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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 June 2013 (27.06.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/095505 Al

(51) International Patent Classification:
G06F 1/32 (2006.01)

(21) International Application Number:
PCT/US201 1/066832

(22) International Filing Date:
22 December 201 1 (22. 12.201 1)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US):* **SCHNEIDER ELECTRIC IT CORPORATION** [US/US]; 132 Fairgrounds Road, West Kingston, RI 02892 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only):* **NIELSEN, Torben. JOHANSEN, Flemming** [DK/DK]; Augustvaenget 32, DK-6000 Kolding (DK). **KIIB, Jens, Engsted** [DK/DK]; Thorbjergvej 11, DK-6000 Kolding (DK). **DALGAS, Mikkel** [DK/DK]; AE Skovbryn 8A, DK-6094 Hejls (DK). **VESTER, Kresten, Peter** [DK/DK]; Syrenvej 16, DK-6000 Kolding (DK).

(74) Agent: **SULLIVAN, Thomas, M.;** Lando & Anastasi LLP, Riverfront Office Park, One Main Street, Suite 1100, Cambridge, MA 02142 (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, 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, IS, JP, KE, KG, KM, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, MD, 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, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: SYSTEMS AND METHODS FOR REDUCING ENERGY STORAGE REQUIREMENTS IN A DATA CENTER

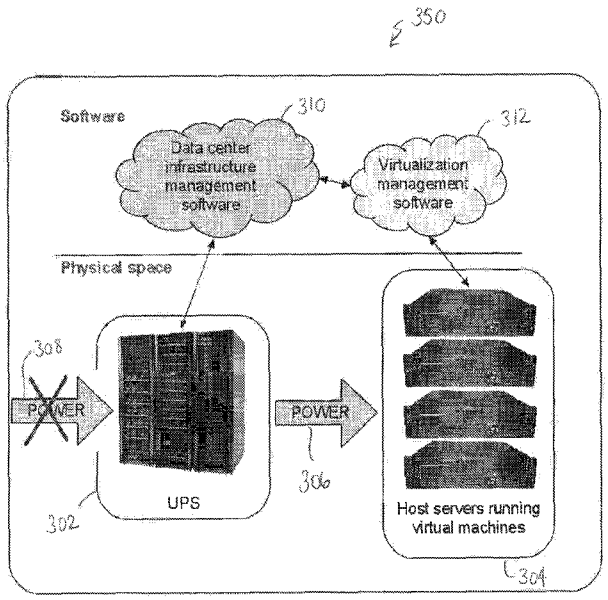


FIG. 4

(57) Abstract: A method for reducing a need for power in a backup mode of operation in a data center includes receiving a first alert from an uninterruptible power supply indicating that the uninterruptible power supply is operating in a first mode, wherein the first mode includes deriving power from a stored energy source, identifying at least one host server configured to receive power from the uninterruptible power supply in the first mode, suspending execution of at least one virtual machine on the at least one host server, receiving a second alert from the uninterruptible power supply indicating that the uninterruptible power supply is operating in a second mode, wherein the second mode includes deriving power from one of a mains power source and a generator, and resuming execution of the at least one virtual machine.

WO 2013/095505 A1

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

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

SYSTEMS AND METHODS FOR REDUCING ENERGY STORAGE REQUIREMENTS IN A DATA CENTER

BACKGROUND

An Uninterruptible Power Supply (UPS) may be installed in a data center to provide back-up power in case of a mains power failure. The UPS is typically configured to provide sufficient power to allow the data center equipment to continue running while a generator is started. Lead-acid batteries are generally used to provide back-up power to the data center equipment. Lead-acid batteries are cheaper than other energy storage systems, and in UPS applications, lead-acid batteries are typically sized to provide enough power to allow the data center equipment to continue to run for a sufficient period of time to start a generator.

SUMMARY

Systems and methods are provided for reducing backup energy storage requirements in a data center. In one embodiment, to benefit from energy storage systems other than lead-acid batteries, such as supercapacitors and flywheels, and to implement other energy storage systems at a price comparable to the price of lead-acid batteries, systems and methods are provided to reduce the amount of stored energy required in a UPS. In one embodiment, systems and methods are provided for reducing the amount of energy used by data center equipment when back-up power is in use. In one example, the execution of virtual machines is suspended when back-up power is in use.

According to one aspect, a method for reducing a need for stored power in a backup mode of operation in a data center includes receiving a first alert from an uninterruptible power supply indicating that the uninterruptible power supply is operating in a first mode, wherein the first mode includes deriving power from a stored energy source, identifying at least one host server configured to receive power from the uninterruptible power supply in the first mode, suspending execution of at least one virtual machine on the at least one host server, receiving a second alert from the uninterruptible power supply indicating that the uninterruptible power supply is operating

in a second mode, wherein the second mode includes deriving power from one of a mains power source and a generator, and resuming execution of the at least one virtual machine.

According to one embodiment, the method also includes, in response to the first alert, migrating operations on the at least one host server to at least one second host server. In another embodiment, the method includes activating a power-capping process in the at least one host server. In a further embodiment, the at least one virtual machine includes a plurality of virtual machines, and the method includes calculating a level of importance for each of the plurality of virtual machines, and suspending execution of a portion of the plurality of virtual machines, based on the level of importance of each virtual machine.

In one embodiment, deriving power from a stored energy source includes deriving power from one of a super capacitor, a lithium ion battery, a flywheel energy source, and a lead-acid battery. In another embodiment, identifying at least one host server configured to receive power from the uninterruptible power supply in the first mode includes identifying at least one host server having a frequency-scaling processor.

According to one aspect, a system for reducing a need for power in a backup mode of operation in a data center includes an uninterruptible power supply including a stored energy source, a plurality of host servers, each host server including at least one virtual machine and a controller coupled to the uninterruptible power supply and the plurality of host servers. The controller is configured to receive a first alert from the uninterruptible power supply indicating that the uninterruptible power supply is operating in a first mode, wherein the first mode includes deriving power from the stored energy source, identify at least one of the plurality of host servers configured to receive power from the uninterruptible power supply in the first mode, and suspend execution of at least one virtual machine on the at least one of the plurality of host servers.

According to one embodiment, the controller is further configured to receive a second alert from the uninterruptible power supply indicating that the uninterruptible power supply is operating in a second mode, wherein the second mode includes deriving power from one of a mains power source and a generator, and resume execution of the at least one virtual machine. In another embodiment, the controller is further configured, in response to the first alert, to migrate operations on the at least one host server to at least

one second host server. In a further embodiment, the controller is further configured, in response to the first alert, to migrate operations on the at least one host server to at least one second host server. In another embodiment, the controller is further configured to activate a power-capping process in the at least one host server.

In one embodiment, the at least one virtual machine includes a plurality of virtual machines, and the controller is further configured to calculate a level of importance for each of the plurality of virtual machines, and suspend execution of a portion of the plurality of virtual machines, based on the level of importance of each virtual machine. In another embodiment, the at least one host server has a frequency-scaling processor. In one embodiment, the stored energy source is one of a super capacitor, a lithium-ion based energy source, a flywheel energy source, and a lead-acid battery.

According to another aspect, a computer readable medium has stored thereon sequences of instruction for reducing a need for power in a backup mode of operation including instructions that will cause a processor to receive a first alert from an uninterruptible power supply indicating that the uninterruptible power supply is operating in a first mode, wherein the first mode includes deriving power from a stored energy source, identify at least one of the plurality of host servers configured to receive power from the uninterruptible power supply in the first mode, and suspend execution of at least one virtual machine on the at least one of the plurality of host servers.

In one embodiment, the computer readable medium also includes instructions that will cause the processor to receive a second alert from the uninterrupted power supply indicating that the uninterrupted power supply is operating in a second mode, wherein the second mode includes deriving power from one of a mains power source and a generator, and resume execution of the at least one virtual machine. In another embodiment, the computer readable medium also includes instructions that will cause the processor, in response to the first alert, to migrate operations on the at least one host server to at least one second host server. In a further embodiment, the computer readable medium also includes instructions that will cause the processor to activate a power-capping process in the at least one host server.

According to one embodiment, the at least one virtual machine includes a plurality of virtual machines, and includes instructions that will cause the processor to

calculate a level of importance for each of the plurality of virtual machines, and suspend execution of a portion of the plurality of virtual machines, based on the level of importance of each virtual machine. According to another embodiment, deriving power from a stored energy source includes deriving power from one of a super capacitor, a lithium ion battery, a flywheel energy source, and a lead-acid battery.

BRIEF DESCRIPTION OF THE FIGURES

The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

FIG. 1 is a block diagram of one example of a computer system with which various aspects in accord with the present invention may be implemented;

FIG. 2 a schematic of one example of a distributed system including a data center management system;

FIG. 3 is a schematic block diagram showing an uninterruptible power supply providing power to data center servers;

FIG. 4 is a schematic block diagram showing an uninterruptible power supply and data center servers; and

FIG. 5 is a flow chart of a method of reducing the amount of energy storage in a data center.

DETAILED DESCRIPTION

At least some embodiments relate to systems and methods for reducing backup storage requirements in a data center. In one embodiment, systems and methods are provided for reducing the amount of energy used by data center equipment when operating from backup power by suspending execution of one or more virtual machines executing on the equipment. According to one feature, reducing the amount of energy used by data center equipment when the equipment loses mains power reduces the amount of energy backup storage needed in the data center. Furthermore, when a data

center uses less stored energy upon mains failure, different types of stored energy devices may be used, such as stored energy modules that provide less overall stored energy.

There are several systems and methods known for reducing energy used in computing equipment. One method for reducing the energy used by a host server includes frequency-scaling the central processing unit. Frequency scaling of a CPU lowers the operating frequency of the CPU, and thus the power draw of the CPU. By decreasing the operating frequency of a CPU when computational load is decreased, the CPU will draw less power.

Another method of reducing the energy used by selected host servers includes migrating operations from the selected host servers to other servers. However, while migrating operations results in less power draw for the selected host servers, the migration operation can draw significant amounts of power. Further, for the migration to occur, powered servers having capacity must be available.

A further method of reducing the energy used by servers is power-capping, which allows a user to limit the maximum power draw of a server to a specific value. While power-capping can be used to lower the power usage of servers that are powered by a UPS, power-capping technology does not detect which servers are impacted by a power failure, or calculate the power-failure impact on specific servers.

A software-based method for reducing energy used by selected servers in accordance with one embodiment includes detecting which servers have lost mains power and are operating on stored energy, and suspending the execution of one or more virtual machines running on the detected servers. Suspending the execution of one or more virtual machines on a server lowers the power draw of the server, resulting in a higher stored energy runtime.

Any of the systems or methods disclosed herein may be used in combination to further reduce energy usage of one or more servers, thereby decreasing the amount of stored energy needed in case of a power mains failure. For example, the UPS load may be lowered by switching off or power-capping servers in accordance with Service Level Agreements, and the deployment of virtual machines may be managed to ensure compliance with a Service Level Agreement's required runtime of each virtual machine.

The aspects and embodiments of the present invention disclosed herein are not limited in their application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. These aspects are capable of assuming other embodiments and of being practiced or of being carried out in various ways. Examples of specific implementations are provided herein for illustrative purposes only and are not intended to be limiting. In particular, acts, elements and features discussed in connection with any one or more embodiments are not intended to be excluded from a similar role in any other embodiments.

For example, according to one embodiment of the present invention, a computer system is configured to perform any of the functions described herein, including but not limited to, detecting a mains power failure, determining the power draw of various servers and suspending the execution of one or more virtual machines on selected servers. Further, computer systems in embodiments may be used to automatically monitor power usage of servers in a data center. Moreover, the systems described herein may be configured to include or exclude any of the functions discussed herein. Thus the embodiments are not limited to a specific function or set of functions. Also, the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use herein of "including," "comprising," "having," "containing," "involving," and variations thereof is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

Computer System

Various aspects and functions described herein in accordance with the present embodiments may be implemented as hardware or software on one or more computer systems. There are many examples of computer systems currently in use. These examples include, among others, network appliances, personal computers, workstations, mainframes, networked clients, servers, media servers, application servers, database servers and web servers. Other examples of computer systems may include mobile computing devices, such as cellular phones and personal digital assistants, and network equipment, such as load balancers, routers and switches. Further, aspects in accordance with the present embodiments may be located on a single computer system or may be

distributed among a plurality of computer systems connected to one or more communications networks.

For example, various aspects and functions may be distributed among one or more computer systems configured to provide a service to one or more client computers, or to perform an overall task as part of a distributed system. Additionally, aspects may be performed on a client-server or multi-tier system that includes components distributed among one or more server systems that perform various functions. Thus, the embodiments are not limited to executing on any particular system or group of systems. Further, aspects may be implemented in software, hardware or firmware, or any combination thereof. Thus, aspects in accordance with the present embodiments may be implemented within methods, acts, systems, system elements and components using a variety of hardware and software configurations, and the embodiments are not limited to any particular distributed architecture, network, or communication protocol.

FIG. 1 shows a block diagram of a distributed computer system 100, in which various aspects and functions in accord with the present embodiments may be practiced. Distributed computer system 100 may include one more computer systems. For example, as illustrated, distributed computer system 100 includes computer systems 102, 104, and 106. As shown, computer systems 102, 104, and 106 are interconnected by, and may exchange data through, communication network 108. Network 108 may include any communication network through which computer systems may exchange data. To exchange data using network 108, computer systems 102, 104, and 106 and network 108 may use various methods, protocols and standards, including, among others, token ring, Ethernet, wireless Ethernet, Bluetooth, TCP/IP, UDP, Http, FTP, SNMP, SMS, MMS, SS7, Json, Soap, and Corba. To ensure data transfer is secure, computer systems 102, 104, and 106 may transmit data via network 108 using a variety of security measures including TLS, SSL or VPN among other security techniques. While distributed computer system 100 illustrates three networked computer systems, distributed computer system 100 may include any number of computer systems and computing devices, networked using any medium and communication protocol.

Various aspects and functions in accordance with the present embodiments may be implemented as specialized hardware or software executing in one or more computer

systems including computer system 102 shown in FIG. 1. As depicted, computer system 102 includes processor 110, memory 112, bus 114, interface 116, and storage 118. Processor 110 may perform a series of instructions that result in manipulated data. Processor 110 may be a commercially available processor such as an Intel Pentium, Motorola PowerPC, SGI MIPS, Sun UltraSPARC, or Hewlett-Packard PA-RISC processor, but may be any type of processor, multi-processor, microprocessor or controller as many other processors and controllers are available. Processor 110 is connected to other system elements, including one or more memory devices 112, by bus 114.

Memory 112 may be used for storing programs and data during operation of computer system 102. Thus, memory 112 may be a relatively high performance, volatile, random access memory such as a dynamic random access memory (DRAM) or static memory (SRAM). However, memory 112 may include any device for storing data, such as a disk drive or other non-volatile, non-transitory, storage device. Various embodiments in accordance with the present invention may organize memory 112 into particularized and, in some cases, unique structures to perform the aspects and functions disclosed herein.

Components of computer system 102 may be coupled by an interconnection element such as bus 114. Bus 114 may include one or more physical busses, for example, busses between components that are integrated within a same machine, but may include any communication coupling between system elements including specialized or standard computing bus technologies such as IDE, SCSI, PCI and InfiniBand. Thus, bus 114 enables communications, for example, data and instructions, to be exchanged between system components of computer system 102.

Computer system 102 also includes one or more interface devices 116 such as input devices, output devices and combination input/output devices. Interface devices may receive input or provide output. More particularly, output devices may render information for external presentation. Input devices may accept information from external sources. Examples of interface devices include keyboards, mouse devices, trackballs, microphones, touch screens, printing devices, display screens, speakers,

network interface cards, etc. Interface devices allow computer system 102 to exchange information and communicate with external entities, such as users and other systems. Storage system 118 may include a computer readable and writeable, nonvolatile, non-transitory, storage medium in which instructions are stored that define a program to be executed by the processor. Storage system 118 also may include information that is recorded, on or in, the medium, and this information may be processed by the program. More specifically, the information may be stored in one or more data structures specifically configured to conserve storage space or increase data exchange performance. The instructions may be persistently stored as encoded signals, and the instructions may cause a processor to perform any of the functions described herein. The medium may, for example, be optical disk, magnetic disk, or flash memory, among others. In operation, the processor or some other controller may cause data to be read from the nonvolatile recording medium into another memory, such as memory 112, that allows for faster access to the information by the processor than does the storage medium included in storage system 118. The memory may be located in storage system 118 or in memory 112, however, processor 110 may manipulate the data within the memory 112, and then may copy the data to the medium associated with storage system 118 after processing is completed. A variety of components may manage data movement between the medium and integrated circuit memory element and the presently described embodiments are not limited thereto. Further, the embodiments are not limited to a particular memory system or data storage system.

Although computer system 102 is shown by way of example as one type of computer system upon which various aspects and functions in accordance with the present embodiments may be practiced, any aspects of the presently disclosed embodiments are not limited to being implemented on the computer system as shown in FIG. 1. Various aspects and functions in accord with the presently disclosed embodiments may be practiced on one or more computers having a different architectures or components than that shown in FIG. 1. For instance, computer system 102 may include specially-programmed, special-purpose hardware, such as for example, an application-specific integrated circuit (ASIC) tailored to perform a particular operation disclosed herein. While another embodiment may perform the same function using

several general-purpose computing devices running MAC OS System X with Motorola PowerPC processors and several specialized computing devices running proprietary hardware and operating systems.

Computer system 102 may be a computer system including an operating system that manages at least a portion of the hardware elements included in computer system 102. Usually, a processor or controller, such as processor 110, executes an operating system which may be, for example, a Windows-based operating system such as Windows NT, Windows 2000 (Windows ME), Windows XP, or Windows Vista operating systems, available from the Microsoft Corporation, a MAC OS System X operating system available from Apple Computer, one of many Linux-based operating system distributions, for example, the Enterprise Linux operating system available from Red Hat Inc., a Solaris operating system available from Sun Microsystems, or a UNIX operating system available from various sources. Many other operating systems may be used, and embodiments are not limited to any particular implementation.

The processor and operating system together define a computer platform for which application programs in high-level programming languages may be written. These component applications may be executable, intermediate, for example, C-, bytecode or interpreted code which communicates over a communication network, for example, the Internet, using a communication protocol, for example, TCP/IP. Similarly, aspects in accord with the presently disclosed embodiments may be implemented using an object-oriented programming language, such as .Net, SmallTalk, Java, C++, Ada, or C# (C-Sharp). Other object-oriented programming languages may also be used. Alternatively, functional, scripting, or logical programming languages may be used.

Additionally, various aspects and functions in accordance with the presently disclosed embodiments may be implemented in a non-programmed environment, for example, documents created in HTML, XML or other format that, when viewed in a window of a browser program, render aspects of a graphical-user interface or perform other functions. Further, various embodiments in accord with the present invention may be implemented as programmed or non-programmed elements, or any combination thereof. For example, a web page may be implemented using HTML while a data object called from within the web page may be written in C++. Thus, the presently disclosed

embodiments are not limited to a specific programming language and any suitable programming language could also be used.

A computer system included within an embodiment may perform additional functions outside the scope of the presently disclosed embodiments. For instance, aspects of the system may be implemented using an existing commercial product, such as, for example, Database Management Systems such as SQL Server available from Microsoft of Seattle WA., Oracle Database from Oracle of Redwood Shores, CA, and MySQL from MySQL AB, a subsidiary of Oracle or integration software such as Web Sphere middleware from IBM of Armonk, NY. However, a computer system running, for example, SQL Server may be able to support both aspects in accord with the presently disclosed embodiments and databases for sundry applications.

Example System Architecture

FIG. 2 presents a context diagram including physical and logical elements of distributed system 200. As shown, distributed system 200 is specially configured in accordance with the presently disclosed embodiments and may be used to control servers in a data center in accordance with methods described herein. The system structure and content recited with regard to FIG. 2 is for exemplary purposes only and is not intended to limit the embodiments to the specific structure shown in FIG. 2. As will be apparent to one of ordinary skill in the art, many variant system structures can be architected without deviating from the scope of the presently disclosed embodiments. The particular arrangement presented in FIG. 2 was chosen to promote clarity.

Information may flow between the elements, components and subsystems depicted in FIG. 2 using any technique. Such techniques include, for example, passing the information over the network via TCP/IP, passing the information between modules in memory and passing the information by writing to a file, database, or some other non-volatile storage device. Other techniques and protocols may be used without departing from the scope of the presently disclosed embodiments.

Referring to FIG. 2, system 200 includes user 202, interface 204, data center design and management system 206, communications network 208, and data center database 210. System 200 may allow user 202, such as a data center architect or other

data center personnel, to interact with interface 204 to create or modify a model of one or more data center configurations. According to one embodiment, interface 204 may include aspects of the floor editor and the rack editor as disclosed in Patent Cooperation Treaty Application No. PCT/US08/63675, titled "Methods and Systems for Managing Facility Power and Cooling," filed on May 15, 2008, which is incorporated herein by reference in its entirety and is hereinafter referred to as PCT/US08/63675. In at least one embodiment, information regarding a data center is entered into system 200 through the interface, and assessments and recommendations for the data center are provided to the user. Further, in at least one embodiment, optimization processes may be performed to optimize cooling performance and energy usage of the data center.

As shown in FIG. 2, data center design and management system 206 presents data design interface 204 to user 202. According to one embodiment, data center design and management system 206 may include the data center design and management system as disclosed in PCT/US08/63675. In this embodiment, design interface 204 may incorporate functionality of the input module, the display module and the builder module included in PCT/US08/63675 and may use the database module to store and retrieve data. As illustrated, data center design and management system 206 may exchange information with data center database 210 via network 208. This information may include any information needed to support the features and functions of data center design and management system 206. For example, in one embodiment, data center database 210 may include at least some portion of the data stored in the data center equipment database described in PCT/US08/63675. In another embodiment, this information may include any information needed to support interface 204, such as, among other data, the physical layout of one or more data center model configurations, the production and distribution characteristics of the cooling providers included in the model configurations, the consumption characteristics of the cooling consumers in the model configurations, and a listing of equipment racks and cooling providers to be included in a cluster.

Data center database 210 may take the form of any logical construction capable of storing information on a computer readable medium including, among other structures, flat files, indexed files, hierarchical databases, relational databases or object oriented databases. The data may be modeled using unique and foreign key relationships and

indexes. The unique and foreign key relationships and indexes may be established between the various fields and tables to ensure both data integrity and data interchange performance.

The computer systems shown in FIG. 2, which include data center design and management system 206, network 208 and data center equipment database 210, each may include one or more computer systems. As discussed above with regard to FIG. 1, computer systems may have one or more processors or controllers, memory and interface devices. The particular configuration of system 200 depicted in FIG. 2 is used for illustration purposes only and embodiments of the invention may be practiced in other contexts. Thus, embodiments of the invention are not limited to a specific number of users or systems.

FIG. 3 is a schematic block diagram 300 showing an uninterruptible power supply (UPS) 302 receiving input power 308 and providing power 306 to data center servers 304 in a data center. In various examples, the input power 308 may be mains power or it may be generator power. Data center infrastructure management software 312, operating for example on a management server in the data center, monitors the UPS 302. Virtualization management software 312 manages and monitors virtual machines running on the data center servers 304. The data center infrastructure management software 310 is integrated with the virtualization infrastructure management software 312, with the data center infrastructure management software 310 communicating with, and exchanging information with, the virtualization infrastructure management software 312 over existing network protocols (for example, TCP/IP-based web services).

FIG. 4 is a schematic block diagram 350 showing the UPS 302 and the data center servers 304 after disruption of input power 308. As indicated by the "x" over the power arrow 308, the UPS 302 has lost input power, and is providing power 306 to the data center servers 304 from a stored energy source. Upon detection of the disruption of input power 308, the management server running the data center infrastructure management software 310 controls the host servers 304 to operate in a reduced power mode of operation as will now be described with reference to FIG. 5, which includes a flow chart of a method 400 for reducing power requirements in a data center when operating in a backup mode of operation.

In the method 400, at block 402, the data center infrastructure management software 310 receives an alert from the UPS 302 indicating that the UPS is operating in a first, stored energy, mode. At block 404, the data center infrastructure management software 310 identifies which host servers 304 are receiving power from the UPS 302.

At block 406, the data center infrastructure management software 310 instructs the virtualization management software 312 to suspend execution of one or more virtual machines on the host servers 304 receiving power from the UPS 302. In one example, it takes less than one second for the data center infrastructure management software 310 to receive the alert from the UPS 302 and suspend execution of one or more virtual machines on the host servers 304. According to one embodiment, the data center infrastructure management software 310 may rank the importance of each virtual machine, and only suspend execution of virtual machines below a selected level of importance.

According to one feature, when execution of the virtual machines is suspended, the computational load of the host servers 304 drops to near-zero. When there is almost no computational load on the host servers 304, the host servers 304 will throttle their computational power, lowering their power draw by between about 50% to about 80%. According to one example, the decrease in the power draw by the host servers 304 depends on the server type. In another example, the decrease in the power draw by the host servers 304 depends on the server load. According to one feature, because the power draw of the host servers 304 drops by between about 50% to about 80%, the energy stored in the back-up power in the UPS 302 equates to a longer runtime. Alternatively, a UPS with a back-up power source having a shorter runtime may be used.

In one embodiment, in response to the alert from the UPS 302 indicating that the UPS is operating in the first, stored energy, mode, the data center infrastructure management software 310 activates a power-capping process in one or more of the host servers 304. The power-capping process limits the amount of power a host server can draw. In another embodiment, in response to the alert from the UPS 302 indicating that the UPS is operating in the stored energy mode, the data center infrastructure management software 310 migrates one or more operations from one or more of the host

servers 304 to one or more second host servers, where the second host servers are not running in a stored energy mode.

At block 408, the data center infrastructure management software 310 receives a second alert from the UPS 302 indicating that input power has been restored to the UPS 302. In some examples, the UPS 302 may be receiving power from a mains power source, or the UPS 302 may be receiving power from a generator. At block 410, the data center infrastructure management software 310 instructs the virtual management software 312 to resume the execution of virtual machines that were previously suspended on the host servers 304 powered by the UPS.

According to one aspect, suspending virtual machines prevents data loss of applications running on the virtual machines. In one embodiment, if power is not restored within a predetermined timeframe, one or more of the suspended virtual machines may have data offloaded to persistent storage. Upon restoration of power virtual machine execution can resume immediately from the data contained in persistent storage.

In one embodiment, because the power draw of the host servers 304 drops by between about 50% to about 80% using the methods described above, less energy may be stored in batteries or other storage devices in the UPS 302. When less back-up power is required by the UPS 302, there are a wider variety of suitable back-up storage options that may be used in the UPS. In various examples, the storage options include a super capacitor, a Lithium-Ion battery, or a flywheel energy source. These alternative back-up storage options may offer several advantages. For example, the alternative backup storage devices can be run at higher temperatures with less derating, they are smaller and take up less floor space, and they weigh less than traditional lead-acid batteries. According to one example, these advantages make the alternative storage devices especially desirable for use in urban datacenters, where the cost per square foot of space is expensive.

In at least some embodiments described above, the operation of virtual machines on computer servers is suspended during operation of the computer servers from backup power. The suspension of operation will typically cause one or more applications to be suspended. In at least one embodiment, virtual machines are selected for suspension

based on the applications running on the virtual machines. For example, although a particular application may be designated as a critical application, suspension of the application during a transition from mains power to generator power may nonetheless be acceptable, and more desirable than terminating the application.

Having thus described several aspects of at least one embodiment, it is to be appreciated various alterations, modifications, and improvements will readily occur to those skilled in the art. Such alterations, modifications, and improvements are intended to be part of this disclosure and are intended to be within the scope of the disclosure. Accordingly, the foregoing description and drawings are by way of example only, and the scope of the disclosure should be determined from proper construction of the appended claims, and their equivalents.

What is claimed is:

CLAIMS

1. A method for reducing a need for stored power in a backup mode of operation in a data center, comprising:

receiving a first alert from an uninterruptible power supply indicating that the uninterruptible power supply is operating in a first mode, wherein the first mode includes deriving power from a stored energy source;

identifying at least one host server configured to receive power from the uninterruptible power supply in the first mode;

suspending execution of at least one virtual machine on the at least one host server;

receiving a second alert from the uninterruptible power supply indicating that the uninterruptible power supply is operating in a second mode, wherein the second mode includes deriving power from one of a mains power source and a generator; and

resuming execution of the at least one virtual machine.

2. The method of claim 1, further comprising, in response to the first alert, migrating operations on the at least one host server to at least one second host server.

3. The method of claim 1, further comprising activating a power-capping process in the at least one host server.

4. The method of claim 1, wherein the at least one virtual machine includes a plurality of virtual machines, and wherein the method further comprises:

calculating a level of importance for each of the plurality of virtual machines; and

suspending execution of a portion of the plurality of virtual machines, based on the level of importance of each virtual machine.

5. The method of claim 1, wherein deriving power from a stored energy source includes deriving power from one of a super capacitor, a lithium ion battery, a flywheel energy source, and a lead-acid battery.

6. The method of claim 1, wherein identifying at least one host server configured to receive power from the uninterruptible power supply in the first mode includes identifying at least one host server having a frequency-scaling processor.

7. A system for reducing a need for power in a backup mode of operation in a data center, comprising:

an uninterruptible power supply including a stored energy source;

a plurality of host servers, each host server including at least one virtual machine;

a controller coupled to the uninterruptible power supply and the plurality of host servers, configured to:

receive a first alert from the uninterruptible power supply indicating that the uninterruptible power supply is operating in a first mode, wherein the first mode includes deriving power from the stored energy source;

identify at least one of the plurality of host servers configured to receive power from the uninterruptible power supply in the first mode; and

suspend execution of at least one virtual machine on the at least one of the plurality of host servers.

8. The system of claim 7, wherein the controller is further configured to:

receive a second alert from the uninterruptible power supply indicating that the uninterruptible power supply is operating in a second mode, wherein the second mode includes deriving power from one of a mains power source and a generator; and

resume execution of the at least one virtual machine.

9. The system of claim 7, wherein the stored energy source is one of a super capacitor, a lithium-ion based energy source, a flywheel energy source, and a lead-acid battery.

10. The system of claim 7, wherein the controller is further configured, in response to the first alert, to migrate operations on the at least one host server to at least one second host server.

11. The system of claim 1, wherein the at least one virtual machine includes a plurality of virtual machines, and wherein the controller is further configured to:

calculate a level of importance for each of the plurality of virtual machines; and

suspend execution of a portion of the plurality of virtual machines, based on the level of importance of each virtual machine.

12. The system of claim 7, wherein the at least one host server has a frequency-scaling processor.

13. The system of claim 7, wherein the controller is further configured to activate a power-capping process in the at least one host server.

14. A computer readable medium having stored thereon sequences of instruction for reducing a need for power in a backup mode of operation including instructions that will cause a processor to:

receive a first alert from an uninterruptible power supply indicating that the uninterruptible power supply is operating in a first mode, wherein the first mode includes deriving power from a stored energy source;

identify at least one of the plurality of host servers configured to receive power from the uninterruptible power supply in the first mode; and

suspend execution of at least one virtual machine on the at least one of the plurality of host servers.

15. The computer readable medium of claim 14, further comprising instructions that will cause the processor to:

receive a second alert from the uninterrupted power supply indicating that the uninterrupted power supply is operating in a second mode, wherein the second mode includes deriving power from one of a mains power source and a generator; and

resume execution of the at least one virtual machine.

16. The computer readable medium of claim 14, further comprising instructions that will cause the processor, in response to the first alert, to migrate operations on the at least one host server to at least one second host server.

17. The computer readable medium of claim 14, wherein deriving power from a stored energy source includes deriving power from one of a super capacitor, a lithium ion battery, a flywheel energy source, and a lead-acid battery.

18. The computer readable medium of claim 14, further comprising instructions that will cause the processor to activate a power-capping process in the at least one host server.

19. The computer readable medium of claim 14, wherein the at least one virtual machine includes a plurality of virtual machines, and further comprising instructions that will cause the processor to:

calculate a level of importance for each of the plurality of virtual machines; and

suspend execution of a portion of the plurality of virtual machines, based on the level of importance of each virtual machine.

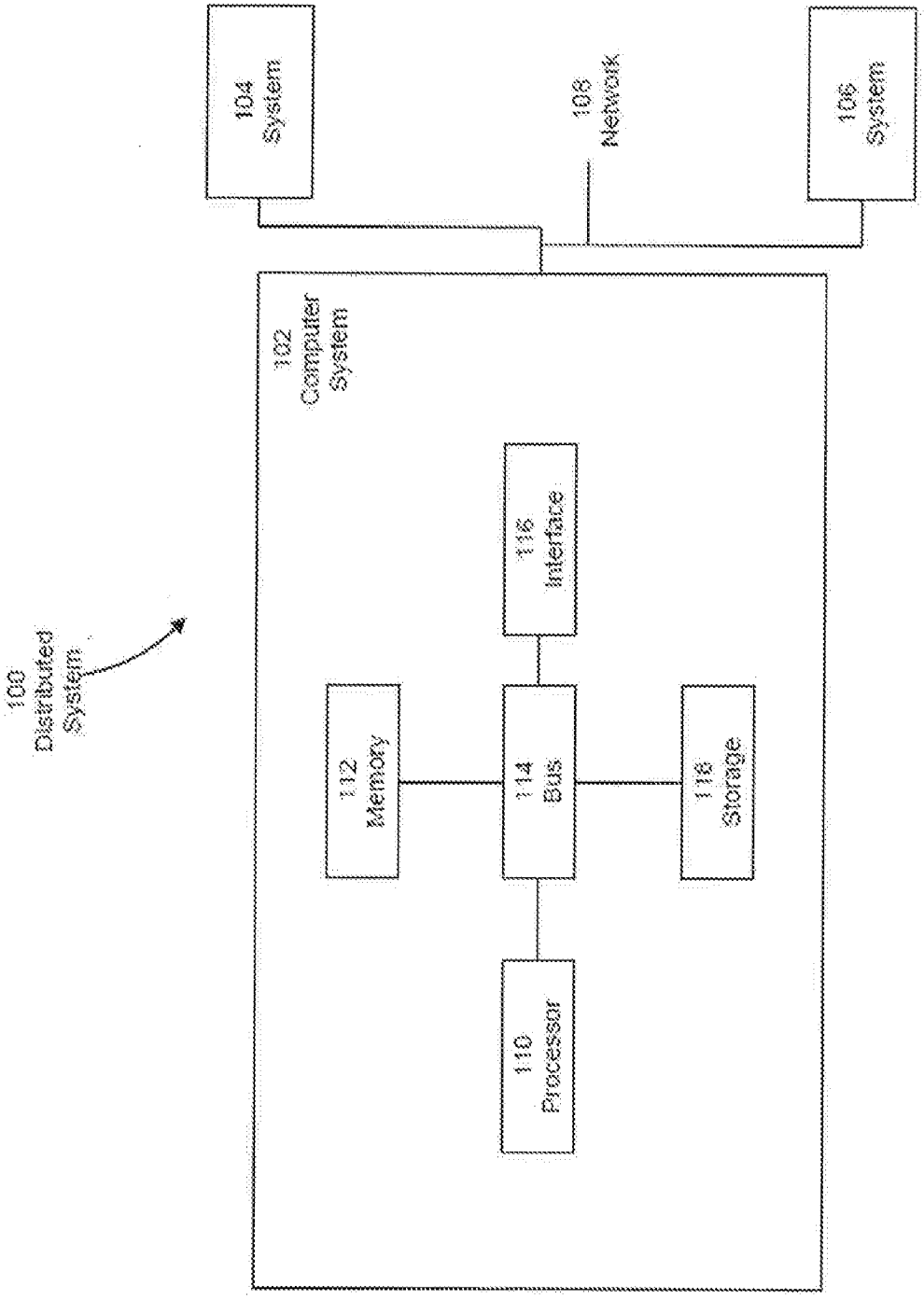

100
Distributed
System
102
Computer
System
104
System
106
System
108
Network
110
Processor
112
Memory
114
Bus
116
Interface
118
Storage

FIG. 1

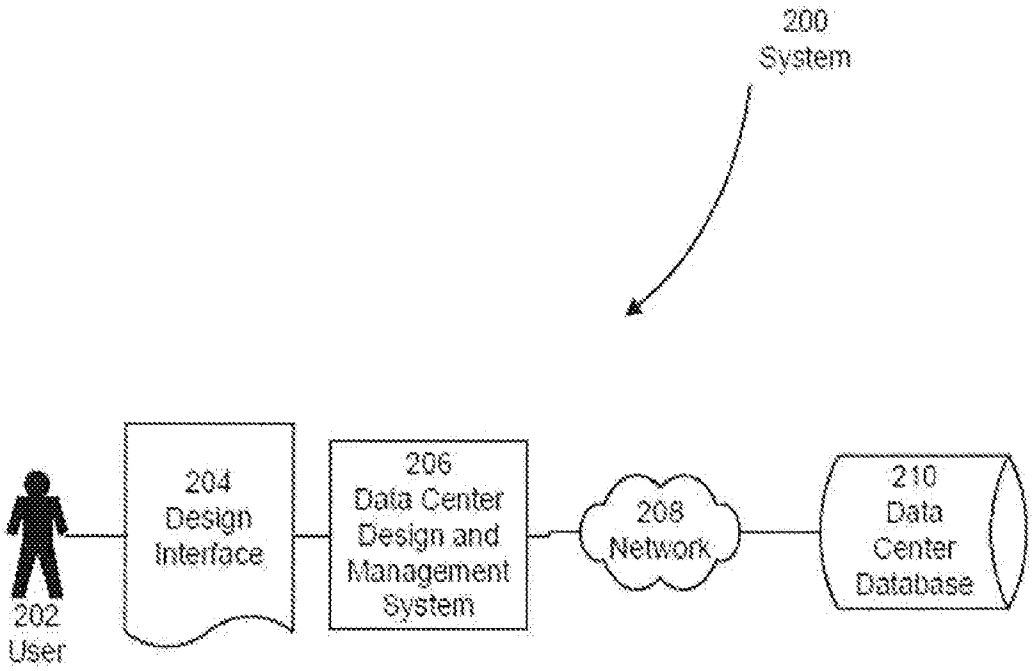

200
System
202
User
204
Design
Interface
206
Data Center
Design and
Management
System
208
Network
210
Data
Center
Database

FIG. 2

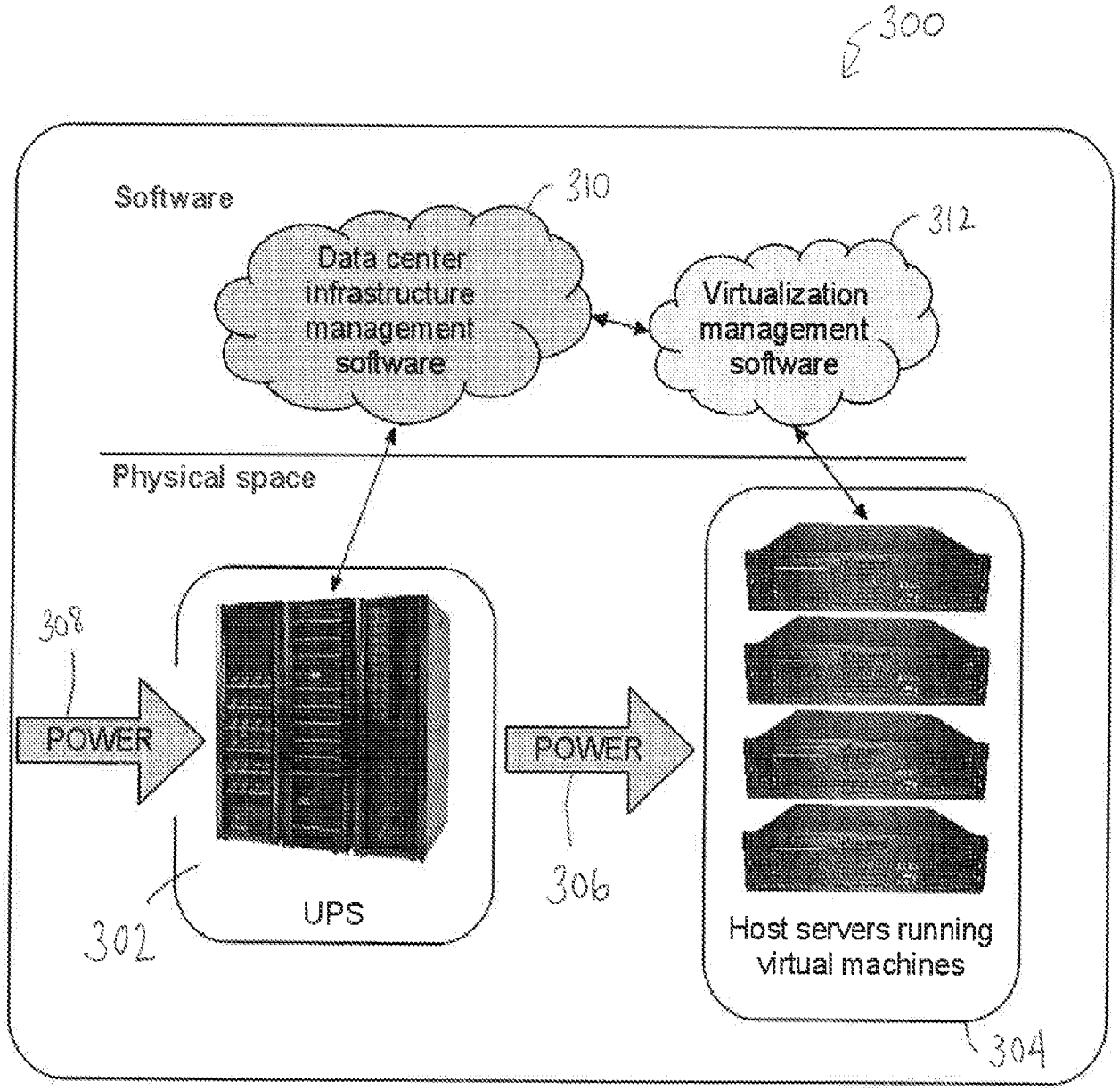

300
Software
310
Data center infrastructure management software
312
Virtualization management software
Physical space
308
POWER
UPS
302
POWER
306
Host servers running virtual machines
304

FIG. 3

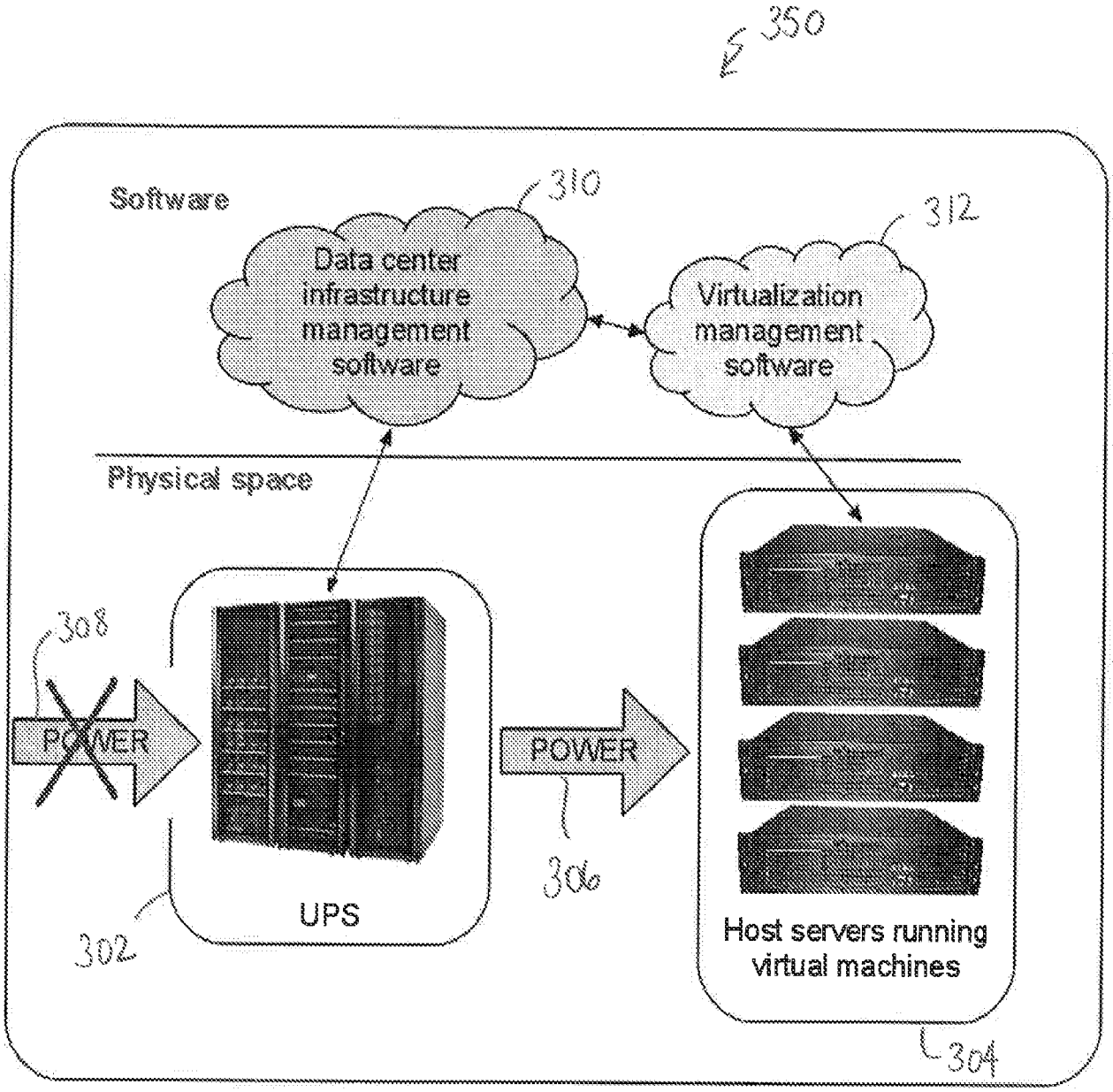

350
Software
310
Data center infrastructure management software
312
Virtualization management software
Physical space
308
POWER
UPS
302
POWER
306
Host servers running virtual machines
304

FIG. 4

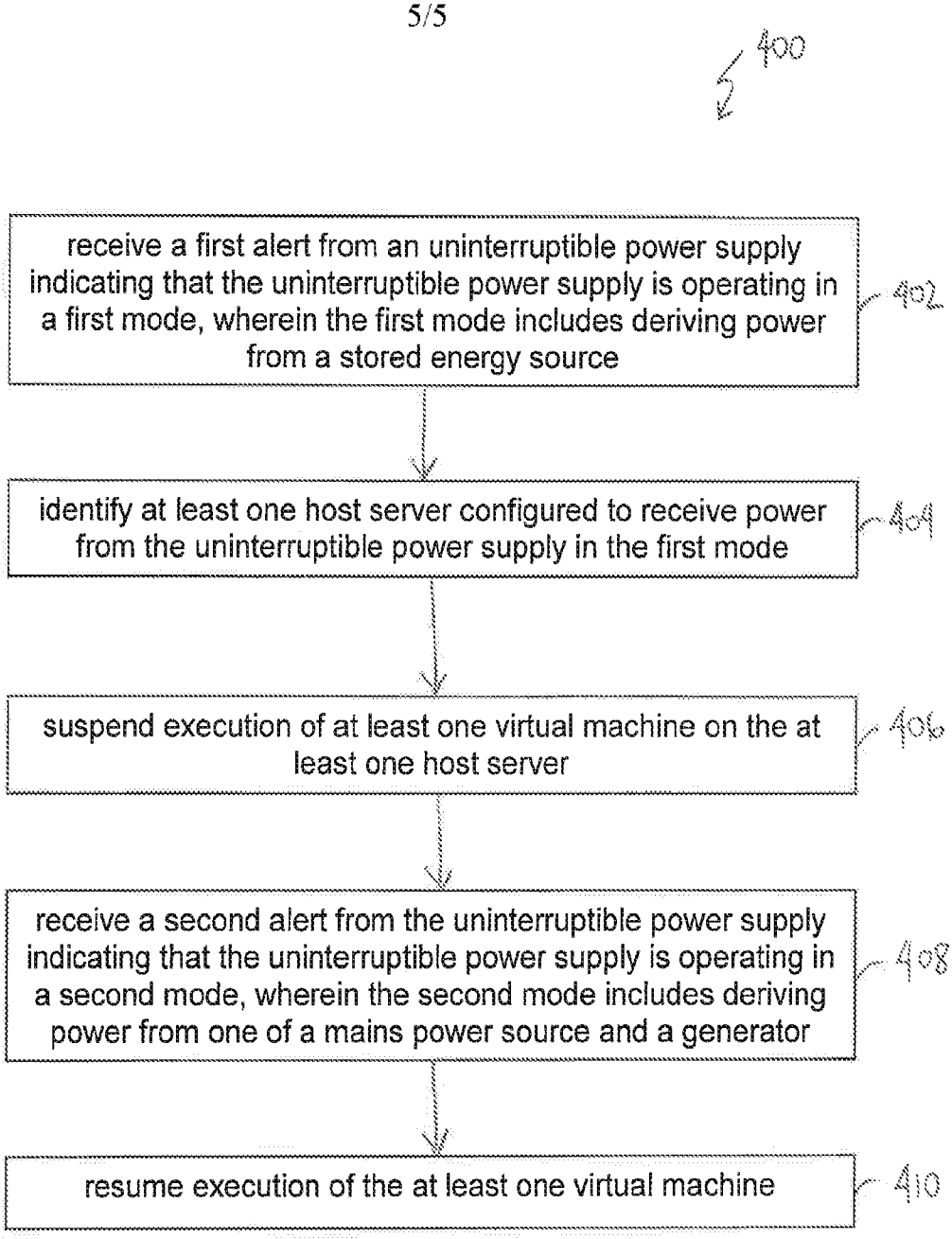

400
receive a first alert from an uninterruptible power supply indicating that the uninterruptible power supply is operating in a first mode, wherein the first mode includes deriving power from a stored energy source
402
identify at least one host server configured to receive power from the uninterruptible power supply in the first mode
404
suspend execution of at least one virtual machine on the at least one host server
406
receive a second alert from the uninterruptible power supply indicating that the uninterruptible power supply is operating in a second mode, wherein the second mode includes deriving power from one of a mains power source and a generator
408
resume execution of the at least one virtual machine
410

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20 11/066832

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 1/32 (201 2.01)
USPC - 713/324
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)
IPC(8) - G06 F 1/00, 1/26, 1/28, 1/30, 1/32 (2012.01)
USPC - 713/300, 320-21 , 324; 714/13-14

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)
MicroPatent, Google Patents, Google Scholar

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 201 1/0072430 A 1 (MANI) 24 March 201 1 (24.03.201 1) entire document	1-19
Y	US 7,451 ,336 B2 (MANUELL et al) 11 November 2008 (11.11.2008) entire document	1-6, 8, 15
Y	US 2009/0249354 A1(YAMAGUCHI et al) 01 October 2009 (01 .10.2009) entire document	7-19
Y	US 2007/0234077 A 1 (ROTHMAN et al) 04 October 2007 (04.10.2007) entire document	2, 10, 16
Y	US 201 1/003561 1 A 1 (BREY et al) 10 February 201 1 (10.02.201 1) entire document	3, 13, 18
Y	US 7,432,617 B2 (KANOUDA.et al) 07 October 2008 (07.10.2008) entire document	5, 9, 17
Y	US 201 1/023901 0 A 1 (JAIN et al) 29 September 201 1 (29.09.201 1) entire document	6, 12
A	US 201 1/0018342 A 1 (PARK et al) 27 January 201 1 (27.01 .201 1) entire document	1-19
A	US 2009/01 13060 A 1 (JACOB et al) 30 April 2009 (30.04.2009) entire document	1-19
A	US 7,703,046 B2 (FALLON et al) 20 April 2010 (20.04.2010) entire document	1-19
A	US 2010/0262974 A 1 (UYEDA) 14 October 2010 (14.10.2010) entire document	1-19
A	US 2010/0141046 A 1 (PAIK) 10 June 2010 (10.06.2010) entire document	1-19

☐ 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
"&" 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 April 2012	0 3 MAY 2012

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774