

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
6 August 2009 (06.08.2009)

PCT

(10) International Publication Number
WO 2009/096944 A1

(51) International Patent Classification:
G06F 9/24 (2006.01) **G06F 15/16** (2006.01)

(21) International Application Number:
PCT/US2008/052235

(22) International Filing Date: 28 January 2008 (28.01.2008)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US):
HEWLETT-PACKARD DEVELOPMENT COMPANY, L.P. [—/US]; 11445 Compaq Center Drive West, Houston, TX 77070 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WATKINS, Mark Robert** [GB/GB]; MS 27, Filton Rd. Stoke Giford, Bristol BS34 8QZ (GB). **CULTER, Bradley** [US/US]; 20555 Tomball Parkway, Houston, Texas 77070 (US).

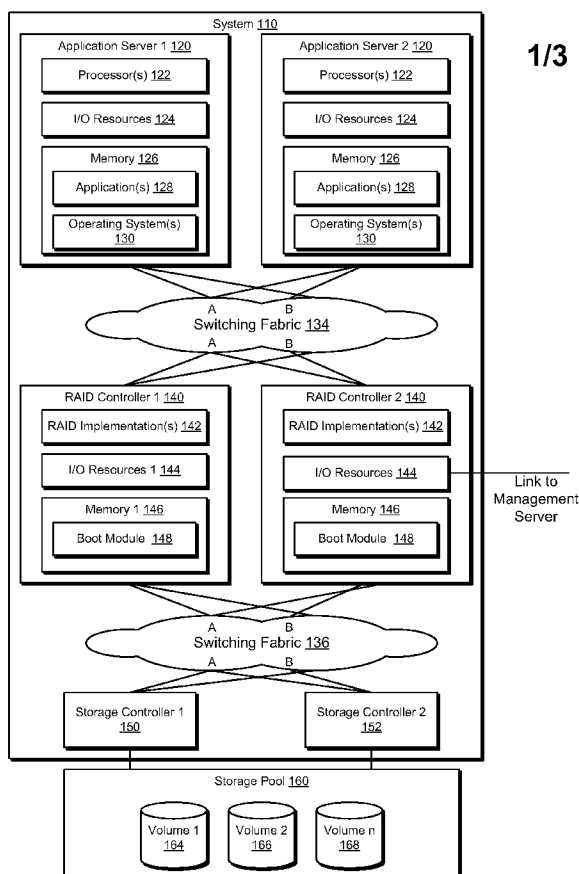
(74) Agent: **WEBB, Steven**; Hewlett-Packard Company, Intellectual Property Administration, Mail Stop 35, Fort Collins, Colorado 80527-2400 (US).

(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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL,

[Continued on next page]

(54) Title: DEPLOYMENT OF BOOT IMAGES IN DISKLESS SERVERS



(57) Abstract: In one embodiment, a system, comprises a first computer system comprising at least a first diskless server, at least a first RAID controller coupled to the first diskless server, at least a first storage pool coupled to the RAID controller, and a remote management server coupled to the RAID controller via an out-of-band communication link. The remote management server comprises a boot management module which, when executed, initiates a command to instruct the RAID controller to create at least a first logical volume in a memory module coupled to the RAID controller, transmits the command to the RAID controller via the out-of-band communication link, and transmits a boot image from the remote management server to the RAID controller via the out-of-band communication link. The RAID controller creates the first logical volume for the boot image in response to the command, and stores the boot image in the first logical volume.



NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG,
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *of inventorship (Rule 4.17(iv))*

Declarations under Rule 4.17:

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

Published:

- *with international search report*

DEPLOYMENT OF BOOT IMAGES IN DISKLESS SERVERS

BACKGROUND

[0001] This application relates to electronic computing and more particularly to deployment of boot images in diskless servers.

[0002] Computing networks may be populated with servers and client computers. Servers are generally more powerful computers that provide common functions such as record sharing and Internet access to the client computers. Client computers may be fully functional computers, each having a processor, hard drive, CD ROM drive, floppy drive and system memory.

[0003] A central server may contain one or more blade computers, which are ultra-dense, low power blade computers designed to provide a high level of computing power in a relatively small space. In some applications hundreds of blade computers may be mounted in a single rack. Because blade computers consume less space, power, and produce less heat than conventional rack-mounted computers, they may result in significant cost savings. Additionally, blade computers may be connected in parallel to form computing engines of immense power.

[0004] Some blade computers utilize servers referred to as “diskless servers” which do not include a hard drive or other magnetic storage media. The absence of a hard drive or other magnetic storage media prevents a typical

disk-based boot operation from being implemented. Thus, additional techniques to implement boot operations in the blade computer systems would find utility.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Fig. 1 is a schematic illustration of a blade computer system, according to embodiments.

[0006] Fig. 2 is a schematic illustration of a remote boot management server system, according to embodiments.

[0007] Fig. 3 is a flowchart illustrating operations associated with deployment of boot images in diskless servers, according to some embodiments.

DETAILED DESCRIPTION

[0008] Described herein are exemplary systems and methods for deploying boot images in a diskless server. The methods described herein may be embodied as logic instructions on a computer-readable medium. When executed on one or more processor(s), the logic instructions cause the processor(s) to be programmed as a special-purpose machine that implements the described methods. The processor(s), when configured by the logic instructions to execute the methods recited herein, constitutes structure for performing the described methods.

[0009] Fig. 1 is a schematic illustration of a blade computer system 110, according to embodiments. Referring to Fig. 1, blade computer system 110 comprises at least one application server 120 coupled by a switching fabric 134 to at least one RAID controller 140, which is in turn coupled by a switching fabric 136 to at least one disk controller 150. The disk controller(s) 150 provide access to storage media in a storage pool 160 blade computer system 110.

[0010] In the embodiment depicted in Fig. 1, the blade computer system 110 comprises two application servers 120. In practice a blade computer system may comprise as few as a single application server 120. There is no theoretical upper limit to the number of application servers 120 which may be incorporated into a blade computer system. In practice, existing blade computer systems can support as many as sixty-four application servers 120 in an enclosure.

[0011] Each application server 120 comprises one or more processors 122, input output resources 124, and a certain amount of memory 126, typically implemented as random access memory (RAM), on which one or more applications 128 and one or more operating systems 130 reside. In an embodiment in which the application server 120 is a diskless server, the application server 120 does not include magnetic media such as a hard drive. Instead, large scale storage needs are provided by the storage pool 160.

[0012] Each RAID controller 140 comprises logic to implement at least one RAID implementation 142. The specific RAID implementation is not critical to the subject matter described herein. RAID controller 140 further comprises input/output resources 144 and a memory module 146, typically implemented as random access memory (RAM), on which a boot module application 148 executes. In one embodiment, input/output resources 144 comprise at least one an out of band management port which provides an out-of-band link to a management server. The particular implementation of the out-of-band link is not critical. For example, an Ethernet link or an RS232 link may be used to implement the out-of-band link.

[0013] Storage controllers 150, 152 may be implemented as conventional disk controllers, network attached storage (NAS) controllers, or storage area network (SAN) controllers, depending upon the complexity and configuration of the system 110. Similarly, the storage pool 160 may be implemented as a JBOD (just a bunch of disks) storage configuration or in a more complex

configuration depending upon the complexity and configuration of the system 110.

[0014] Switching fabrics 134, 136 may be implemented in accordance with one or more of several protocols. For example, switching fabrics 134, 136 may operate according to Ethernet protocols, Fibre Channel protocols, InfiniBand Protocols, SAS (Serial-Attached-SCSI (small computer serial interface)) protocols, or the like.

[0015] In operation, applications 128 executing on an application server 120 generate input/output requests to storage provided by storage 160. The input/output requests are transmitted by switching fabric 134 to a RAID controller 140, which performs a level of RAID processing to the input/output requests. RAID controller 140 transmits the requests via the switching fabric 136 to at least one of the storage controllers 150, 152, which in turn accesses data in the storage pool 160.

[0016] As described above, when application server 120 is implemented as a diskless server issues are created with respect to booting the first server in the system 110. In conventional operation, the first server is booted manually by a system administrator or other network technician, typically by coupling a boot media to the application server 120. This manual process, while effective, is slow and expensive to implement.

[0017] In one embodiment, the system 110 depicted in Fig. 1 resolves this problem by enabling the boot module 148 in the RAID controller 142 to obtain a boot image for at least one application server 120 from a remote

management server via the out-of-band communication link between the I/O resources 144 of raid controller 140, and to store the boot image in a logical volume in the storage pool 160. Once the boot image is stored in a logical volume, a server 120 can boot from the boot image.

[0018] Fig. 2 is a schematic illustration of a computing system 200 in which the remote management system may be implemented, according to some embodiments. In the illustrated embodiment, system 200 may be embodied as a server, a desktop PC, a notebook computer, personal digital assistant, or any other processing devices.

[0019] The computing system 200 includes a computer 208 and one or more accompanying input/output devices 206 including a display 202 having a screen 204, a keyboard 210, other I/O device(s) 212, and a mouse 214. The other device(s) 212 can include a touch screen, a voice-activated input device, a track ball, and any other device that allows the system 200 to receive input from a developer and/or a user. The computer 208 includes system hardware 220 including a processing unit 226, a basic input/output system (BIOS) 222, and random access memory and/or read-only memory 230. A file store 280 is communicatively connected to computer 208. File store 280 may be internal such as, e.g., one or more hard drives, or external such as, e.g., one or more external hard drives, network attached storage, or a separate storage network.

[0020] Memory 230 includes an operating system 240 for managing operations of computer 208. In one embodiment, operating system 240 includes a hardware interface module 254 that provides an interface to system

hardware 220. In addition, operating system 240 includes a kernel 244, one or more file systems 246 that manage files used in the operation of computer 208 and a process control subsystem 248 that manages processes executing on computer 208. Operating system 240 further includes one or more device drivers 250 and a system call interface module 242 that provides an interface between the operating system 240 and one or more application modules 262 and/or libraries 264. The various device drivers 250 interface with and generally control the hardware installed in the computing system 200.

[0021] In operation, a boot management module 262 and at least one boot image repository 264 resides in memory 230 of computer 208. In operation, applications executing on computer 200 make calls to the system call interface module 242 to execute one or more commands on the computer's processor. The system call interface module 242 invokes the services of the file systems 246 to manage the files required by the command(s) and the process control subsystem 248 to manage the process required by the command(s). The file system(s) 246 and the process control subsystem 248, in turn, invoke the services of the hardware interface module 254 to interface with the system hardware 220. The operating system kernel 244 can be generally considered as one or more software modules that are responsible for performing many operating system functions.

[0022] The particular embodiment of operating system 240 is not critical to the subject matter described herein. Operating system 240 may be

embodied as a UNIX operating system or any derivative thereof (e.g., Linux, Solaris, *etc.*) or as a Windows® brand operating system.

[0023] In one embodiment, boot management module 262 cooperates with the boot module 148 of RAID controller 140 to implement operations to enable deployment of boot images to the server. Fig. 3 is a flowchart illustrating operations associated with deployment of boot images in diskless servers, according to some embodiments. In some embodiments, the operations depicted in Fig. 3 may be implemented by the boot module 148 in the RAID controller 142 in cooperation with the boot management module 262 of the server 200.

[0024] Referring to Fig. 3, at operation 310 the boot management module 262 initiates a command to create a logical volume in the storage pool 160 for a boot image. At operation 315 the boot management module 262 transmits the command to a RAID controller 140 in the diskless server system 110 via the out-of-band communication link.

[0025] At operation 320 the RAID controller 140 receives the command to create a logical volume for the boot image, and at operation 325 the RAID controller creates a logical volume for the boot image. In some embodiments, the RAID controller may transmit an acknowledgment (ACK) via the out-of-band communication link to signal to the boot management module 262 that the logical volume has been created.

[0026] At operation 330 the boot management module 262 retrieves a boot image from the boot image repository 264. In some embodiments, the

boot management module 262 may compress the boot image (operation 335) before the boot image is transmitted (operation 340) to the RAID controller via the out-of-band communication link.

[0027] At operation 345 the RAID controller receives the boot image from the remote server. If the boot image was compressed (operation 335), then the RAID controller decompresses (operation 350) the boot image. At operation 355 the boot image is stored in the logical volume created for the boot image in operation 325.

[0028] At operation 360 the RAID controller 140 assigns ownership of the logical volume to a particular server 120 in the system 110. At operation 365 the boot management module 262 initiates a boot operation for a server 120. The server 120 would boot in a normal fashion using the boot image in the logical volume created in operation 325.

[0029] In some embodiments, the RAID controller 140 may generate one or more snapshots of the boot image (operation 370) which can be stored in additional logical volumes 166, 168, in the storage pool 160 (operation 375).

[0030] Thus, the operations depicted in Fig. 3 enables a diskless blade server system such as the system 110 depicted in Fig. 1 to deploy boot images from a remote source such as a remote boot image server. Because the boot image is retrieved by the RAID controller 140 via a management interface the boot image can be retrieved and stored before the server 120 is powered up, or even before the server 120 is present in the system 110. In addition, in some embodiments the management interface operates according to a high speed

communication protocol such as, for example, an Ethernet protocol. Such high-speed protocols permit the RAID controller 140 to retrieve large boot images quickly.

[0031] The terms "logic instructions" as referred to herein relates to expressions which may be understood by one or more machines for performing one or more logical operations. For example, logic instructions may comprise instructions which are interpretable by a processor compiler for executing one or more operations on one or more data objects. However, this is merely an example of machine-readable instructions and embodiments are not limited in this respect.

[0032] The terms "computer readable medium" as referred to herein relates to media capable of maintaining expressions which are perceivable by one or more machines. For example, a computer readable medium may comprise one or more storage devices for storing computer readable instructions or data. Such storage devices may comprise storage media such as, for example, optical, magnetic or semiconductor storage media. However, this is merely an example of a computer readable medium and embodiments are not limited in this respect.

[0033] The term "logic" as referred to herein relates to structure for performing one or more logical operations. For example, logic may comprise circuitry which provides one or more output signals based upon one or more input signals. Such circuitry may comprise a finite state machine which receives a digital input and provides a digital output, or circuitry which provides one or

more analog output signals in response to one or more analog input signals. Such circuitry may be provided in an application specific integrated circuit (ASIC) or field programmable gate array (FPGA). Also, logic may comprise machine-readable instructions stored in a memory in combination with processing circuitry to execute such machine-readable instructions. However, these are merely examples of structures which may provide logic and embodiments are not limited in this respect.

[0034] Some of the methods described herein may be embodied as logic instructions on a computer-readable medium. When executed on a processor, the logic instructions cause a processor to be programmed as a special-purpose machine that implements the described methods. The processor, when configured by the logic instructions to execute the methods described herein, constitutes structure for performing the described methods. Alternatively, the methods described herein may be reduced to logic on, e.g., a field programmable gate array (FPGA), an application specific integrated circuit (ASIC) or the like.

[0035] In the description and claims, the terms coupled and connected, along with their derivatives, may be used. In particular embodiments, connected may be used to indicate that two or more elements are in direct physical or electrical contact with each other. Coupled may mean that two or more elements are in direct physical or electrical contact. However, coupled may also mean that two or more elements may not be in direct contact with each other, but yet may still cooperate or interact with each other.

[0036] Reference in the specification to “one embodiment” or “some embodiments” means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least an implementation. The appearances of the phrase “in one embodiment” in various places in the specification may or may not be all referring to the same embodiment.

[0037] Although embodiments have been described in language specific to structural features and/or methodological acts, it is to be understood that claimed subject matter may not be limited to the specific features or acts described. Rather, the specific features and acts are disclosed as sample forms of implementing the claimed subject matter.

CLAIMS

What is claimed is:

1. A method to deploy a boot image in a diskless server in a computer system, comprising:
 - initiating, in a remote management server coupled to a RAID controller in the computer system by an out-of-band communication link, a command to instruct the RAID controller to create at least a first logical volume in a memory module coupled to the RAID controller;
 - transmitting the command to the RAID controller via the out-of-band communication link;
 - in response to the command, creating, in the RAID controller, the first logical volume for the boot image;
 - transmitting a boot image from the remote management server to the RAID controller via the out-of-band communication link;
 - storing the boot image in the first logical volume;
 - assigning ownership of the first logical volume to a first diskless server in the computer system; and
 - initiating a boot operation for the first server, wherein the first server boots from the boot image in the first logical volume.
2. The method of claim 1, wherein transmitting a boot image from the remote management server to the RAID controller to a remote server comprises transmitting the boot image via the out-of-band communication link.
3. The method of claim 1, wherein transmitting a boot image from the remote management server to the RAID controller to a remote server comprises:
 - retrieving the boot image from a repository of approved boot images;
 - and;
 - compressing the boot image.
4. The method of claim 3, further comprising:
 - receiving the boot image at the RAID controller; and
 - decompressing the boot image.

5. The method of claim 4, further comprising:
 passing a pointer to the logical volume from the RAID controller to the diskless server.
6. The method of claim 1, further comprising:
 generating a snapshot of the logical volume; and
 storing the snapshot in a memory module.

7. A system, comprising:
 - a first computer system comprising:
 - at least a first diskless server;
 - at least a first RAID controller coupled to the first diskless server;
 - at least a first storage pool coupled to the RAID controller;
 - a remote management server coupled to the RAID controller via an out-of-band communication link and comprising a boot management module which, when executed:
 - initiates a command to instruct the RAID controller to create at least a first logical volume in a memory module coupled to the RAID controller;
 - transmits the command to the RAID controller via the out-of-band communication link; and
 - transmits a boot image from the remote management server to the RAID controller via the out-of-band communication link; and
 - wherein the RAID controller
 - creates the first logical volume for the boot image in response to the command; and
 - stores the boot image in the first logical volume.
8. The system of claim 7, wherein the boot management module:
 - assigns ownership of the first logical volume to a first diskless server in the computer system; and
 - initiates a boot operation for the first server, wherein the first server boots from the boot image in the first logical volume.
9. The system of claim 7, wherein the boot management module transmits a boot image from the remote management server to the RAID controller to a remote server comprises transmitting the boot image via the out-of-band communication link.
10. The system of claim 7, wherein the boot management module:
 - retrieves the boot image from a repository of approved boot images; and;
 - compresses the boot image.

11. The system of claim 10, wherein the RAID controller:
 - receives the boot image at the RAID controller; and
 - decompresses the boot image.
12. The system of claim 11, wherein the RAID controller:
 - passes a pointer to the logical volume from the RAID controller to the diskless server.
13. The system of claim 7, wherein the RAID controller:
 - generates a snapshot of the logical volume; and
 - stores the snapshot in a memory module.

14. A computer system to manage boot operations of a first computer system comprising at least a first diskless server, at least a first RAID controller coupled to the first diskless server, and at least a first storage pool coupled to the RAID controller, the computer system comprising:

at least one processor;

a memory module coupled to the at least one processor and comprising logic instructions which, when executed by the processor, configure the processor to:

initiate a command to instruct the RAID controller to create at least a first logical volume in a memory module coupled to the RAID controller;

cause a command to be transmitted to the RAID controller via an out-of-band communication link; and

cause a boot image to be transmitted from the remote management server to the RAID controller via the out-of-band communication link.

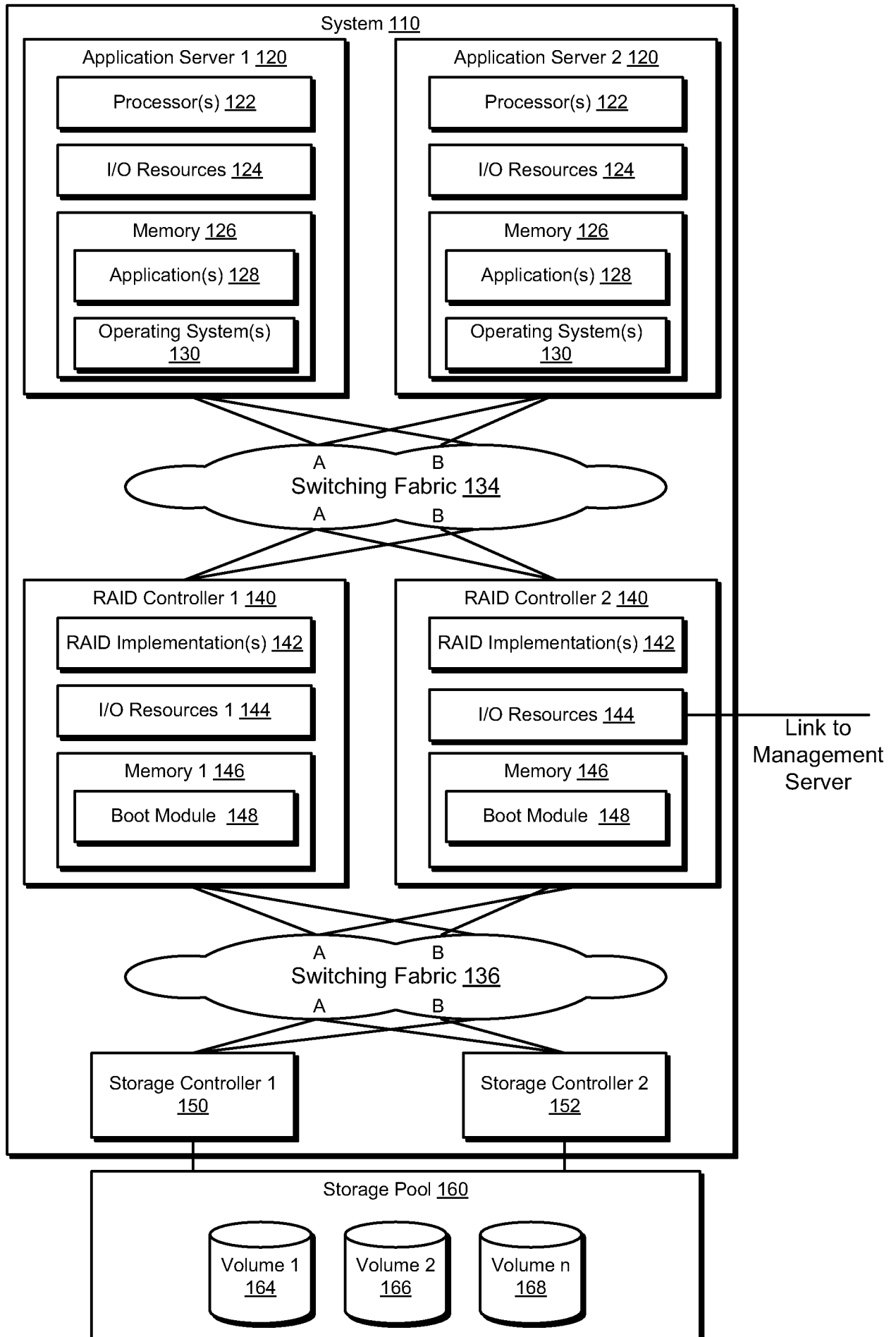
15. The computer system of claim 14, further comprising logic instructions which, when executed by the processor, configure the processor to:

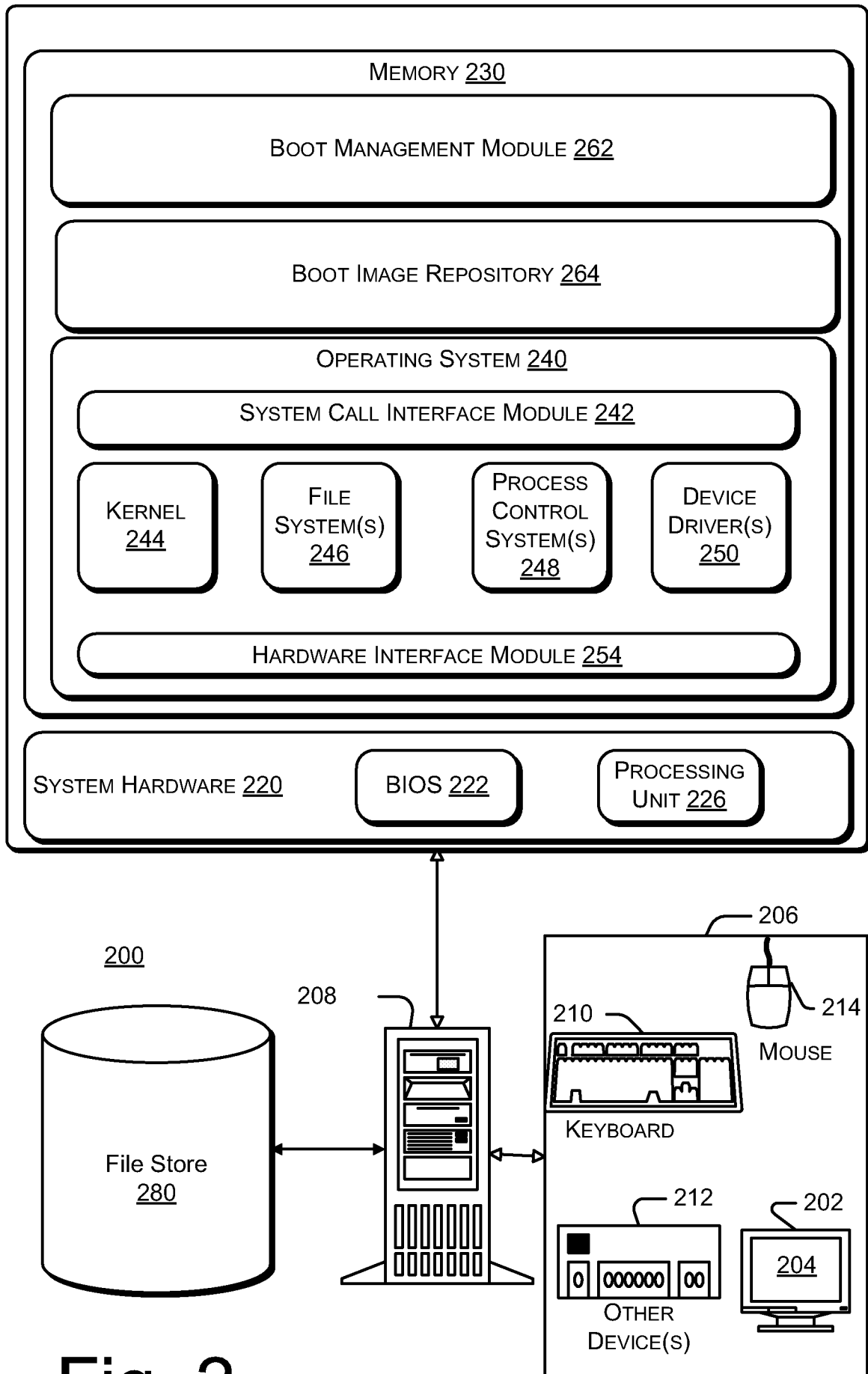
assign ownership of the first logical volume to a first diskless server in the first computer system; and

initiate a boot operation for the first diskless server, wherein the first diskless server boots from the boot image in the first logical volume.

16. The computer system of claim 14, further comprising logic instructions which, when executed by the processor, configure the processor to:

retrieve the boot image from a repository of approved boot images; and;
compress the boot image.



**Fig. 2**

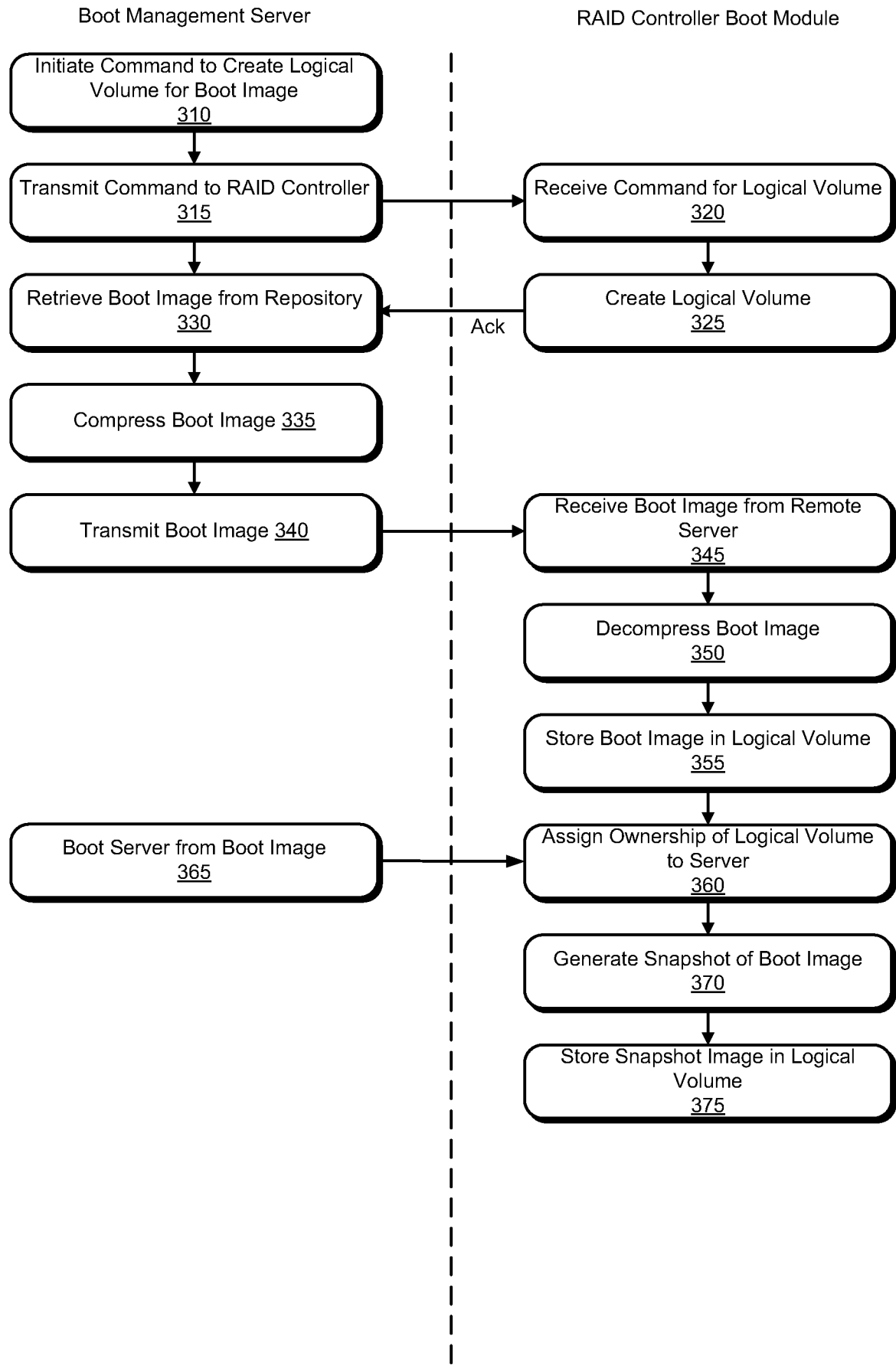


Fig. 3

A. CLASSIFICATION OF SUBJECT MATTER**G06F 9/24(2006.01)i, G06F 15/16(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)

IPC 8 G06F

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 since 1975

Japanese Utility models and applications for Utility models since 1975

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

eKIPASS (KIPO internal) & keyword : "remote boot"

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2006-0020848 A1 (SANJOY MAITY et al.) 26 January 2006 See the abstract, figure 4 and its corresponding specification.	1-15
A	US 2007-0268917 A1 (STEVEN WONG et al.) 22 November 2007 See the abstract and figures 3, 4.	1-15
A	US 2005-0138346 A1 (DAVID M. CAUTHRON) 23 June 2005 See the abstract and figures 1, 2.	1-15
A	US 6810478 B1 (VAIJAYANTHIMALA K. ANAND et al.) 26 October 2004 See the abstract and figure 5.	1-15
A	US 2006-0117172 A1 (YAOXUE ZHANG et al.) 01 June 2006 See figure 4 and its corresponding specification.	1-15
A	US 2003-0208675 A1 (GINTAUTAS BUROKAS et al.) 06 November 2003 See the abstract.	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 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

08 OCTOBER 2008 (08.10.2008)

Date of mailing of the international search report

08 OCTOBER 2008 (08.10.2008)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

KIM, Jung Hun

Telephone No. 82-42-481-5949



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2008/052235

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2006-0020848 A1	26.01.2006	US 6973587 B1 US 7310744 B2	06.12.2005 18.12.2007
US 2007-0268917 A1	22.11.2007	None	
US 2005-0138346 A1	23.06.2005	CA 2578017 A1 WO 2006-025840 A2	09.03.2006 09.03.2006
US 6810478 B1	26.10.2004	None	
US 2006-0117172 A1	01.06.2006	None	
US 2003-0208675 A1	06.11.2003	None	