IMPROVED REMOTE COPY SYNCHRONIZATION IN DISASTER RECOVERY COMPUTER SYSTEMS

Field of the Invention

The present invention relates to the field of disaster recovery computer systems, and more particularly to the resynchronization of primary and secondary copies of data after a disconnection and reestablishment of the Remote Copy pair.

Background of the Invention

In the field of computer data processing there is increasing demand for ways for organizations to continue doing business even after the loss of use of data or processing capability at the main business data processing site. The technique used in typical disaster recovery solutions is known in the art as "Remote Copy", or sometimes "Peer-to-Peer Remote Copy (PPRC)".

In a typical Remote Copy solution, one storage controller is designated as holding the primary disk of a remote copy relationship. The primary disk of the relationship will be referred to herein as the Master. The Master is the disk normally used by a person or organization for day-to-day processing. A second storage controller holds the secondary disk of the remote copy relationship, which will be known as the Auxiliary. The Auxiliary is the disk normally not used by a person or organization for day-to-day processing, but held in reserve in case of a need for disaster recovery or business continuity operations after the loss of use of the Master. Both Master and Auxiliary are the same size. Many solutions allow multiple sets of disks to be managed in a coordinated fashion, and often a controller might hold Masters for one relationship, and Auxiliaries for others, but for clarity and conciseness the present description will focus on a single relationship comprising two disks. In normal operation, the Master is used as the primary source and target of all host I/O requests. In these circumstances, the term Master/primary will be used in this description. Similarly, in normal operation, the Auxiliary is not used as the source or target of host I/O requests, but is used to hold a copy of the data from the Master/primary and to accept changes passed on to it from the Master/primary as a result of writes directed to the Master/primary. In these circumstances, the term Auxiliary/secondary will be used in this description.

The Master/primary is thus the disk that normally, in the absence of a disaster, holds the application data. The function of Remote Copy is to maintain a copy of that data on the Auxiliary/secondary disk.

To establish initial synchronization, all the data is copied from Master/primary to Auxiliary/secondary.

Once synchronization has been established, each write I/O received at the Master/primary is sent to the Auxiliary/secondary disk as well as to the Master/primary. Under normal situations, the Auxiliary/secondary does not receive writes from applications directly, but only indirectly from writes issued at the primary and forwarded to it.

In the event of a loss of connection between the two sites, a conventional technique that is well known is to use change recording at the Master/primary. This typically uses a bitmap to record which regions of the disk at the Master/primary have received write I/O. It is common to map a single bit to 32k of data, or some similar fairly small amount. Once the link is reestablished, the bitmap is used to resynchronize the Auxiliary/secondary, bringing it fully up to date with the Master/primary, by transferring data corresponding to every bit marked as changed in the bitmap.

However, there are uses of Remote Copy where this well-known scheme by itself is insufficient. One example is what is done after a disaster. Typically, if a disaster occurs at the Master/primary controller, then access to the Auxiliary/secondary controller is enabled, and the application is restarted using the storage there. This situation will be referred to in this description by using the term Auxiliary/primary.

However, the next thing that is needed is to reestablish a disaster recovery capability. In many 'disasters', the Master site is in fact physically intact, possibly only having suffered a power failure or a similar short-term failure. It is thus possible to use the Master (old primary) as the secondary of the relationship (thus creating a Master/secondary), and to have the Auxiliary become the primary (as an Auxiliary/primary, as defined above), essentially reversing the flow of data. While this is possible with today's products, they require that a full copy be performed from Auxiliary/primary to Master/secondary, repeating the problem faced by the user in the initial setup.

While this cost may at first appear to be acceptable because a real disaster is an infrequent occurrence, it must be borne in mind that testing the disaster recovery system is an essential part of any disaster recovery plan. Some companies and other organizations are required to demonstrate their disaster recovery capability in order to pass an audit, possibly as frequently as once a month. If the disaster recovery test involves carrying out a complete failing-over of the business as described above, the cost of a full copy from Auxiliary/primary to Master/secondary to reestablish synchronization is very heavy.

All known conventional schemes require a full copy after such a failover scenario, unless great care was taken to ensure that the application was completely halted at the old primary with no outstanding, "in-flight" updates, before switching the primary/secondary roles. This, however, is untypical of the way in which complex systems fail. Frequently, failures are of the type known as "rolling failures", where parts of the original Master/primary system fail over a period of time before the failover is triggered. In these circumstances, there may be changes made at the old Master/primary during the rolling failure of which the original Auxiliary/secondary has not been made aware.

It might be thought that the solution to the problem would be to set up the remote copy in reverse, and simply use change recording on the Auxiliary/primary to define what must be copied back to the Master/secondary after a disaster has been recovered. This is inadequate, because, as described above, changes might have happened at the original Master/primary during the failure, which were not change-recorded at the original Auxiliary/secondary. If these are not corrected, then the Master and the Auxiliary may never become truly synchronized.

A different scenario, but again one which might occur in the context of a disaster recovery or other form of test (such as an upgrade test), is where the Auxiliary/secondary is broken away from the Master/primary, and then directly receives write I/Os in its isolated state, perhaps from a test application, while the business continues to run as normal at the Master. Here, the resynchronization after reestablishment of the connection must be from Master to Auxiliary, even though the Auxiliary has been temporarily treated as an Auxiliary/primary while the Master was simultaneously being treated as a Master/primary. It is essential in this case that the real application data at the Master not be overwritten by the test data that has been applied at the Auxiliary during the period of its isolation from the Master.

It is therefore desirable to have an efficient means of Remote Copy resynchronization while alleviating the disadvantages of applying costly full copies of data as in the conventional systems described above.

Summary of the Invention

The present invention accordingly provides, in a first aspect a method, in a remote copy pair, for resynchronizing data held at a primary and secondary after the pair has undergone separation and reestablishment of a remote copy relationship, comprising the steps of sending a change record identifying data segments changed at said secondary from said secondary to said primary; merging said change record from said secondary with a change record identifying data segments changed at said primary into a merged change record; and copying all identified changed data segments indicated in said merged change record from said primary to said secondary.

Preferably, said step of merging is performed before the start of resynchronization.

Preferably, said change record identifying data segments changed at said secondary and said change record identifying data segments changed at said primary are bitmaps.

Preferably, said bitmaps are merged during said step of merging by the application of a logical OR operation.

Preferably, resynchronizing data held at a primary and secondary after the pair has undergone separation and reestablishment of a remote copy relationship is performed after a disaster has caused a failover.

Preferably, resynchronizing data held at a primary and secondary after the pair has undergone separation and reestablishment of a remote copy relationship is performed after completion of a test procedure.

The present invention provides, in a second aspect, an apparatus in a storage system for a remote copy pair, for resynchronizing data held at a primary and secondary after the pair has undergone separation and reestablishment of a remote copy relationship, comprising a communication component for sending a change record identifying data segments changed at said secondary from said secondary to said primary; a logic component for merging said change record from said secondary with a change record

identifying data segments changed at said primary into a merged change record; and a data copying component for copying all identified changed data segments indicated in said merged change record from said primary to said secondary.

The present invention provides, in a third aspect, a computer program product tangibly embodied in a computer-readable storage medium, comprising computer program code means to, when loaded into a computer system and executed thereon, cause said computer system to perform resynchronization of data held at a primary and secondary of a remote copy pair after the pair has undergone separation and reestablishment of a remote copy relationship by performing the steps of sending a change record identifying data segments changed at said secondary from said secondary to said primary; merging said change record from said secondary with a change record identifying data segments changed at said primary into a merged change record; and copying all identified changed data segments indicated in said merged change record from said primary to said secondary.

Brief Description of the Drawings

A preferred embodiment of the present invention will now be described by way of example only, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram illustrating an arrangement of apparatus in accordance with a preferred embodiment of the present invention.

Figures 2a and 2b show a flow diagram illustrating steps of a method in accordance with a preferred embodiment of the present invention.

Detailed Description of the Preferred Embodiment

An arrangement of apparatus according to a preferred embodiment of the present invention is shown in Figure 1. The arrangement is responsive to application I/O requests 100 to operate communication means 101 between a Master disk 102 and an Auxiliary disk 104. The normal communications flow responsive to application I/O requests 100 is shown by arrow 106 where the Master 102 is the primary disk in the relationship and the Auxiliary 104 is the secondary. On the occurrence of a disaster, the Master is failed and application I/O requests 100 follow the path

represented by arrow 108 directly to the Auxiliary 104 which operates as an isolated primary disk.

Resynchronization after the reestablishment of the Remote Copy relationship takes place when the change bitmap recording changes made at the secondary disk is sent to the primary disk, as represented by arrow 110. It will be clear to one of ordinary skill in the art that at resynchronization, it is a user decision as to which disk shall be designated the primary disk and which the secondary. In the example shown here, the Master has been designated as the secondary disk, and the Auxiliary has been designated as the primary disk, but the relationship could equally be established in reverse.

When the bitmap representing changes made at the Master/secondary 102 has been received by the Auxiliary/primary 104 and merged with the bitmap representing changes made at the Auxiliary/primary, all the changes from the resulting merged bitmap are copied from the Auxiliary/primary 104 to the Master/secondary 102, as shown at arrow 112. At some later stage, the roles of the Master 102 and the Auxiliary 104 may be reversed to return to normal operation, where the Master/primary 102 communicates with the Auxiliary/secondary 104 using communication means 101 to pass remote copy activity along the path represented by arrow 106'. This reversal of roles to revert to normal is not a requirement of the preferred embodiment of the present invention, but is to be expected in the normal course of user activity, as the Master/primary is likely to have been selected on some logical basis, such as locality or available day-to-day processing power.

Turning now to Figures 2a and 2b, there is shown a flow diagram representing steps of a method, which may be embodied using hardware, software or a combination thereof, according to a preferred embodiment of the present invention.

At step 200, the system is initialized, with, at step 201, the assignment of a Master/primary disk (effected in the preferred embodiment by write enabling the Master disk). Responsive to the receipt of Host I/O 202 the I/O activity (here exemplified as a write request) causes a write 204 to the Master/primary, followed by a copy step 206 to the Auxiliary/secondary. This process continues until a loss of connection is registered at test step 208.

In response to the loss of connection, the Auxiliary is designated as a primary disk (in the preferred embodiment, by write enabling the Auxiliary). On receipt of Host I/O 212, for example a write request, data is written directly to the Auxiliary/primary at step 214. This process continues until reestablishment of the Master/Auxiliary relationship is registered at test step 216. At step 220, a primary and secondary are designated. As described above, this designation may assign the Master or the Auxiliary as the primary, and its counterpart as the secondary. Prior to resynchronization, at step 222, the change recording bitmap from the secondary is sent to the primary and merged 224 with the change recording bitmap from the primary. At step 226, all data that was marked as changed in the merged bitmap is copied from the primary to the secondary. The resynchronization is complete when all changed data has been copied, and at step 228 processing is resumed.

Thus, in the preferred embodiment whenever a disk is enabled for write access, it is placed in a mode where change recording is enabled (in the most preferred embodiment, this is synonymous with being accorded the primary role). Whenever, a disk is inhibited from accepting writes (becomes a secondary), the change recording maps from both disks are merged and established as the new change recording bitmap. This is most preferably done before any resynchronization takes place, although the resynchronization may be begun in parallel in an alternative embodiment. For example, while the change recording map from the secondary is being sent to the primary, the process of applying a first set of changes that have been recorded at the primary may be started.

The term "merge" is used here to mean that, for each region of the disk encoded in the map, if either map indicates the region has changed, then the new map is set to indicate it has changed. This can conveniently be achieved, if the maps are bitmaps, by using a logical OR operation.

The merge requires that the change recording map be transferred from the secondary controller to the primary controller. In the preferred embodiment, this occurs over the same Fiber Channel connection as is used for data traffic, using a control connection that is established there. In one alternative, the map could be read into a host system, and then the merge could be driven to the primary controller from there, using, for example, a TCP/IP network for both transfers.

The resynchronization process is that for each region marked as changed in the merged map, data is copied from primary to secondary.

Thus, invalid or unwanted changes at the secondary are overwritten with valid data from the primary, and valid changes to data from the primary are applied to the secondary.

It will be appreciated that the method described above will typically be carried out in software running on one or more processors (not shown), and that the software may be provided as a computer program element carried on any suitable data carrier (also not shown) such as a magnetic or optical computer disc. The channels for the transmission of data likewise may include storage media of all descriptions as well as signal carrying media, such as wired or wireless signal media.

The present invention may suitably be embodied as a computer program product for use with a computer system. Such an implementation may comprise a series of computer readable instructions either fixed on a tangible medium, such as a computer readable medium, for example, diskette, CD-ROM, ROM, or hard disk, or transmittable to a computer system, via a modem or other interface device, over either a tangible medium, including but not limited to optical or analogue communications lines, or intangibly using wireless techniques, including but not limited to microwave, infrared or other transmission techniques. The series of computer readable instructions embodies all or part of the functionality previously described herein.

Those skilled in the art will appreciate that such computer readable instructions can be written in a number of programming languages for use with many computer architectures or operating systems. Further, such instructions may be stored using any memory technology, present or future, including but not limited to, semiconductor, magnetic, or optical, or transmitted using any communications technology, present or future, including but not limited to optical, infrared, or microwave. It is contemplated that such a computer program product may be distributed as a removable medium with accompanying printed or electronic documentation, for example, shrink-wrapped software, pre-loaded with a computer system, for example, on a system ROM or fixed disk, or distributed from a server or electronic bulletin board over a network, for example, the Internet or World Wide Web.

It will be appreciated that various modifications to the embodiment described above will be apparent to a person of ordinary skill in the art.

CLAIMS

1. A method, in a remote copy pair, for resynchronizing data held at a primary and secondary after the pair has undergone separation and reestablishment of a remote copy relationship, comprising the steps of:

sending a change record identifying data segments changed at said secondary from said secondary to said primary;

merging said change record from said secondary with a change record identifying data segments changed at said primary into a merged change record; and

copying all identified changed data segments indicated in said merged change record from said primary to said secondary.

2. A method as claimed in claim 1, wherein said step of merging is performed before the start of resynchronization.

3. A method as claimed in claim 1, wherein said change record identifying data segments changed at said secondary and said change record identifying data segments changed at said primary are bitmaps.

4. A method as claimed in claim 3, wherein said bitmaps are merged during said step of merging by the application of a logical OR operation.

5. A method as claimed in claim 1, wherein resynchronizing data held at a primary and secondary after the pair has undergone separation and reestablishment of a remote copy relationship is performed after a disaster has caused a failover.

6. A method as claimed in claim 1, wherein resynchronizing data held at a primary and secondary after the pair has undergone separation and reestablishment of a remote copy relationship is performed after completion of a test procedure.

7. An apparatus in a storage system for a remote copy pair, for resynchronizing data held at a primary and secondary after the pair has undergone separation and reestablishment of a remote copy relationship, comprising:

a communication component for sending a change record identifying data segments changed at said secondary from said secondary to said primary;

a logic component for merging said change record from said secondary with a change record identifying data segments changed at said primary into a merged change record; and

a data copying component for copying all identified changed data segments indicated in said merged change record from said primary to said secondary.

8. A computer program product tangibly embodied in a computer-readable storage medium, comprising computer program code means to, when loaded into a computer system and executed thereon, cause said computer system to perform resynchronization of data held at a primary and secondary of a remote copy pair after the pair has undergone separation and reestablishment of a remote copy relationship by performing the steps of:

sending a change record identifying data segments changed at said secondary from said secondary to said primary;

merging said change record from said secondary with a change record identifying data segments changed at said primary into a merged change record; and

copying all identified changed data segments indicated in said merged change record from said primary to said secondary.