

## Understanding Controller-based Replication VS. Host-based Replication

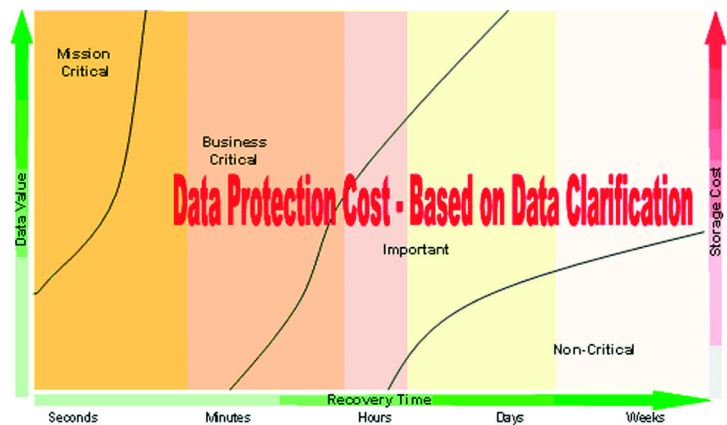


# Understanding Controller-based Replication versus Host-based Replication

## Introduction

Controller and host-based replication are common forms of data protection that are used in tiered storage architectures. Each has its positives and negatives. Controller replication is host-free, which improves system performance, but with a lower recovery point objective. Host-based has a better recovery point objective, but at the expense of system performance and significant additional cost. This technology brief discusses the pros and cons of each solution.

In general there are four classes of data to consider, as shown in Figure 1: non-critical; important - files; business critical - emails, schedules; and mission critical - online transaction processing (OLTP), enterprise resource planning (ERP), financial. Within the replication realm these can be divided into three groups, non-critical data in which some data loss is acceptable, important and business-critical data in which recovery tools are more than sufficient, and business and mission-critical data in which a host-based system is required. Data classification is all about managing costs and providing high return on investment (ROI) and enables a tiered storage architecture to be realized. An Intransa™ IntraStor™ storage area network (SAN) provides superior survivability for the first three classes and Intransa solutions partners provide effective solutions for the last class.



**Figure 1. Data classification structure.**

## Tiered Storage Architectures

The best practice for data protection is to utilize data classification to create a tiered storage recovery architecture<sup>1</sup>, which creates a highly survivable and cost-effective structure that can handle a wealth of data protection requirements. To better understand how to implement a tiered storage recovery architecture, this technology brief discusses two key replication technologies, controller-based replication, and host-based replication.

## Controller-based Replication Overview

Controller-based replication means transparent, host-free replication in which the application host is not impacted. No special agents are required on the host. As shown in Figure 2, there are four basic steps involved in asynchronous replication at the storage systems level. The first two steps are for the host to write the data to the primary storage system and for the system to return an acknowledgement that the data has been written to disk. The last two steps are for the primary storage system to write the data to the replica, over a network link such as a local area network (LAN), wide area network (WAN), metropolitan area network (MAN), or campus network and the acknowledgement to be returned to the primary storage system.

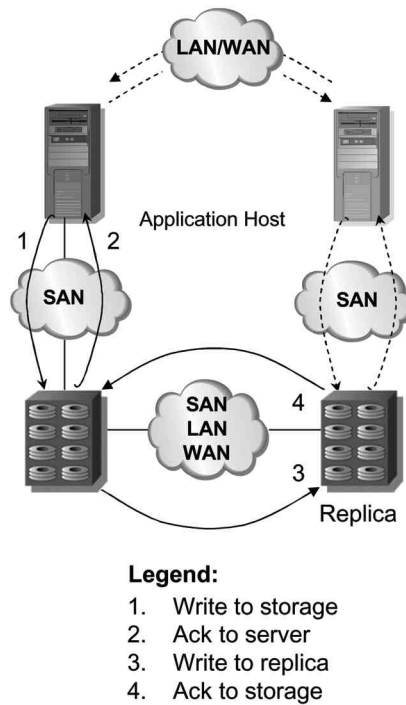


Figure 2. Controller-based replication.

**Data Consistency Issues**

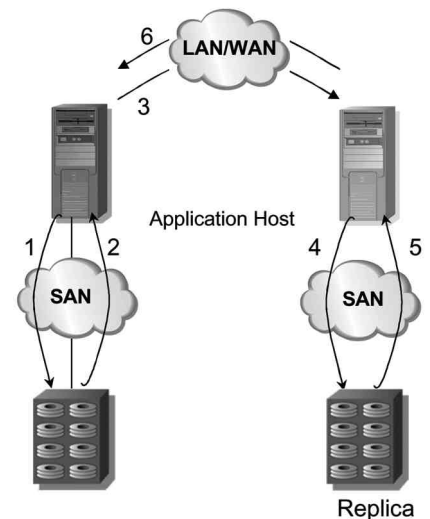
The benefit of the host-based replication approach is that the agent can help ensure that data is consistent across the board. This is particularly important in certain types of database transactions in which data is entered into the database, but may not yet be flushed to storage. The caching of the data in host memory improves host performance, but means that the database image on the storage is not consistent with the image in memory. Thus, the possibility for data corruption exists.

To appropriately scope the protection requirements, understanding what type of data is in use is critical. Three data groupings exist for consideration in replication, non-critical data that can suffer some degree of corruption, important and business-critical data that can be recovered using standard tools, and mission-critical data in which minimal corruption can be tolerated. Understanding the value of the data is important because it can be cost-prohibitive to provide mission-critical level of support to non-critical data. The first three classes of users cover 80 to 90 percent of data needs.

For synchronous and point-in-time replication, the steps are similar but the ordering is slightly different. In the asynchronous and point-in-time replication case, there is no impact on host performance to have data protection in the form of controller-based replication. The second host is only required to access the data in case of an outage. It is not required for replication to occur.

**Host-based Replication Overview**

Alternatively, host-based replication can be used to provide a similar type of protection. In host-based replication, two hosts are required and each must have a special agent installed. These agents are custom pieces of software that hook into the storage system datapath. Some may be agents that interact with application programming interfaces (APIs) such as Microsoft® Volume Shadow Copy Service (VSS) and Oracle® RecoveryManager (RMAN), others may appear to be device drivers or file system drivers. Upon each write to the storage, the agent will intercept the write and send it to another host over the network. This host will then write it to the storage to which it is attached. In this case six steps are used, as indicated in Figure 3. More importantly, host performance is impacted as multiple copies are required to write each item of data.



- Legend:**
1. Write to storage
  2. Ack to server
  3. Write to replica server
  4. Write to replica
  5. Ack to replica server
  6. Ack to server

Figure 3. Host-based replication.

For the first data grouping, such as file users, this is more than adequate as there are other data protection schemes in effect, such as backups. The additional costs for increased data corruption protection simply are not worth it. Most storage replication solutions, such as those provided by Engenio™ for StorageTek® and IBM®, provide this level of protection<sup>2</sup>. Intransa's IntraStor product family, coupled with StorAR<sup>3</sup>, provides a highly cost-effective and survivable solution for this class of data.

To handle the second data grouping, those with important and business-critical data, databases provide technology to ensure that the storage image is consistent with the database image. They also provide tools to recover from events that cause data corruption, whether in the form of system crashes or other failures. File systems have the same issues and provide the same tools, such as the "scandisk" program for Windows®. For the majority of database and file users this is more than sufficient and it is not worth the extra cost and overhead to completely eliminate all sources of data corruption. Intransa's IntraStor product family with StorAR and standard recovery tools is more than adequate to meet these needs — a complete solution without any additional costs over the solution for the first class of users.

However, these recovery tools are not sufficient for all users and a more robust solution is required. Specifically, financial transactions and other limited circumstances in which a single error results in significant pain, a host-based replication solution should be utilized. In this environment, what storage administrators are really looking for is hot-stand-by copies of the data. For this mission-critical environment, and third group of data, Intransa works with solutions partners to provide a qualified and proven host-based solution. (More information about these solutions can be found at the Intransa website, [www.intransa.com](http://www.intransa.com).)

## Conclusion

In the evolving world of tiered storage architectures, controller and host-based replication are solutions to different problems within the complicated realm of data protection. Controller-based replication protects the 80 to 90 percent of users who have either non-critical or important data. For these users data corruption can be minimized through the use of standard recovery tools that come with the applications that are run on the hosts. Intransa's IntraStor storage solution provides a highly survivable and cost-effective solution for these cases. For the remaining set of users that have mission-critical data that cannot tolerate any loss, Intransa solution partners and the IntraStor product family provide affordable and reliable data protection.

<sup>1</sup> See [www.intransa.com](http://www.intransa.com) for further information.

<sup>2</sup> See <http://www.engenio.com/default.aspx?pageID=359>.

<sup>3</sup> See Intransa StorAR datasheet for further information.

Copyright 2005 Intransa, Inc. 2870 Zanker Road, Ste. 200, San Jose, CA 95134 U.S.A. All rights reserved.

Intransa, the Intransa logo, Simply Smarter, IntraStor and StorControl are registered trademarks of Intransa, Inc. in the United States and other countries. Oracle is a registered trademark of Oracle Corporation. Microsoft and Windows are registered trademarks of Microsoft. IBM is a registered trademark of IBM. All other trademarks are the property of their respective owners.

THIS PUBLICATION IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NON-INFRINGEMENT. THIS DOCUMENT COULD INCLUDE TECHNICAL INACCURACIES OR TYPOGRAPHICAL ERRORS. CHANGES ARE PERIODICALLY ADDED TO THE INFORMATION HEREIN; THESE CHANGES WILL BE INCORPORATED IN NEW EDITIONS OF THE DOCUMENT. INTRANSA, INC. MAY MAKE IMPROVEMENTS AND/OR CHANGES IN THE PRODUCT(S) AND/OR THE PROGRAM(S) DESCRIBED IN THIS DOCUMENT AT ANY TIME.