications (e.g., processing power, storage capacity, etc.), or the type of service provided by the

ailing server. In some examples, the server analyzer 220 may determine specifications of the ailing server or component types of the ailing server (e.g., processor type, storage type, storage capacity, processing/memory speed, etc.). The server analyzer 220 may then forward the corresponding server information (e.g., software information, specifications of the server, components of the server, etc.) of the ailing server to the server selector 230 of FIG. 2.

[0022] In some examples, the server analyzer 220 refers to a mapping of software to the cloud servers 120 to determine service components (e.g., applications, software, etc.) corresponding to a particular server or ailing server. For example, the mapping may be generated by an operator or processor using any suitable techniques. An example mapping of N servers is illustrated in FIG. 3. In FIG. 3, service components 302 (e.g., applications, software, etc.), service specifications 304, and server types are mapped for servers 1, 2, 3, ... N of the cloud servers 120 having N number of servers. Accordingly, the server analyzer 220 may identify the ailing server (e.g., based on the server number 1, 2, 3, ..., N (i.e., a server ID)) in the mapping 300 and determine the corresponding service component, corresponding service usage specification, or corresponding server type of the ailing server. As used herein, a server type may be based on specifications (e.g., speed, storage, etc.) or a model, manufacturer, etc. The server analyzer 220 may then provide the information 302, 304, 306 to the server selector 230 to select a replacement server from the heterogeneous hot spare server pool 130 to replace the ailing server in the group of cloud servers 120.

[0023] The example server selector 230 receives the information corresponding to the ailing server to select a replacement server from the heterogeneous hot spare server pool 130. The example server selector 230 analyzes the heterogeneous hot spare server pool 130 to identify and select a server that matches or is best suited to replace the ailing server in the group of cloud servers 120 relative to the other servers of the heterogeneous hot spare server pool 130. For example, the server selector 230 may refer to a mapping of the heterogeneous hot spare server pool 130 (e.g., a mapping similar to the mapping 300 of FIG. 3 that indicates server type, server specifications, etc.). In

some examples, the server selector 230 may compare software parameters (e.g., requirements for effectively executing the software) with specifications (e.g., processing speed, storage size, memory capabilities, accessibility, security, etc.) of the servers in the heterogeneous hot spare server pool 130 to determine an appropriate server for replacing the ailing server in the group of cloud servers 120 to restore the services of the ailing server.

[0024] The replacement manager 240 of the example hot spare manager 110 of FIG. 2 prepares the selected replacement server from the heterogeneous hot spare server pool 130 to replace the ailing server in the group of cloud servers 120. For example, the replacement manager 240 may assign an appropriate address, identifier, etc. to the selected replacement server of the heterogeneous hot spare server pool 130 such that the selected replacement server is virtually transferred from the heterogeneous hot spare server pool 130 to the group of cloud servers 120. In some examples, the replacement manager 240 may retrieve software (e.g., software corresponding to the software of the ailing server) to be uploaded to the replacement server in the group of cloud servers 120. For example, the software may be retrieved from a database of the group of cloud servers 120 or a database in communication with the hot spare manager 110 that manages software for the services of the cloud servers 120.

[0025] In some examples, the replacement manager 240 of FIG. 2 may remove the ailing server from the group of cloud servers 120 by decommissioning the ailing server or placing the ailing server in a standby state or repair state such that the ailing server is no longer actively providing any services. In such examples, a system administrator may be notified that the ailing server is to be repaired or that the ailing server has been removed from the group of cloud servers 120. In some examples, the replacement manager 240 may reassign the ailing server to the group of cloud servers 120 after the ailing server has been repaired to remove any detected errors or issues. Additionally or alternatively, the replacement manager 240 may remove the replacement server from the heterogeneous hot spare server pool 130. In such examples, the replacement manager 240 may notify a system administrator that

the replacement sever has been removed from the heterogeneous hot spare server pool 130.

[0026] Accordingly, the hot spare manager 110 of FIG. 2 monitors the health of the cloud servers 120 to identify errors in the group of cloud servers 120 or to identify errors in the cloud servers 120, themselves. Further, in response to detecting the errors, the hot spare manager 110 replaces the ailing server with a replacement server from the heterogeneous hot spare server pool 130 that corresponds to the ailing server (e.g., based on software requirements, service type, server type, etc.) in accordance with the teachings of this disclosure.

[0027] While an example manner of implementing the hot spare manager 110 of FIG. 1 is illustrated in FIG. 2, at least one of the elements, processes and/or devices illustrated in FIG. 2 may be combined, divided, re-arranged, omitted, eliminated and/or implemented in any other way. Further, the server monitor 210, the server analyzer 220, the server selector 230, the replacement manager 240, and/or, more generally, the example hot spare manager 110 of FIG. 2 may be implemented by hardware and/or any combination of hardware and executable instructions (e.g., software and/or firmware). Thus, for example, any of the server monitor 210, the server analyzer 220, the server selector 230, the replacement manager 240, and/or, more generally, the example hot spare manager 110 could be implemented by at least one of an analog or digital circuit, a logic circuit, a programmable processor, an application specific integrated circuit (ASIC), a programmable logic device (PLD) and/or a field programmable logic device (FPLD). When reading any of the apparatus or system claims of this patent to cover a purely software and/or firmware implementation, at least one of the server monitor 210, the server analyzer 220, the server selector 230, and the replacement manager 240 is/are hereby expressly defined to include a tangible computer readable storage device or storage disk such as a memory, a digital versatile disk (DVD), a compact disk (CD), a Blu-ray disk, etc. storing the executable instructions. Further still, the example hot spare manager 110 of FIG. 2 may include at least one element, process, and/or device in addition to, or instead of, those illustrated in FIG. 2,

and/or may include more than one of any or all of the illustrated elements, processes and devices.

[0028] FIG. 4 schematically illustrates a detailed example of the server system 100 of FIG. 1. Accordingly, the server system 100 of FIG. 4 includes the hot spare manager 110, the heterogeneous hot spare server pool 130, and cloud servers 421-426. In FIG. 4, the cloud servers 421-426 may implement the cloud servers 120 of FIG. 1 and the heterogeneous hot spare server pool 130 includes replacement servers 431, 432, 433, 434, 435. In the illustrated example of FIG. 4, the cloud servers 421, 422, 423 are labeled as type A servers, the server 424 is labeled as a type B server, and the servers 425, 426 are labelled as type C servers. For sake of this example, the type A cloud servers 421-423 are high compute servers, the type B server is an average compute server, and the type C servers 425, 426 are high storage servers. The example heterogeneous hot spare server pool 130 includes type A replacement servers 431, 432, type B replacement servers 433, 434, and type C replacement servers 435, 436.

[0029] In the illustrated example of FIG. 4, the example hot spare manager 110, which may be implemented by the hot spare manager 110 of FIG. 2, may replace an ailing server 423 with a replacement server 432 in accordance with the teachings of this disclosure. For example, in FIG. 4, the server monitor 210 of the hot spare manager 110 may detect that the server 423 is experiencing errors. In response to detecting the ailing server 423, the server analyzer 220 may analyze the ailing server 423 by identifying a service type or server type of the ailing server 423. For example, the server analyzer 220 may refer to a mapping (e.g., similar to the mapping of FIG. 3) of the cloud servers 120 to identify a service type or server type of the ailing server 423. In some examples, the server analyzer 220 may capture a snapshot (e.g., a golden image) of the service provided by the ailing server 423 that summarizes/identifies the service components of the ailing server 423.

[0030] The server selector 230 of the hot spare manager 110 of FIG. 4 selects a server from the replacement servers 431-436 of the heterogeneous hot spare server pool 130. In examples disclosed herein, the server selector

230 analyzes the replacement servers 431-436 of the heterogeneous hot spare server pool 130 to select a replacement server for the ailing server 423. In the illustrated example of FIG. 3, the server selector 230 may select the server 432 based on a similar type (e.g., both the ailing server 423 and the replacement server 432 are type A servers (high compute servers)). In some examples, the server selector 230 may refer to a mapping (e.g., similar to the mapping 300 of FIG. 3) of the replacement servers to corresponding services or server types of the servers 421-426. Accordingly, the server selector 230 may select an appropriate server from at least one of the servers 431-436 to replace any of the servers 421-426 if they encounter errors, fail, etc.

[0031] The replacement manager 240 of the hot spare manager 110 of FIG. 4 replaces the ailing server 423 with the replacement server 432. In examples disclosed herein, the hot spare manager 110 may decommission the ailing server 423 or place the ailing server 423 in a standby state until the ailing server 423 is repaired. The hot spare manager 110 may then activate the replacement server 432 (e.g., by assigning an address to the replacement server corresponding to an address of the ailing server 423) and deploy software (e.g., an application, executable instructions, etc.) to the replacement server 432. The deployed software may correspond to the service of the ailing server 423. For example, the replacement manager 240 may retrieve the deployed software from a database of the hot spare manager 110 or a data base of the server system 100 storing software that is executable to provide services via the servers 421-426 or via the replacement servers 431-436 of the heterogeneous hot spare server pool 130. The retrieved software may then be uploaded or installed on the replacement server 432 to restore the service of the ailing server 423. Accordingly, the hot spare manager 110 inserting the replacement server 432 into group of cloud servers 120 may allow the services of the ailing server 423 (or a cloud of the ailing server) to continue.

[0032] A flowchart representative of example machine readable instructions for implementing the hot spare manager 110 of FIG. 2 is shown in FIG. 5. In this example, the machine readable instructions comprise a program/process for execution by a processor such as the processor 612 shown

in the example processor platform 600 discussed below in connection with FIG. 6. The program/process may be embodied in executable instructions (e.g., software) stored on a tangible computer readable storage medium such as a CD-ROM, a floppy disk, a hard drive, a digital versatile disk (DVD), a Blu-ray disk, or a memory associated with the processor 612, but the entire program/process and/or parts thereof could alternatively be executed by a device other than the processor 612 and/or embodied in firmware or dedicated hardware. Further, although the example program is described with reference to the flowchart illustrated in FIG. 5, many other methods of implementing the example A1 may alternatively be used. For example, the order of execution of the blocks may be changed, and/or some of the blocks described may be changed, eliminated, or combined.

[0033] The process 500 of FIG. 5 begins with an initiation of the hot spare manager 110 (e.g., upon startup, upon instructions from a user, upon startup of cloud servers 120, upon startup of a device implementing the hot spare manager 110 (e.g., a server or controller of the group of cloud servers 120), etc.). The example process 500 of FIG. 5 may be executed to replace an ailing server in the group of cloud servers 120 with a replacement server from the heterogeneous hot spare server pool 130 of FIG. 1. At block 510, the server monitor identifies an ailing server in the group of cloud servers 120. In some examples at block 510, the server monitor 210 may monitor the cloud servers 120 for errors and in response to detecting the errors identify the ailing server causing the error. Additionally or alternatively, the server monitor 210 may monitor individual servers periodically or aperiodically (e.g., after identifying an error in at least one of the cloud servers 120, after start up or restart of the cloud servers 120, etc.) to identify an ailing server in the group of cloud servers 120.

[0034] At block 520 of FIG. 5, the server analyzer 220 determines server specifications of the ailing server. For example, the server analyzer 220 may refer to a mapping of the cloud servers 120 that indicates service components, software requirements, service components, server types, etc. In such examples, the server analyzer 220 may identify the ailing server in the mapping (e.g., based on a server ID) and determine the server specifications provided

based on the listed service components or service usage specifications. At block 530, the server selector 230 selects a replacement server from the heterogeneous hot spare server pool based on the determined service. For example, the server selector 230 may identify available spare or replacement servers in the heterogeneous hot spare server pool having a same or similar type or features (e.g., speed capabilities, processing capabilities, storage capabilities, size, etc.) of the ailing server.

[0035] At block 540, the server manager 540 replaces the ailing server with the replacement server in the group of cloud servers 120. For example, the server manager 540 may deactivate the ailing server or place the ailing server in a standby state and reassign (e.g., by assigning a new address) the replacement server to the group of cloud servers 120. In some examples, the replacement manager 240 may remove the replacement server from the heterogeneous hot spare server pool 130 such that the selected replacement server cannot be assigned or reassigned to any other service or cloud or be used to replace any other server in the group of cloud servers 120. After block 540, the example process 500 ends.

[0036] As mentioned above, the example process of FIG. 5 may be implemented using coded instructions (e.g., computer and/or machine readable instructions) stored on a tangible computer readable storage medium such as a hard disk drive, a flash memory, a read-only memory (ROM), a compact disk (CD), a digital versatile disk (DVD), a cache, a random-access memory (RAM) and/or any other storage device or storage disk in which information is stored for any duration (e.g., for extended time periods, permanently, for brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term tangible computer readable storage medium is expressly defined to include any type of computer readable storage device and/or storage disk and to exclude propagating signals and to exclude transmission media. As used herein, "tangible computer readable storage medium" and "tangible machine readable storage medium" are used interchangeably. Additionally or alternatively, the example processes of FIG. 5 may be implemented using coded instructions (e.g., computer and/or machine readable instructions) stored

on a non-transitory computer and/or machine readable medium such as a hard disk drive, a flash memory, a read-only memory, a compact disk, a digital versatile disk, a cache, a random-access memory and/or any other storage device or storage disk in which information is stored for any duration (e.g., for extended time periods, permanently, for brief instances, for temporarily buffering, and/or for caching of the information). As used herein, the term non-transitory computer readable medium is expressly defined to include any type of computer readable storage device and/or storage disk and to exclude propagating signals and to exclude transmission media. As used herein, when the phrase "at least" is used as the transition term in a preamble of a claim, it is open-ended in the same manner as the term "comprising" is open ended. As used herein the term “a” or “an” may mean “at least one,” and therefore, “a” or “an” do not necessarily limit a particular element to a single element when used to describe the element. As used herein, when the term “or” is used in a series, it is not, unless otherwise indicated, considered an “exclusive or.”

[0037] FIG. 6 is a block diagram of an example processor platform 600 capable of executing the instructions of FIG. 5 to implement the hot spare manager 110 of FIG. 2. The example processor platform 600 may be or may be any apparatus or may be included in any type of apparatus, such as a server, a personal computer, a mobile device (e.g., a cell phone, a smart phone, a tablet, etc.), or any other type of computing device.

[0038] The processor platform 600 of the illustrated example of FIG. 6 includes a processor 612. The processor 612 of the illustrated example is hardware. For example, the processor 612 can be implemented by at least one integrated circuit, logic circuit, microprocessor or controller from any desired family or manufacturer.

[0039] The processor 612 of the illustrated example includes a local memory 613 (e.g., a cache). The processor 612 of the illustrated example is in communication with a main memory including a volatile memory 614 and a non-volatile memory 616 via a bus 618. The volatile memory 614 may be implemented by Synchronous Dynamic Random Access Memory (SDRAM), Dynamic Random Access Memory (DRAM), RAMBUS Dynamic Random

Access Memory (RDRAM) and/or any other type of random access memory device. The non-volatile memory 616 may be implemented by flash memory and/or any other desired type of memory device. Access to the main memory 614, 616 is controlled by a memory controller.

[0040] The processor platform 600 of the illustrated example also includes an interface circuit 620. The interface circuit 620 may be implemented by any type of interface standard, such as an Ethernet interface, a universal serial bus (USB), and/or a peripheral component interconnect (PCI) express interface.

[0041] In the illustrated example, at least one input device 622 is connected to the interface circuit 620. The input device(s) 622 permit(s) a user to enter data and commands into the processor 612. The input device(s) can be implemented by, for example, an audio sensor, a microphone, a camera (still or video), a keyboard, a button, a mouse, a touchscreen, a track-pad, a trackball, isopoint and/or a voice recognition system.

[0042] At least one output device 624 is also connected to the interface circuit 620 of the illustrated example. The output device(s) 624 can be implemented, for example, by display devices (e.g., a light emitting diode (LED), an organic light emitting diode (OLED), a liquid crystal display, a cathode ray tube display (CRT), a touchscreen, a tactile output device, a light emitting diode (LED), a printer and/or speakers). The interface circuit 620 of the illustrated example, thus, may include a graphics driver card, a graphics driver chip or a graphics driver processor.

[0043] The interface circuit 620 of the illustrated example also includes a communication device such as a transmitter, a receiver, a transceiver, a modem and/or network interface card to facilitate exchange of data with external machines (e.g., computing devices of any kind) via a network 626 (e.g., an Ethernet connection, a digital subscriber line (DSL), a telephone line, coaxial cable, a cellular telephone system, etc.).

[0044] The processor platform 600 of the illustrated example also includes at least one mass storage device 628 for storing executable instructions (e.g., software) and/or data. Examples of such mass storage

device(s) 628 include floppy disk drives, hard drive disks, compact disk drives, Blu-ray disk drives, RAID systems, and digital versatile disk (DVD) drives.

[0045] The coded instructions 632 of FIGS. 5 may be stored in the mass storage device 628, in the local memory 613 in the volatile memory 614, in the non-volatile memory 616, and/or on a removable tangible computer readable storage medium such as a CD or DVD.

[0046] From the foregoing, it will be appreciated that the above disclosed methods, apparatus and articles of manufacture provide a heterogeneous hot spare server pool including replacement servers to automatically replace ailing servers in a group of cloud servers. In examples disclosed herein, the cloud servers are monitored for errors to detect the ailing servers. The ailing servers are replaced with replacement servers from the heterogeneous hot spare server pool by deploying software corresponding to services executed by the ailing servers to the replacement servers. Accordingly, the replacement servers may relatively seamlessly restore services of the ailing server and/or group of cloud servers.

[0047] Although certain example methods, apparatus and articles of manufacture have been disclosed herein, the scope of coverage of this patent is not limited thereto. On the contrary, this patent covers all methods, apparatus and articles of manufacture fairly falling within the scope of the claims of this patent.

CLAIMS

What Is Claimed Is:

1. A method comprising:

identifying an ailing server in a group of cloud servers;

determining server specifications of the ailing server by analyzing services provided by the ailing server;

selecting a replacement server from a heterogeneous hot spare server pool based on the server specifications; and

replacing the ailing server with the replacement server in the group of cloud servers.

2. The method as defined in claim 1, further comprising:

identifying software for the replacement server based on the services provided by the ailing server; and

deploying the software to the replacement server for execution to implement a service provided by the ailing server.

3. The method as defined in claim 1, further comprising:

monitoring the group of cloud servers to identify the ailing server; and

identifying the ailing server in response to detecting that operating parameters of the ailing server are outside of an expected threshold.

4. The method as defined in claim 1, further comprising:

referring to a mapping of the group of cloud servers to identify the service type of the ailing server.

5. The method as defined in claim 1, further comprising capturing a snapshot of the ailing server in response to detecting the ailing server, the snapshot comprising an image of components of the service provided by the ailing server.

6. The method as defined in claim 1, further comprising decommissioning the ailing server by placing the ailing server in a standby state.

7. An apparatus comprising:

a server monitor to monitor a group of cloud servers and identify an ailing server in the group of cloud servers;

a server analyzer to identify server specifications of the ailing server based on a service provided by the ailing server;

a server selector to select a replacement server from a heterogeneous hot spare server pool based on the server specification; and

a replacement manager to replace the ailing server with the replacement server such that the replacement server restores the service provided by the ailing server.

8. The apparatus as defined in claim 7, wherein the server manager is to replace the ailing server with the replacement server by decommissioning the ailing server and deploying software corresponding to the service to the replacement server.

9. The apparatus as defined in claim 7, wherein the server analyzer is to determine the type of service provided by the ailing server by identifying software executed by the ailing server based on a service component in a mapping of the group of cloud servers.

10. The apparatus as defined in claim 7, wherein the server monitor is to detect an error in the group of cloud serves to identify the ailing server.

11. The apparatus as defined in claim 7, wherein the server selector is to identify the service type in a mapping of the heterogeneous hot spare server pool to select the replacement server.

12. The apparatus as defined in claim 7, wherein the plurality of servers in the heterogeneous hot spare server pool includes at least one high compute server, at least one average compute server, and at least one high storage server.

13. A non-transitory machine readable medium comprising instructions that, when executed, cause a machine to at least:

identify an ailing server in a group of cloud servers based on an error detected in the group of cloud servers;

determine a service provided by the ailing server;

select a replacement server from a heterogeneous hot spare server pool based on the service; and

replace the ailing server with the replacement server by deploying software to the replacement server for execution by the replacement server, the software corresponding to the service provided by the ailing server.

14. The non-transitory machine readable medium of claim 13, wherein the instructions, when executed, cause the machine to:

determine the service provided by the ailing server by referring to a mapping of the group of cloud servers including corresponding services provided by servers of the group of cloud servers.

15. The non-transitory machine readable medium of claim 12, wherein instructions when executed further cause the machine to:

deactivate the ailing server prior to replacing the ailing server with the replacement server.

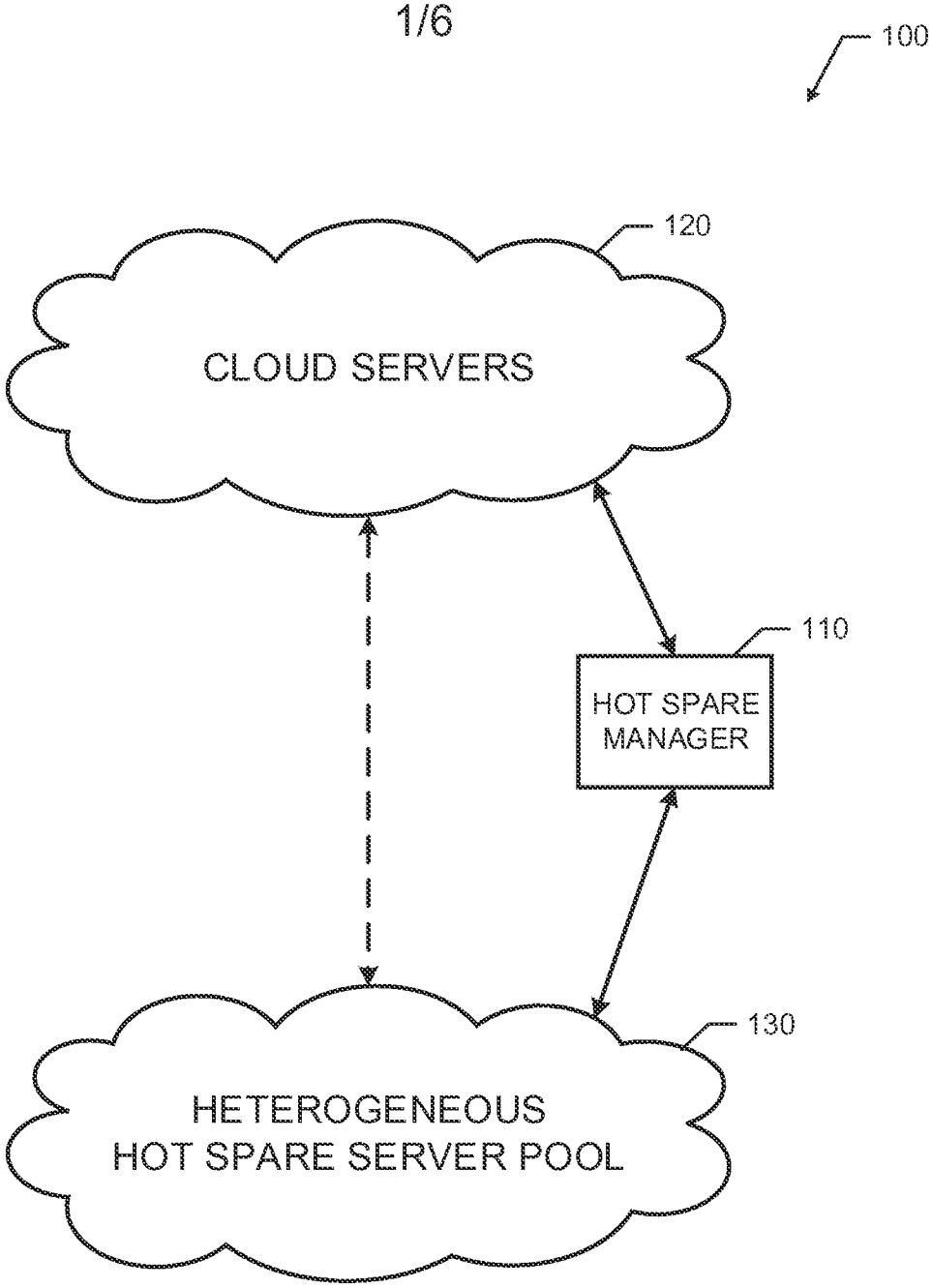

100
120
CLOUD SERVERS
110
HOT SPARE MANAGER
130
HETEROGENEOUS HOT SPARE SERVER POOL

FIG. 1

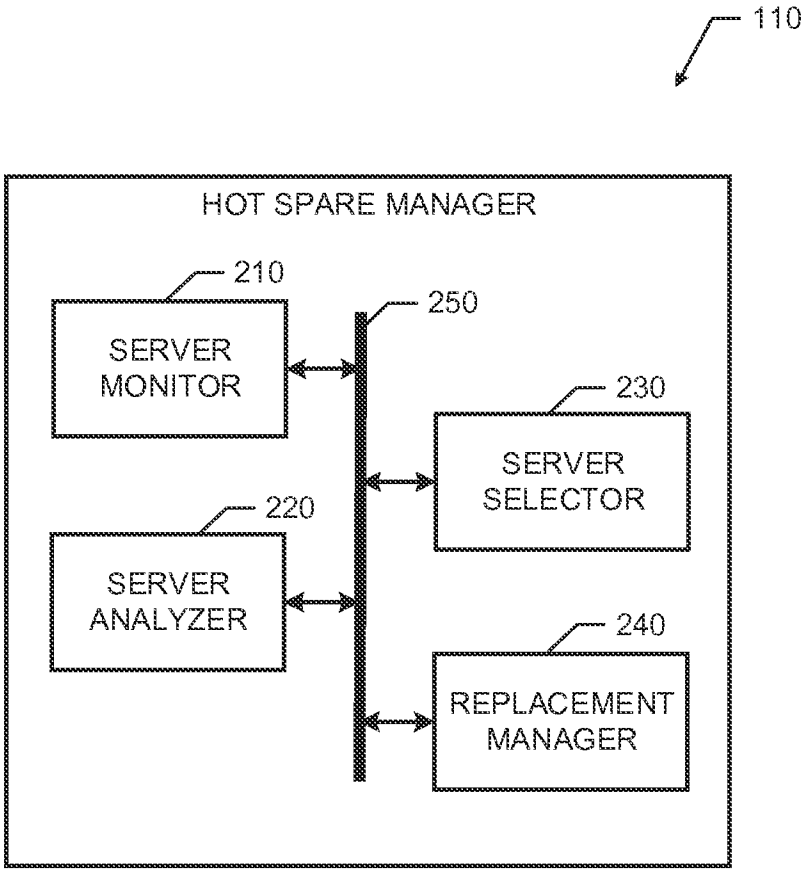

110
HOT SPARE MANAGER
210
SERVER MONITOR
250
230
SERVER SELECTOR
220
SERVER ANALYZER
240
REPLACEMENT MANAGER

FIG. 2

	SERVICE COMPONENT	SERVICE USAGE SPECs	SERVER TYPE
1	COMPUTE API	HIGH COMPUTE	A
2	BLOCK STORAGE	HIGH STORAGE	C
3	WEB SERVICE (UI)	AVERAGE COMPUTE	B
4	COMPUTE HYPERVISOR	HIGH COMPUTE	A
5	OBJECT STORAGE API	HIGH STORAGE	C
6	PROXY	AVERAGE COMPUTE	B
⋮	⋮	⋮	⋮
N	COMPONENT N	SPECS N	SERVER TYPE N

FIG. 3

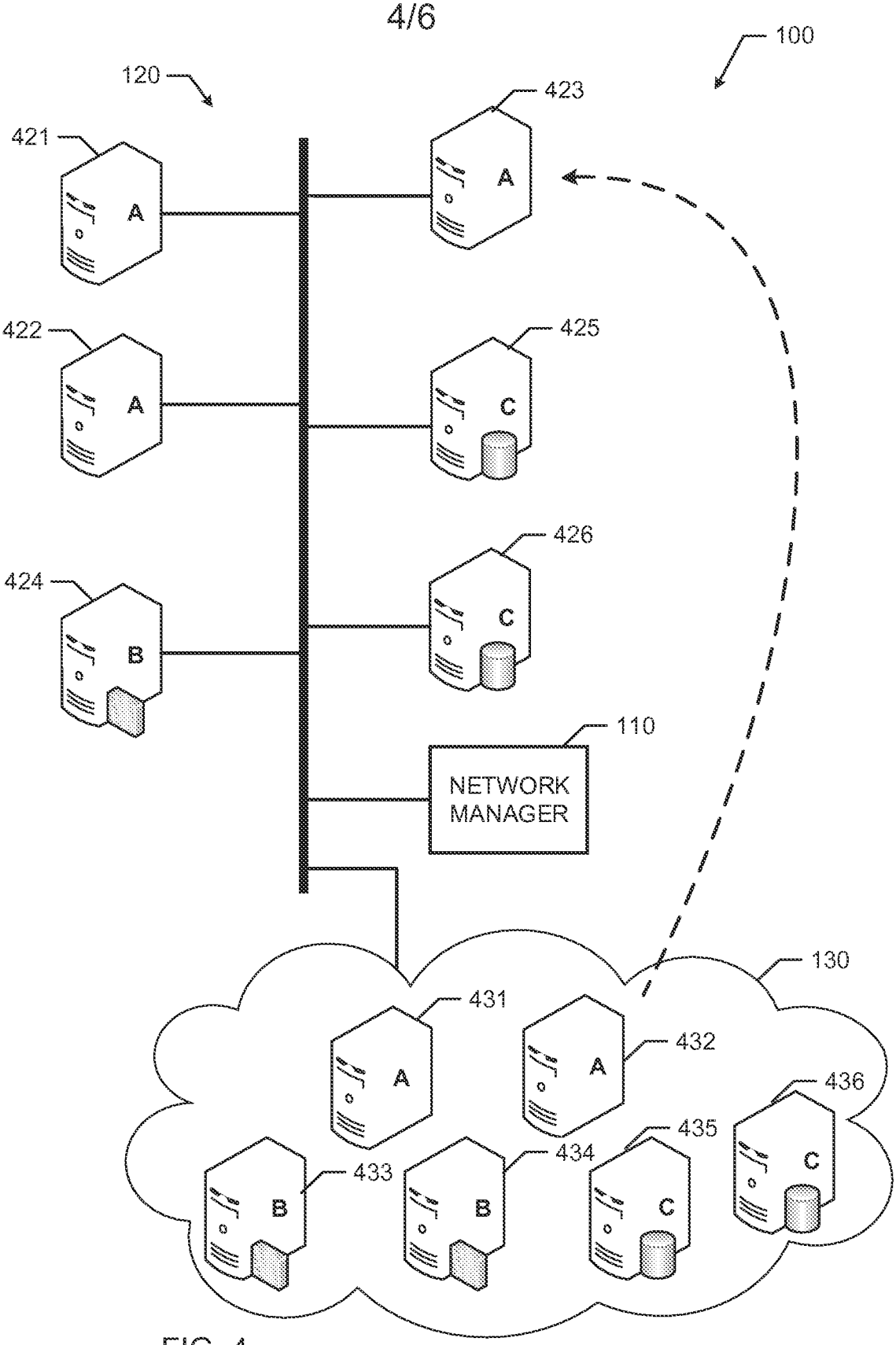

100
120
423
421
A
A
422
425
A
C
426
424
C
B
110
NETWORK
MANAGER
130
431
A
432
A
436
435
434
433
B
B
C
C

FIG. 4

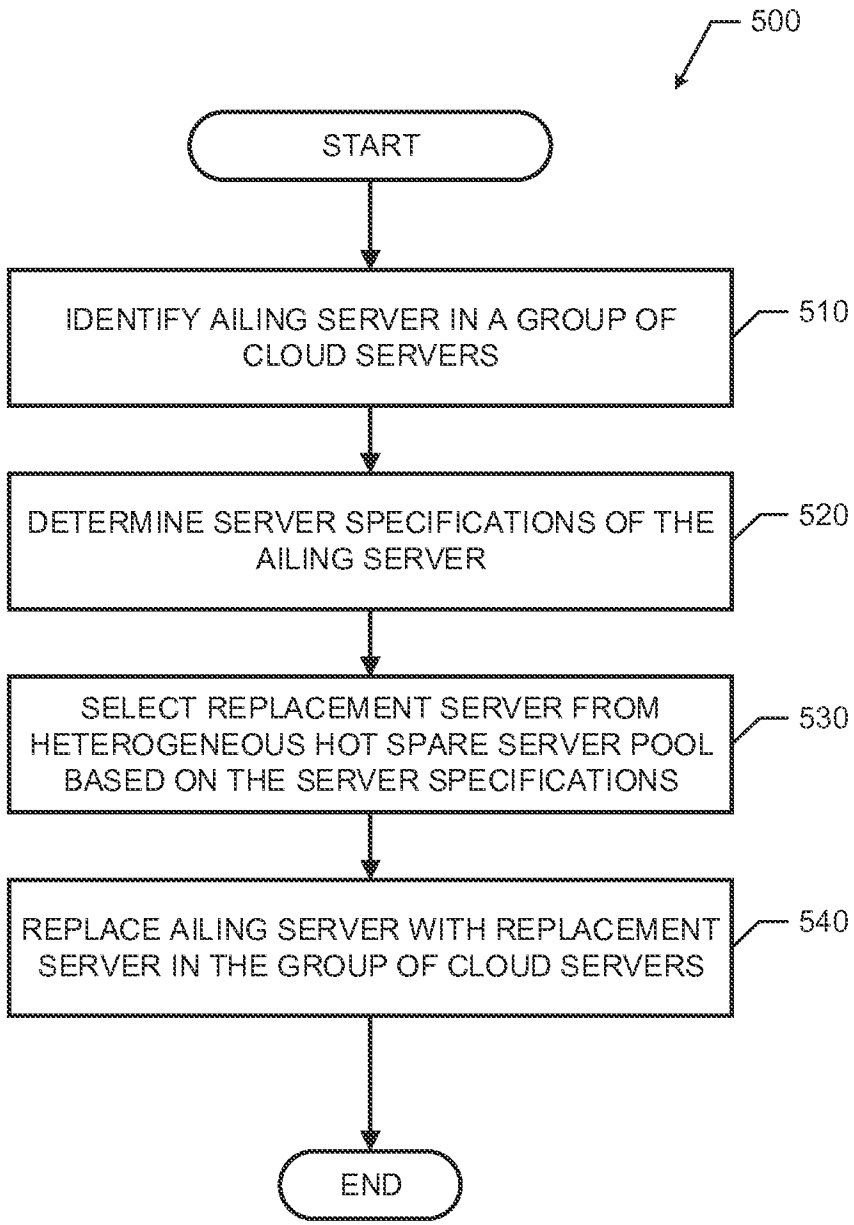

500
START
IDENTIFY AILING SERVER IN A GROUP OF CLOUD SERVERS
510
DETERMINE SERVER SPECIFICATIONS OF THE AILING SERVER
520
SELECT REPLACEMENT SERVER FROM HETEROGENEOUS HOT SPARE SERVER POOL BASED ON THE SERVER SPECIFICATIONS
530
REPLACE AILING SERVER WITH REPLACEMENT SERVER IN THE GROUP OF CLOUD SERVERS
540
END

FIG. 5

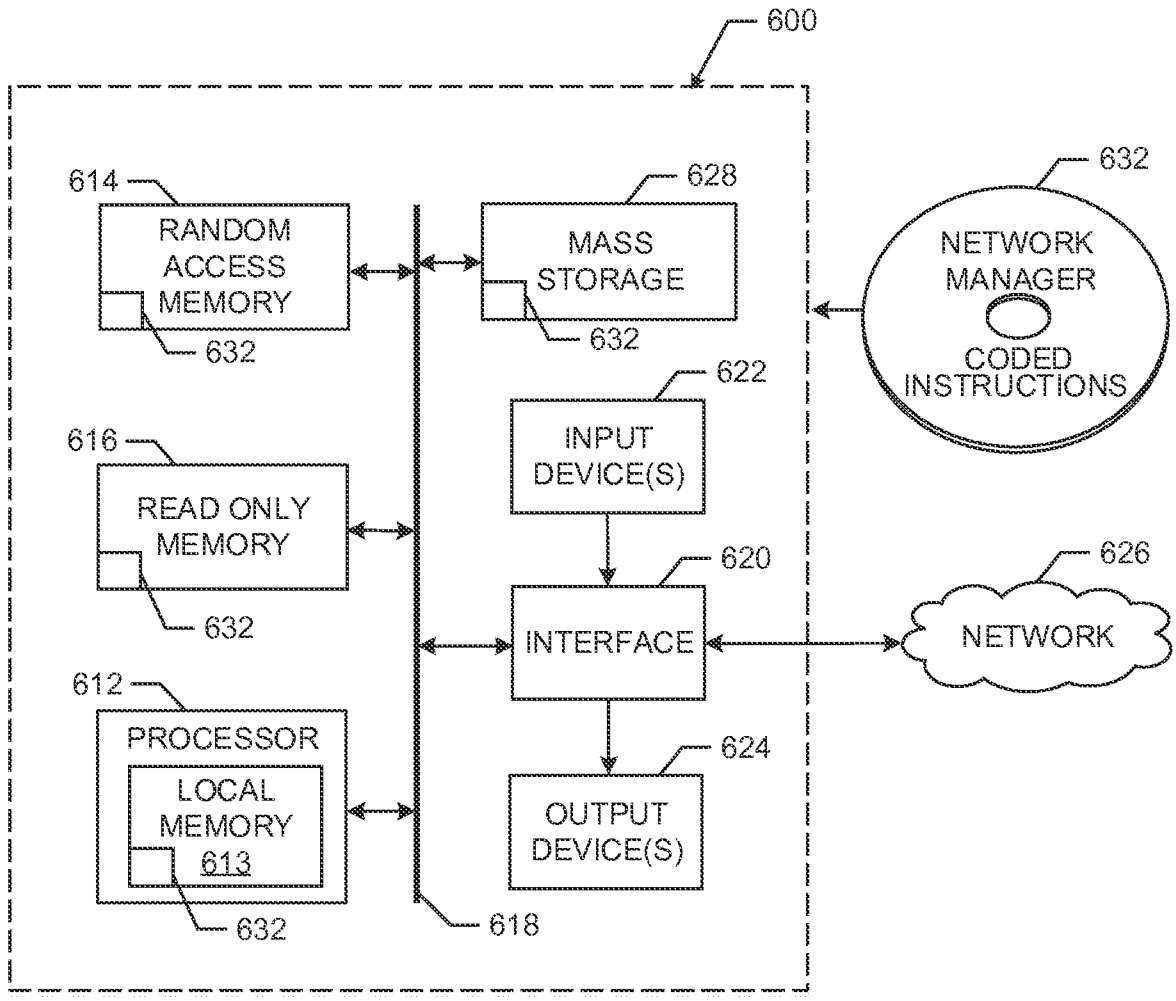

600
614
RANDOM ACCESS MEMORY
632
628
MASS STORAGE
632
632
NETWORK MANAGER
CODED INSTRUCTIONS
622
INPUT DEVICE(S)
616
READ ONLY MEMORY
632
620
INTERFACE
626
NETWORK
612
PROCESSOR
LOCAL MEMORY 613
632
624
OUTPUT DEVICE(S)
618

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/042340

A. CLASSIFICATION OF SUBJECT MATTER

H04L 12/24(2006.01)i, H04L 12/26(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)
H04L 12/24; G06F 15/16; G06F 15/173; G06F 11/14; G06F 11/20; H04L 12/26

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: cloud, failure, server, replacement, hot, spare, 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.
Y	US 2003-0191838 A1 (SHENG (TED) TAI TSAO) 09 October 2003 See paragraph [0062]; and figure 4.	1-15
A	US 2015-0006951 A1 (NAGENDRA K. GURRAM et al.) 01 January 2015 See paragraphs [0021]-[0051]; and figures 1A-5	1-15
A	US 2013-0117448 A1 (NELSON NAHUM et al.) 09 May 2013 See paragraphs [0032]-[0077]; and figures 1, 5-7.	1-15
A	US 05996086 A (WILLIAM P. DELANEY et al.) 30 November 1999 See column 3, line 43 - column 6, line 51; and figures 1-3.	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
23 February 2016 (23.02.2016)	**24 February 2016 (24.02.2016)**

Name and mailing address of the ISA/KR
International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 35208,
Republic of Korea
Facsimile No. +82-42-472-7140

Authorized officer
COMMISSIONER
Telephone No. +82-42-481-5916

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2015/042340

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2011-0258488 A1	20/10/2011	None	
US 2003-0191838 A1	09/10/2003	CN 101416173 A	22/04/2009
		CN 101416173 B	15/12/2010
		CN 102098172 A	15/06/2011
		CN 103593394 A	19/02/2014
		US 07379990 B2	27/05/2008
		US 07418702 B2	26/08/2008
		US 07734778 B2	08/06/2010
		US 07945652 B2	17/05/2011
		US 08473596 B2	25/06/2013
		US 08566463 B2	22/10/2013
		US 08589527 B2	19/11/2013
		US 08639788 B2	28/01/2014
		US 08645542 B2	04/02/2014
		US 08799473 B2	05/08/2014
		US 08812640 B2	19/08/2014
		US 09015321 B2	21/04/2015
		US 2004-0215749 A1	28/10/2004
		US 2004-0215791 A1	28/10/2004
		US 2007-0198713 A1	23/08/2007
		US 2008-0091761 A1	17/04/2008
		US 2008-0115144 A1	15/05/2008
		US 2008-0177821 A1	24/07/2008
		US 2008-0178184 A1	24/07/2008
		US 2008-0184228 A1	31/07/2008
		US 2010-0211879 A1	19/08/2010
		US 2010-0242099 A1	23/09/2010
		US 2011-0066948 A1	17/03/2011
		US 2011-0093740 A1	21/04/2011
		US 2011-0119353 A1	19/05/2011
		US 2011-0119594 A1	19/05/2011
		US 2011-0125889 A1	26/05/2011
		US 2011-0197142 A1	11/08/2011
		US 2011-0213832 A1	01/09/2011
		WO 2007-123785 A2	01/11/2007
		WO 2007-123785 A3	03/04/2008
US 2015-0006951 A1	01/01/2015	US 2015-006950 A1	01/01/2015
US 2013-0117448 A1	09/05/2013	CN 103999031 A	20/08/2014
		EP 2774045 A1	10/09/2014
		JP 05762643 B2	12/08/2015
		JP 2015-501494 A	15/01/2015
		US 08819230 B2	26/08/2014
		US 2014-366121 A1	11/12/2014
		WO 2013-066482 A1	10/05/2013
US 05996086 A	30/11/1999	None	