

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
23 December 2004 (23.12.2004)

PCT

(10) International Publication Number
WO 2004/111849 A2

(51) International Patent Classification⁷: **G06F 11/00**

BARTFAI, Robert, Francis [US/US]; 6340 North Placita
Derrumbadera, Tucson, AZ 85750 (US).

(21) International Application Number:

PCT/EP2004/051115

(74) **Agent: LING, Christopher, John**; IBM United Kingdom
Limited, Intellectual Property Law, Hursley Park, Winches-
ter Hampshire SO21 2JN (GB).

(22) International Filing Date: 15 June 2004 (15.06.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/464,024 17 June 2003 (17.06.2003) US

(71) **Applicant (for all designated States except US): INTER-
NATIONAL BUSINESS MACHINES CORPORA-
TION** [US/US]; New Orchard Road, Armonk, NY 10504
(US).

(71) **Applicant (for MG only): IBM UNITED KINGDOM
LIMITED** [GB/GB]; PO Box 41, North Harbour,
Portsmouth Hampshire PO6 3AU (GB).

(72) **Inventors; and**

(75) **Inventors/Applicants (for US only): STANLEY, Warren**
[US/US]; 1461 North Purple Cactus Place, Tucson, AZ
85715 (US). **MICKA, William, Frank** [US/US]; 3921
E. LaEspalda, Tucson, AZ 85718 (US). **SPEAR, Gail,
Andrea** [US/US]; 6822 E. Fieldstone Lane, Tucson, AZ
85750 (US). **WARNER, Sam, Clark** [US/US]; 4284 West
Calva Draw Place, Tucson, AZ 85745 (US). **GLUCK,
Olympia** [IL/IL]; Rh.Hatishbi 51A/AP2, Haifa (IL).
FACTOR, Michael [IL/IL]; Rachel 32, 34402 Haifa (IL).

(81) **Designated States (unless otherwise indicated, for every
kind of national protection available):** AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, 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, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,
SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ,
GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) **Title:** METHOD, SYSTEM AND ARTICLE OF MANUFACTURE FOR REMOTE COPYING OF DATA

(57) **Abstract:** Provided are a method, system, and article of manufacture for copying storage. Data sent from a first storage unit is synchronously copied at a second storage unit. The copied data is sent asynchronously from the second storage unit to a third storage unit.



WO 2004/111849 A2

Description

METHOD, SYSTEM, AND ARTICLE OF MANUFACTURE FOR REMOTE COPYING OF DATA

Technical Field

- [001] The present invention relates to a method, system, and an article of manufacture for remote copying of data.

Background Art

- [002] Information technology systems, including storage systems, may need protection from site disasters or outages, where outages may be planned or unplanned. Furthermore, information technology systems may require features for data migration, data backup, or data duplication. Implementations for disaster or outage recovery, data migration, data backup, and data duplication may include mirroring or copying of data in storage systems. Such mirroring or copying of data may involve interactions among hosts, storage systems and connecting networking components of the information technology system.
- [003] An enterprise storage server (ESS), such as the IBM® TotalStorage Enterprise Storage Server®, may be a disk storage server that includes one or more processors coupled to storage devices, including high capacity scalable storage devices, Redundant Array of Independent Disks (RAID), etc. The enterprise storage servers are connected to a network and include features for copying data in storage systems.
- [004] PeertoPeer Remote Copy (PPRC) is an ESS function that allows the shadowing of application system data from a first site to a second site. The first site may be referred to as an application site, a local site, or a primary site. The second site may be referred to as a recovery site, a remote site, or a secondary site. The logical volumes that hold the data in the ESS at the local site are called local volumes, and the corresponding logical volumes that hold the mirrored data at the remote site are called remote volumes. High speed links, such as ESCON links may connect the local and remote ESS systems.
- [005] In the synchronous type of operation for PPRC, i.e., synchronous PPRC, the updates done by a host application to the local volumes at the local site are synchronously shadowed onto the remote volumes at the remote site. As synchronous PPRC is a synchronous copying solution, write updates are ensured on both copies (local and remote) before the write is considered to be completed for the host application. In synchronous PPRC the host application does not get the Awrite complete@ condition until the update is synchronously done in both the local and the remote volumes. Therefore, from the perspective of the host application the data at the

remote volumes at the remote site is equivalent to the data at the local volumes at the local site.

[006] Synchronous PPRC increases the response time as compared to an asynchronous copy operation, and this is inherent to the synchronous operation. The overhead comes from the additional steps that are executed before the write operation is signaled as completed to the host application. Also the PPRC activity between the local site and the remote site may be comprised of signals that travel through the links that connect the sites, and the overhead on the response time of the host application write operations will increase proportionally with the distance between the sites. Therefore, the distance affects a host application's write response time. In certain implementations, there may be a maximum supported distance for synchronous PPRC operations referred to as the synchronous communication distance.

[007] In the Extended Distance PPRC (also referred to as PPRC Extended Distance) method of operation, PPRC mirrors the updates of the local volume onto the remote volumes in an asynchronous manner, while the host application is running. In Extended

[008] Distance PPRC, the host application receives a write complete response before the update is copied from the local volumes to the remote volumes. In this way, when in Extended Distance PPRC, a host application's write operations are free of the typical synchronous overheads. Therefore, Extended Distance PPRC is suitable for remote copy solutions at very long distances with minimal impact on host applications. There is no overhead penalty upon the host application's write such as in synchronous PPRC. However, Extended Distance PPRC does not continuously maintain an equivalent copy of the local data at the remote site.

[009] Further details of the PPRC are described in the IBM publication AIBM TotalStorage Enterprise Storage Server: PPRC Extended Distance, @ IBM document no. SG24656800 (Copyright IBM, 2002), which publication is incorporated herein by reference in its entirety.

Disclosure of Invention

[010] Provided are a method, system, and article of manufacture for copying storage. Data sent from a first storage unit is synchronously copied at a second storage unit. The copied data is sent asynchronously from the second storage unit to a third storage unit.

[011] Preferably the first storage unit is a first storage volume in a first storage control unit, wherein the second storage unit is a second storage volume in a second storage control unit, wherein the third storage unit is a third storage volume in a third storage control unit, and wherein the third storage control unit is beyond a synchronous communication distance to the first storage control unit.

- [012] Alternatively the first storage unit is a storage control unit, wherein the second storage unit is a second storage control unit, wherein the third storage unit is a third storage control unit, wherein the second storage unit is within a synchronous communication distance to the first storage unit, and wherein the third storage unit is beyond the synchronous communication distance to the second storage unit.
- [013] Preferably the copied data is received at the third storage unit. The data received from the second storage unit, is asynchronously copied at the third storage unit.
- [014] Optionally a write request is received at the first storage unit from a host application coupled to a host system. Data corresponding to the write request is written at a cache and a non-volatile storage coupled to the first storage unit. The data corresponding to the write request is sent synchronously to the second storage unit. Optionally a response, indicative of the write request being completed, from the second storage unit is received at the first storage unit. An indication is sent to the host application that the write request is completed.
- [015] Optionally, in response to receiving the data at the second storage unit, the data is copied to a cache and a non-volatile storage at the second storage unit. The data is marked as modified at the second storage unit. A response is sent from the second storage unit to the first storage unit, wherein the response indicates that the received data has been copied at the second storage unit.
- [016] Optionally a determination is made of the modified data at the second storage unit, wherein sending the copied data, further comprises sending the modified data asynchronously from the second storage unit to the third storage unit.
- [017] Optionally the first storage unit is coupled to a host that sends Input/Output requests to the first storage unit, wherein an update from the host at the first storage unit is synchronously reflected at the second storage unit and asynchronously reflected at the third storage unit.
- [018] Preferably a host application that performs I/O with the first storage unit completes a write operation faster in comparison to the time taken if an update corresponding to the write operation were copied synchronously to the third storage unit.
- [019] Optionally recovering from a disaster at the first storage unit is performed by substituting replicated data from the second or third storage units.
- [020] Optionally additional storage units are cascaded to the first, second, and third storage units, wherein the additional storage units may communicate synchronously or asynchronously. In yet further implementations, functions of the first and second storage units are integrated into a single storage control unit.
- [021] Accordingly, an embodiment of the invention may be used to create a long distance disaster recovery solution by copying synchronously from a local storage control unit to an intermediate storage control unit, and in parallel copying asynchronously from

the intermediate storage control unit to a remote storage control unit.

Brief Description of the Drawings

- [022] The invention will now be described, by way of example only, with reference to a preferred embodiment thereof, as illustrated in the accompanying drawings, in which:
- [023] FIG. 1 illustrates a block diagram of a computing environment, in accordance with certain described aspects of the invention;
- [024] FIG. 2 illustrates a block diagram of a cascading copy application, in accordance with certain described implementations of the invention;
- [025] FIG. 3 illustrates logic implemented in a local storage control unit, in accordance with certain described implementations of the invention;
- [026] FIG. 4 illustrates logic for receiving data synchronously as implemented in an intermediate storage control unit, in accordance with certain described implementations of the invention;
- [027] FIG. 5 illustrates logic for copying data asynchronously as implemented in the intermediate storage control unit, in accordance with certain described implementations of the invention; and
- [028] FIG. 6 illustrates a block diagram of a computer architecture in which certain described aspects of the invention are implemented.
- [029] Note that in the drawings like reference numbers represent corresponding parts throughout.

Best Mode for Carrying Out the Invention

- [030] FIG. 1 illustrates a computing environment utilizing three storage control units, such as a local storage control unit 100, an intermediate storage control unit 102, and a remote storage control unit 104 connected by data interface channels 106, 108, such as, the Enterprise System Connection (ESCON)* channel or any other data interface mechanism known in the art (e.g., fibre channel, Storage Area Network (SAN) interconnections, etc.).
- [031] The three storage control units 100, 102, 104 may be at three different sites with the local storage control unit 100 and the intermediate storage control unit 102 being within a synchronous communication distance of each other. The synchronous communication distance between two storage control units is the distance up to which synchronous communication is feasible between the two storage control units. The remote storage control unit 104 may be a long distance away from the intermediate storage control unit 102 and the local storage control unit 100, such that, synchronous copying of data from the intermediate storage control unit 102 to the remote storage control unit 104 may be time consuming or impractical. Additionally, the intermediate storage control unit 102 may be in a secure environment separated from the local

storage control unit 100 and with separate power to reduce the possibility of an outage affecting both the local storage control unit 100 and the intermediate storage control unit 102. Certain implementations of the invention create a three site (local, intermediate, remote) disaster recovery solution where there may be no data loss if the local storage control unit 100 is lost. In the three site disaster recovery solution, the local storage control unit 100 is kept at the local site, the intermediate storage control unit 102 is kept at the intermediate site, and the remote storage control unit 104 is kept at the remote site. Data copied on the intermediate storage control unit 102 or the remote storage control unit 104 may be used to recover from the loss of the local storage control unit 100. In certain alternative implementations, there may be less than three sites. For example, the local storage control unit 100 and the intermediate storage control unit 102 may be at the same site. In additional alternative implementations of the invention, there may be more than three storage control units distributed among three or more sites. Furthermore, functions of a plurality of storage control units may be integrated into a single storage control unit, e.g., functions of the local storage control unit 100 and the intermediate storage control unit 102 may be integrated into a single storage control unit.

[032] The local storage control unit 100 is coupled to a host 110 via data interface channel 112. While only a single host 110 is shown coupled to the local storage control unit 100, in certain implementations of the invention, a plurality of hosts may be coupled to the local storage control unit 100. The host 110 may be any computational device known in the art, such as a personal computer, a workstation, a server, a mainframe, a hand held computer, a palm top computer, a telephony device, network appliance, etc. The host 110 may include any operating system (not shown) known in the art, such as the IBM OS/390* operating system. The host 110 may include at least one host application 114 that sends Input/Output (I/O) requests to the local storage control unit 100.

[033] The storage control units 100, 102, and 104 are coupled to storage volumes, such as, local site storage volumes 116, intermediate site storage volumes 118, and remote site storage volumes 120, respectively. The storage volumes 116, 118, 120 may be configured as a Direct Access Storage Device (DASD), one or more RAID ranks, Just a bunch of disks (JBOD), or any other data repository system known in the art.

[034] The storage control units 100, 102, and 104 may each include a cache, such as, cache 122, 124, 126 respectively. The caches 122, 124, 126 comprise volatile memory to store tracks. The storage control units 100, 102, and 104 may each include a non-volatile storage (NVS), such as non-volatile storage 128, 130, 132 respectively. The non-volatile storage 128, 130, 132 elements may buffer certain modified tracks in the caches 122, 124, 126 respectively.

- [035] The local storage control unit 100 additionally includes an application, such as, a local application 134, for synchronous copying of data stored in the cache 122, non-volatile storage 128, and local site storage volumes 116 to another storage control unit, such as, the intermediate storage control unit 102. The local application 134 includes copy services functions that execute in the local storage control unit 100. The local storage control unit 100 receives I/O requests from the host application 114 to read and write to the local site storage volumes 116.
- [036] The intermediate storage control unit 102 additionally includes an application, such as a cascading PPRC application 136. The cascading PPRC application 136 includes copy services functions that execute in the intermediate storage control unit 102. The cascading PPRC application 136 can interact with the local storage control unit 100 to receive data synchronously. The cascading PPRC application 136 can also send data asynchronously to the remote storage control unit 104. Therefore, the cascading PPRC application 136 cascades a first pair of storage control units formed by the local storage control unit 100 and the intermediate storage control unit 102, and a second pair of storage control units formed by the intermediate storage control unit 102 and the remote storage control unit 104. In alternative implementations of the invention, additional storage control units may be cascaded.
- [037] The remote storage control unit 104 additionally includes an application, such as a remote application 138, that can receive data asynchronously from another storage control unit, such as, the intermediate storage control unit 102. The remote application 138 includes copy services functions that execute in the remote storage control unit 104.
- [038] Therefore, FIG. 1, illustrates a computing environment where a host application 114 sends I/O requests to a local storage control unit 100. The local storage control unit 100 synchronously copies data to the intermediate storage control unit 102, and the intermediate storage control unit 104 asynchronously copies data to the remote storage control unit 104.
- [039] FIG. 2 illustrates a block diagram that illustrates communications between the local application 134, the cascading PPRC application 136 and the remote application 138, in accordance with certain implementations of the invention.
- [040] The local application 134 performs a synchronous data transfer, such as, via synchronous PPRC 200, to a synchronous copy process 202 that may be generated by the cascading PPRC application 136. The synchronous data transfer 200 takes place over the data interface channel 106.
- [041] A background asynchronous copy process 204 that may be generated by the cascading PPRC application 136 performs an asynchronous data transfer, such as, via Extended Distance PPRC 206, to the remote application 138. The asynchronous data

transfer takes place over the data interface channel 108.

- [042] Since data from the local storage control unit 100 are copied synchronously to the intermediate storage control unit 102, the intermediate site storage volumes 118 may include a copy of the local site storage volumes 116. In certain implementations of the invention the distance between the local storage control unit 100 and the intermediate storage control unit is kept as close as possible to minimize the performance impact of synchronous PPRC. Data is copied asynchronously from the intermediate storage control unit 102 to the remote storage control unit 104. As a result, the effect of long distance on the host response time is eliminated.
- [043] Therefore, FIG. 2 illustrates how the cascading PPRC application 136 on the intermediate storage control unit 102 receives data synchronously from the local storage control unit 100, and transmits data asynchronously to the remote storage control unit 104.
- [044] FIG. 3 illustrates logic implemented in the local storage control unit 100, in accordance with certain implementations of the invention. In certain implementations of the invention the logic of FIG. 3 may be implemented in the local application 134 resident in the local storage control unit 100.
- [045] Control starts at block 300, where the local application 134 receives a write request from the host application 114. The local application 134 writes (at block 302) data corresponding to the write request on the cache 122 and the non-volatile storage 128 on the local storage control unit 100. Additional applications (not shown), such as, caching applications and non-volatile storage applications, in the local storage control unit 100 may manage the data in the cache 122 and the data in the non-volatile storage 128 and keep the data in the cache 122 and the non-volatile storage 128 consistent with the data in the local site storage volumes 116.
- [046] The local application 134 determines (at block 304) if the local storage control unit 100 is a primary PPRC device, i.e., the local storage control unit includes source data for a PPRC transaction. If so, the local application 134 sends (at block 306) the written data to the intermediate storage control unit 102 via a new write request. The local application 134 waits (at block 308) for a write complete acknowledgment from the intermediate storage control unit 102. The local application 134 receives (at block 310) a write complete acknowledgment from the intermediate storage control unit 102. Therefore, the local application 134 has transferred the data written by the host application 114 on the local storage control unit 100 to the intermediate storage control unit 102 via a synchronous copy.
- [047] The local application 134 signals (at block 312) to the host application 114 that the write request from the host application 114 has been completed at the local storage control unit 100. The local application 134 receives (at block 300) a next write request

from the host application 114.

[048] If the local application 134 determines (at block 304) that the local storage control unit 100 is not a primary PPRC device, i.e., the local storage control unit is not a source device for a PPRC transaction, then the local application 134 does not have to send any data to the intermediate storage control unit 102, and the local application 134 signals (at block 312) to the host application 114 that the write request from the host application 114 has been completed at the local storage control unit 100.

[049] Therefore, FIG. 3 illustrates a logic for receiving a write request from the host application 114 to the local storage control unit 100 and synchronously copying the data corresponding to the write request from the local storage control unit 100 to the intermediate storage control unit 102. The host application 114 waits for the write request to be completed while the synchronous copying of data takes place. Since the local storage control unit 100 and the intermediate storage control unit 102 are within a synchronous communication distance of each other, the synchronous copying of data from the local storage control unit 100 to the intermediate storage control unit 102 takes a smaller amount of time when compared to the situation where the local storage control unit 100 is beyond a synchronous communication distance to the intermediate storage control unit 102. Since the copy of the data on the intermediate storage control unit 102 is written synchronously, the intermediate storage control unit 102 includes an equivalent copy of the data on the local storage control unit 100.

[050] FIG. 4 illustrates logic for receiving data synchronously as implemented in the intermediate storage control unit 102, in accordance with certain implementations of the invention. The cascading PPRC application 136 may perform the logic illustrated in FIG. 4.

[051] Control starts at block 400 where the cascading PPRC application 136 receives a write request from the local application 134. For example, the write request sent at block 306 of FIG. 3 to the intermediate storage control unit 102 may be received by the cascading PPRC application 136. The cascading PPRC application 136 writes (at block 402) data corresponding to the write request to the cache 124 and the non-volatile storage 130. The intermediate storage control unit 102 may keep the cache 124 and the non-volatile storage 130 consistent with the intermediate site storage volumes 118.

[052] The cascading PPRC application 136 determines (at block 404) if data on the intermediate storage control unit 102 is to be cascaded, i.e., the data is to be sent to the remote storage control unit 104. If so, the synchronous copy process 202 of the cascading PPRC application 136 marks (at block 406) data as PPRC modified. The synchronous copy process 202 of the cascading PPRC application 136 signals (at block 408) a write complete acknowledgment to the local application 134. The cascading PPRC application 136 receives (at block 400) the next write request from the local ap-

plication 134.

- [053] If the cascading PPRC application 136 determines (at block 404) that data on the intermediate storage control unit 102 does not have to be cascaded, then the synchronous copy process 202 of the cascading PPRC application 136 signals (at block 408) a write complete acknowledgment to the local application 134 and the cascading PPRC application 136 receives (at block 400) the next request from the local application 134.
- [054] Therefore, FIG. 4 illustrates how the intermediate storage control unit 102 receives a write request from the local storage control unit 100, where the write request corresponds to a host write request. The intermediate storage control unit 102 marks data corresponding to the host write request as PPRC modified.
- [055] FIG. 5 illustrates logic for copying data asynchronously as implemented in the intermediate storage control unit 102, in accordance with certain implementations of the invention. The logic illustrated in FIG. 5 may be performed by the background asynchronous copy process 204 of the cascading PPRC application 136.
- [056] Control starts at block 500 where the background asynchronous copy process 204 of the cascading PPRC application 136 determines the PPRC modified data stored in the cache 124, non-volatile storage 130, and the intermediate site storage volumes 118 of the intermediate storage control unit 102.
- [057] The background asynchronous copy process 204 of the cascading PPRC application 136 sends (at block 502) the PPRC modified data to the remote storage control unit 104 asynchronously, i.e., the background asynchronous copy process 204 keeps sending the PPRC modified data stored in the cache 124, non-volatile storage 130, and the intermediate site storage volumes 118 of the intermediate storage control unit 102.
- [058] After the PPRC modified data has been sent, the background asynchronous copy process 204 determines (at block 504) if write complete acknowledgment has been received from the remote storage control unit 104. If not, the background asynchronous copy process 204 again determines (at block 504) if the write complete acknowledgment has been received.
- [059] If after all PPRC modified data has been sent, the background asynchronous copy process 204 determines (at block 504) that write complete acknowledgment has been received from the remote storage control unit 104 then the background asynchronous copy process 204 determines (at block 500) the PPRC modified data once again.
- [060] The logic of FIG. 5, illustrates how the background asynchronous copy process 204 while executing in the background copies data asynchronously from the intermediate storage control unit 102 to the remote storage control unit 104. Since the copying is asynchronous, the intermediate storage control unit 102 and the remote storage control

unit 104 may be separated by long distances, such as, the extended distances allowed by Extended Distance PPRC.

[061] In certain implementations of the invention, if the local storage control unit 100 stops sending updates to the intermediate storage control unit 102 because of an outage at the local site that has the local storage control unit 100, then the background asynchronous copy process 204 may quickly complete the copy of all remaining modified data to the remote storage control unit 104. At the completion of the copy, the remote site storage volumes 120 will include an equivalent copy of all updates up to the time of the outage. If there are multiple failures, such that both the local storage control unit 100 and the intermediate storage control unit 102 are lost then there may be data loss at the remote site.

[062] Since the remote storage control unit 104 is updated asynchronously, the data on the remote storage control unit 104 may not be equivalent to the data on the local storage control unit 100, unless all of the data from the intermediate storage control unit 102 has been copied up to some point in time. To maintain an equivalent copy of data at the remote storage control unit 104 in case of failure of both the local storage control unit 100 and the intermediate storage control unit 102, certain implementations of the invention may force the data at the remote storage control unit to contain all dependent updates up to some specified time. The consistent copy at the remote storage control unit may be preserved via a point in time copy, such as FlashCopy*. One method may include quiescing the host I/O temporarily at the local site while the remote storage control unit 104 catches up with the updates. Another method may prevent writes to the intermediate storage control unit 102 while the remote storage control unit 104 catches up with the updates.

[063] The implementations create a long distance disaster recovery solution by first copying synchronously from a local storage control unit to an intermediate storage control unit, and subsequently copying asynchronously from the intermediate storage control unit to a remote storage control unit. The distance between the local storage control unit and the intermediate storage control unit may be small enough such that copying data synchronously does not cause a significant performance impact on applications that perform I/O operations on the local storage control unit.

[064] In implementations of the invention, if either the local storage control unit 100 or data on the local storage control unit 100 is lost, then the data can be recovered from replicated copies of the data on either the intermediate storage control unit 102 or the remote storage control unit 104. In certain implementations, it may be preferable to recover the data from the intermediate storage control unit 102 as the data on the intermediate storage control unit 102 is always equivalent to the data on the local storage control unit 100 since data is copied synchronously from the local storage control unit

100 to the intermediate storage control unit 102.

[065] Additional Implementation Details

[066] The described techniques may be implemented as a method, apparatus or article of manufacture using standard programming and/or engineering techniques to produce software, firmware, hardware, or any combination thereof. The term Article of manufacture as used herein refers to code or logic implemented in hardware logic (e.g., an integrated circuit chip, Programmable Gate Array (PGA), Application Specific Integrated Circuit (ASIC), etc.) or a computer readable medium (e.g., magnetic storage medium, such as hard disk drives, floppy disks, tape), optical storage (e.g., CD-ROMs, optical disks, etc.), volatile and non-volatile memory devices (e.g., EEPROMs, ROMs, PROMs, RAMs, DRAMs, SRAMs, firmware, programmable logic, etc.). Code in the computer readable medium is accessed and executed by a processor. The code in which implementations are made may further be accessible through a transmission media or from a file server over a network. In such cases, the article of manufacture in which the code is implemented may comprise a transmission media, such as a network transmission line, wireless transmission media, signals propagating through space, radio waves, infrared signals, etc. The article of manufacture may comprise any information bearing medium known in the art.

[067] FIG. 6 illustrates a block diagram of a computer architecture in which certain aspects of the invention are implemented. FIG. 6 illustrates one implementation of the host 110, and the storage control units 100, 102, 104. The host 110, and the storage control units 100, 102, 104 may implement a computer architecture 600 having a processor 602, a memory 604 (e.g., a volatile memory device), and storage 606 (e.g., a non-volatile storage, magnetic disk drives, optical disk drives, tape drives, etc.). The storage 606 may comprise an internal storage device, an attached storage device or a network accessible storage device. Programs in the storage 606 may be loaded into the memory 604 and executed by the processor 602 in a manner known in the art. The architecture may further include a network card 608 to enable communication with a network. The architecture may also include at least one input 610, such as a keyboard, a touchscreen, a pen, voice-activated input, etc., and at least one output 612, such as a display device, a speaker, a printer, etc.

[068]

[069] In alternative implementations of the invention the data transfer between the local storage control unit 100 and the intermediate storage control unit 102 may be via Extended Distance PPRC. However, there may be data loss if there is an outage at the local storage control unit 100. Additionally, in alternative implementations of the invention the data transfer between the intermediate storage control unit 102 and the remote storage control unit 104 may be via synchronous PPRC. However, there may

be performance impacts on the I/O from the host 110 to the local storage control unit 100.

[070] In alternative implementations of the invention, the functions of the local storage control unit 100 and the intermediate storage control unit 102 may be implemented in a single storage control unit. Furthermore, in additional implementations of the invention there may be more than three storage control units cascaded to each other. For example, a fourth storage control unit may be coupled to the remote storage control unit 104 and data may be transferred from the remote storage control unit 104 to the fourth storage control unit. In certain implementations of the invention, a chain of synchronous data transfers and a chain of asynchronous data transfers may take place among a plurality of cascaded storage control units. Furthermore, while the implementations have been described with storage control units, the storage control units may be any storage unit known in the art.

[071] The logic of FIGs. 3, 4, and 5 describe specific operations occurring in a particular order. Further, the operations may be performed in parallel as well as sequentially. In alternative implementations, certain of the logic operations may be performed in a different order, modified or removed and still implement implementations of the present invention. Moreover, steps may be added to the above described logic and still conform to the implementations. Yet further steps may be performed by a single process or distributed processes.

[072] Many of the software and hardware components have been described in separate modules for purposes of illustration. Such components may be integrated into a fewer number of components or divided into a larger number of components. Additionally, certain operations described as performed by a specific component may be performed by other components.

[073] *IBM, IBM TotalStorage Enterprise Storage Server, Enterprise System Connection (ESCON), OS/390, FlashCopy are trademarks of International Business Machines Corp.

Claims

- [001] A method for copying storage, comprising: synchronously copying data, sent from a first storage unit, at a second storage unit; and sending the copied data asynchronously from the second storage unit to a third storage unit.
- [002] The method of claim 1, wherein the first storage unit is a first storage volume in a first storage control unit, wherein the second storage unit is a second storage volume in a second storage control unit, wherein the third storage unit is a third storage volume in a third storage control unit, and wherein the third storage control unit is beyond a synchronous communication distance to the first storage control unit.
- [003] The method of claim 1, wherein the first storage unit is a storage control unit, wherein the second storage unit is a second storage control unit, wherein the third storage unit is a third storage control unit, wherein the second storage unit is within a synchronous communication distance to the first storage unit, and wherein the third storage unit is beyond the synchronous communication distance to the second storage unit.
- [004] The method of claim 1, further comprising: receiving a write request at the first storage unit from a host application coupled to a host system; writing data corresponding to the write request at a cache and a non-volatile storage coupled to the first storage unit; and sending the data corresponding to the write request synchronously to the second storage unit.
- [005] The method of claim 1, further comprising: in response to receiving the data at the second storage unit, copying the data to a cache and a non-volatile storage at the second storage unit; and marking the data as modified at the second storage unit; sending a response from the second storage unit to the first storage unit, wherein the response indicates that the received data has been copied at the second storage unit.
- [006] The method of claim 5, further comprising: determining the modified data at the second storage unit; and wherein sending the copied data, further comprises sending the modified data asynchronously from the second storage unit to the third storage unit.
- [007] The method of claim 1, further comprising: recovering from a disaster at the first storage unit by substituting replicated data from the second or third storage units.
- [008] The method of claim 1, wherein additional storage units are cascaded to the first, second, and third storage units, wherein the additional storage units may communicate synchronously or asynchronously.
- [009] A system for copying storage, comprising: a first storage unit; a second storage

unit coupled to the first storage unit; a third storage unit coupled to the second storage unit; means for synchronously copying data, sent from the first storage unit, at the second storage unit; and means for sending the copied data asynchronously from the second storage unit to the third storage unit.

[010] The system of claim 9, wherein the first storage unit is a first storage volume in a first storage control unit, wherein the second storage unit is a second storage volume in a second storage control unit, wherein the third storage unit is a third storage volume in a third storage control unit, and wherein the third storage control unit is beyond a synchronous communication distance to the first storage control unit.

[011] The system of claim 9, wherein the first storage unit is a storage control unit, wherein the second storage unit is a second storage control unit, wherein the third storage unit is a third storage control unit, wherein the second storage unit is within a synchronous communication distance to the first storage unit, and wherein the third storage unit is beyond the synchronous communication distance to the second storage unit.

[012] The system of claim 9, further comprising: a host system coupled to the first storage unit; a host application coupled to the host system; means for receiving a write request at the first storage unit from the host application; means for writing data corresponding to the write request at a cache and a non-volatile storage coupled to the first storage unit; and means for sending the data corresponding to the write request synchronously to the second storage unit.

[013] The system of claim 9, further comprising: means for copying the data to a cache and a non-volatile storage at the second storage unit, in response to receiving the data at the second storage unit; and means for marking the data as modified at the second storage unit; means for sending a response from the second storage unit to the first storage unit, wherein the response indicates that the received data has been copied at the second storage unit.

[014] The system of claim 13, further comprising: means for determining the modified data at the second storage unit; and wherein the means for sending the copied data, further comprises means for sending the modified data asynchronously from the second storage unit to the third storage unit.

[015] The system of claim 9, further comprising: means for recovering from a disaster at the first storage unit by substituting replicated data from the second or third storage units.

[016] The system of claim 9, wherein additional storage units are cascaded to the first, second, and third storage units, wherein the additional storage units may communicate synchronously or asynchronously.

[017] An article of manufacture for copying storage, wherein the article of manufacture includes at least one program which comprises operation capable of performing a method according to any one of claims 1 to 8.

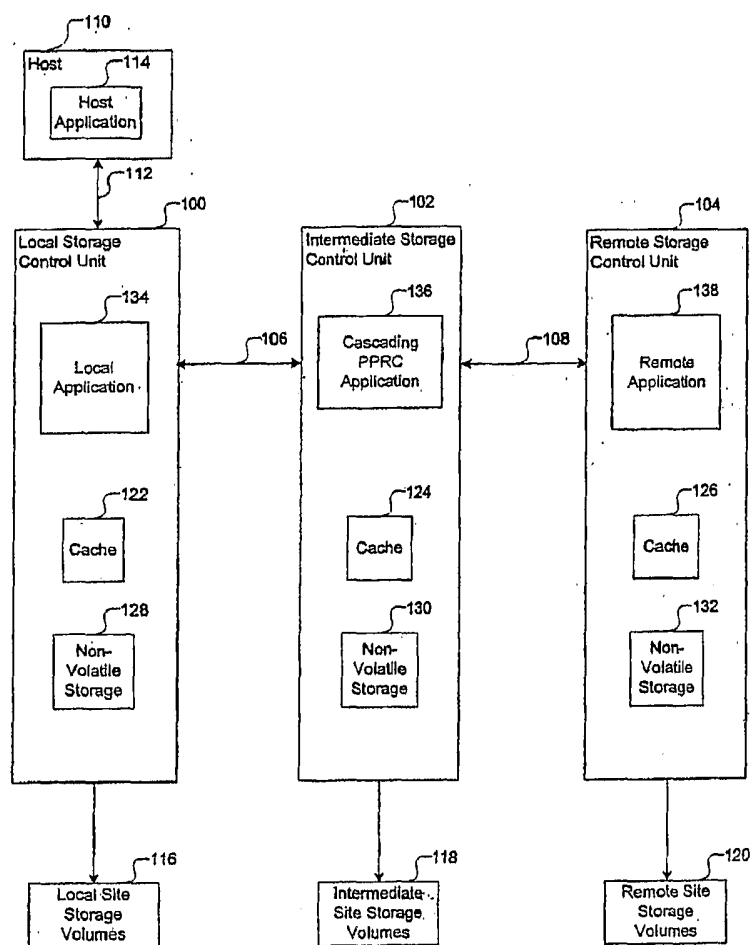
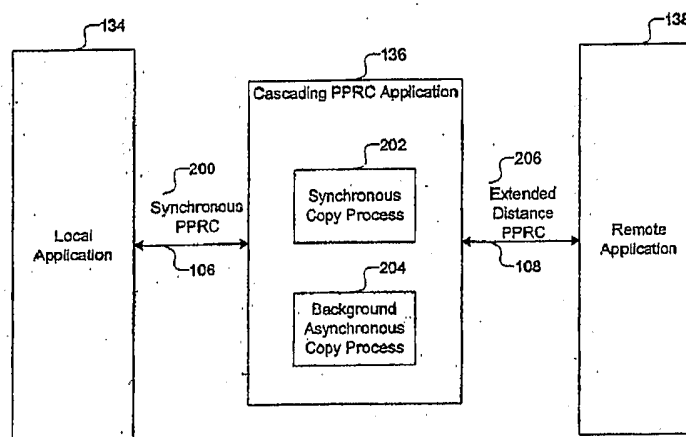


FIG. 1

2/6

**FIG. 2**

3/6

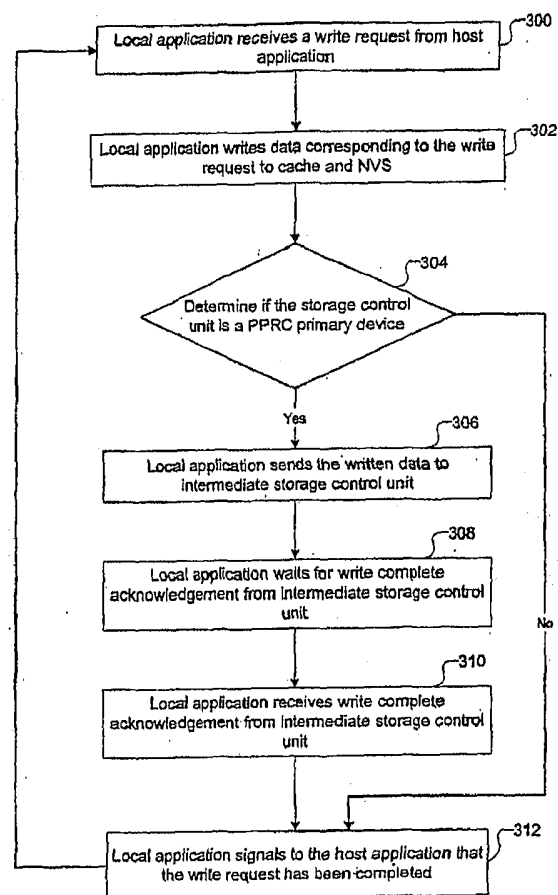
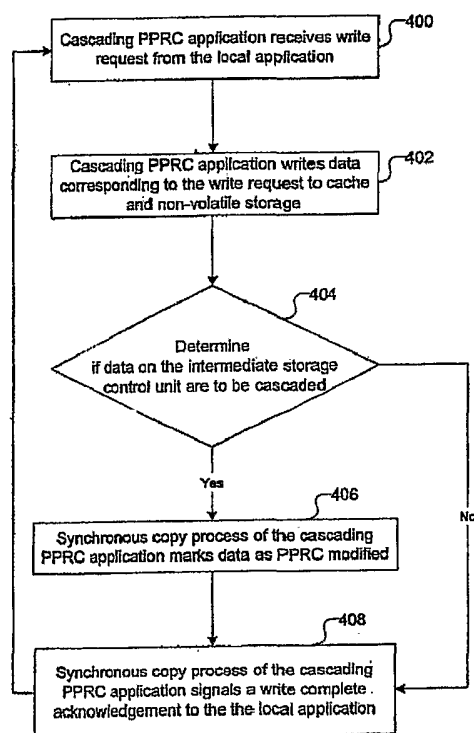
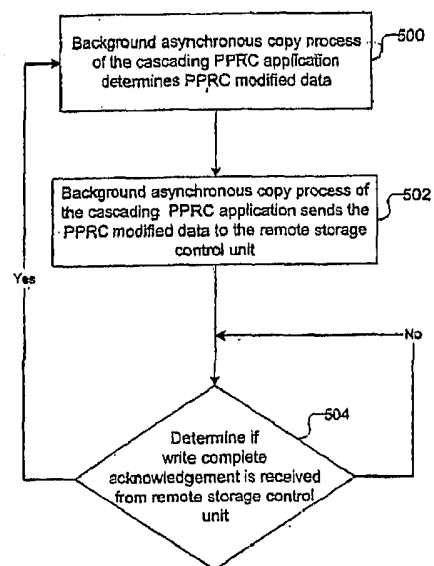


FIG. 3

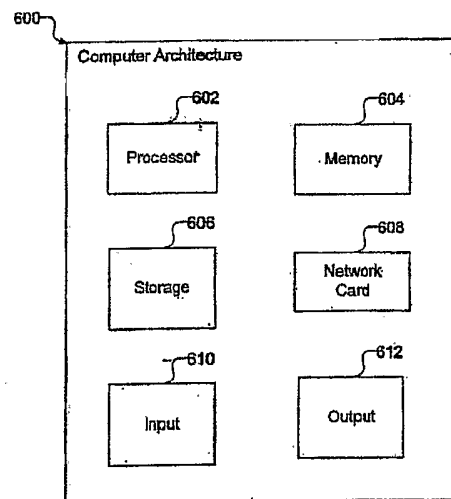
4/6

FIG. 4

5/6

FIG. 5

6/6

FIG. 6