



Data Protection with IBM TotalStorage™ NAS and NSI Double- Take® Data Replication Software

September 2002
IBM Storage Products Division
Raleigh, NC
<http://www.storage.ibm.com>

Data protection with IBM Total Storage NAS Storage and NSI Double-Take Data Replication Software

Table of contents

Introduction	3
Key Storage Challenges	3
Double-Take Architecture	4
Technology	4
Key Components	5
Mirroring	5
Replication	6
Failover	6
Restoration	7
Data Replication Scenarios.....	8
One-to-One.....	8
One-to-Many.....	8
Many-to One.....	9
Chained	9
Applications.....	9
High Availability.....	10
Storage Consolidation	11
Business Continuity.....	11
Enhanced Backup	12
Summary.....	12

Introduction

Combining the reliability and scalability of the IBM NAS storage and capabilities of Double-Take deliver solutions that:

- Enhance disaster recovery solutions
- Provide efficient answers to the shrinking backup window
- Offer lower Total Cost of Ownership alternatives to growth

This paper introduces and explains how the IBM TotalStorage Network Attached Storage (NAS) appliances in conjunction with the replication engine Double-Take can provide a complete solution for enterprise storage networks.

The IBM NAS portfolio of products is designed to deliver user value, advanced systems management capabilities and task-optimized operating system technology. The latest portfolio offers several enhancements to earlier product models, including:

- Enhanced hardware redundancy
- Improved client performance for Windows and UNIX® systems
- Expanded tape backup capabilities
- Instant volume- and drive-level data restoration

Double-Take is a real-time data replication and failover software product that allows replication of mission-critical data from a production appliance, a.k.a. *the source*, to a backup appliance, a.k.a. *the target*.

Key Storage Challenges

Enterprise and non-enterprise companies are accumulating files and looking for a means to efficiently manage these files. Today, it is not uncommon for a company to have multiple servers with different operating systems and thousands or millions of files scattered across the network. Administering and managing these servers can be a challenge, especially when the IT staff is already overworked.

- Geographical Dispersion

In their quest for efficiency, businesses are becoming more geographically dispersed. Offices commonly cross time zones and employees work varying schedules with the advent of employee-friendly flexible work hours. The result is that the window of time available for administrative tasks, such as backups of the servers, is getting smaller. The shrinking window cannot accommodate the demands of increased storage capacity. Considering that data growth occurs where storage capacity is readily available, such as on any server or storage device, growth is taking place at different locations on the network, almost randomly, making it increasingly difficult for IT to manage the data.

- Data Protection

As seen in recent current events, disaster recovery is an important issue for many businesses, and it is a requirement rather than an option. Businesses that already have a disaster recovery solution are reevaluating their existing plans for adequacy, and many are finding that they are not. As such, many businesses are searching for ways to increase the protection of their data by making their existing solutions disaster proof.

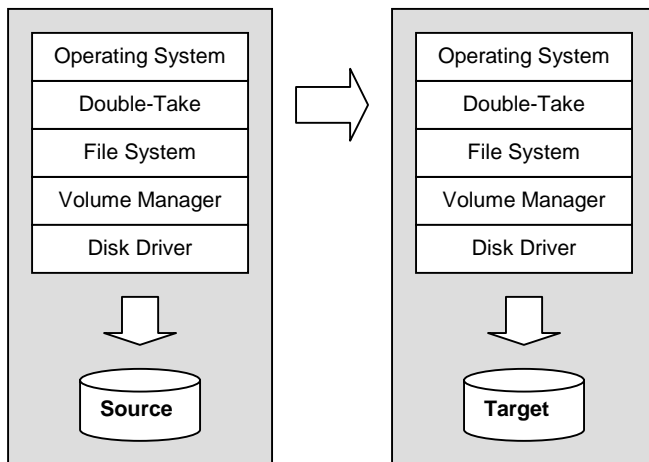
- Time to Recover

The need to access more data faster, even when the tape backup technology is advancing rapidly, is constantly becoming more important. For example, the total tape restoration times continue to increase. A typical 50 GB server could be restored in just over 16 hours at a rate of around 50 MB per minute with traditional tape backup technology. Even with the more advanced tape backup technologies, restoration of a 1 TB server at 900 MB per minute, or 52 GB per hour, will take over 19 hours to restore. During this time the data would be unavailable to users, making the minimum amount of time for users to access their data nearly a day.

Double-Take Architecture

Double-Take uses asynchronous byte-level replication to transfer data, allowing just the file changes to be captured and replicated. This means that if a 10 KB change is made to a 20 MB file, only the 10 KB of changed data is replicated across the network. Unlike synchronous technologies which can actually slow down the production server, the Double-Take replication process has little impact on the network and server resources during replication. On the target side, the data is readily available and accessible. There is no need to mount the volume or run any disk utilities, like checkdisk.

Double-Take is a real-time replication engine that captures write operations at the file system level and before the operation is written to disk, it replicates it to the target. Because of where Double-Take resides within the operating system, it even replicates open files, like database files, without the need for an application backup agent.

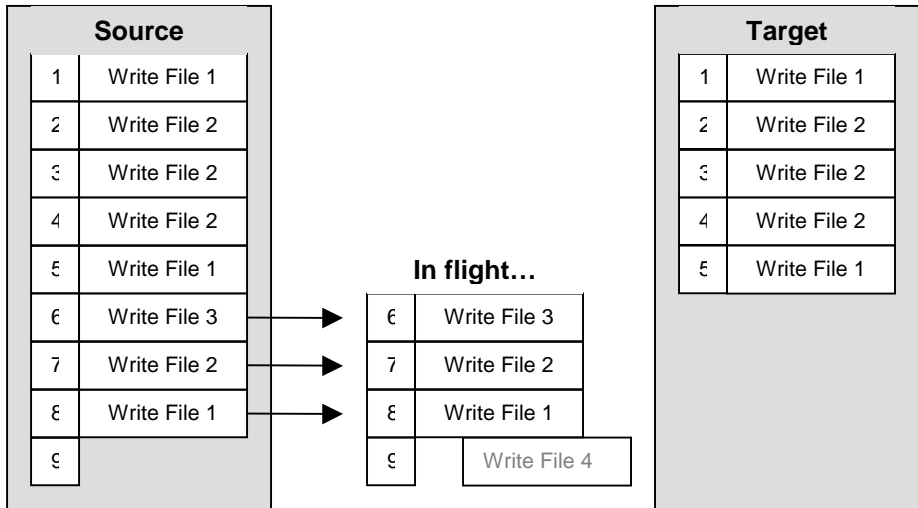


Double-Take resides at the operating system level and, as files write requests pass through the file system stack, it captures those requests and transmits them to the target.

Technology

Double-Take is based on the NSI patented *Sequential Transfer Asynchronous Replication* (STAR) technology:

- **Sequential Transfer**
The replication follows the same write sequence within and across multiple files, providing complete data integrity at all times.
- **Asynchronous Replication**
A combination of source and target-based queuing capabilities allow the server to continue processing while replication occurs and allows for maximum performance.



The technology utilizes sequenced write-order-preservation when replicating data. This means that data is always written to the target in the same order it is written on the source. By preserving the order of how data is written when it replicates Double-Take avoids the possibility of data corruption.

Key Components

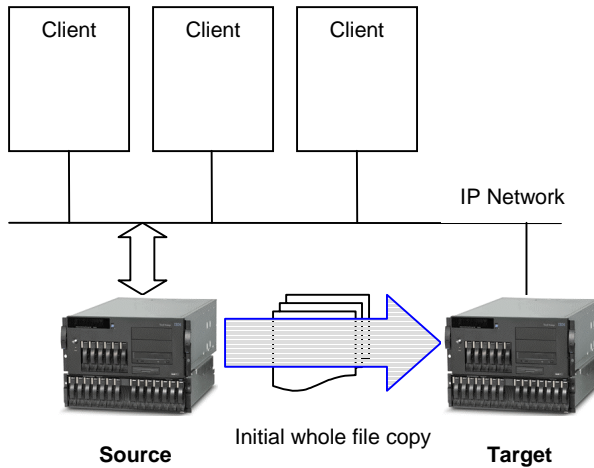
Double-Take is based on four key components: mirroring, replication, failover, and restoration.

The IBM NAS storage is based on Microsoft’s Windows 2000 Advanced Server operating system or Windows Powered OS. A NAS storage product is a dedicated server that delivers high-performance, file-level access to network clients through the standard TCP/IP network protocol. Although they are dedicated file servers, the IBM NAS storage provide the same set of integrated scalability, reliability, availability, and management components as general-purpose Windows 2000 based servers. As such, Double-Take may be installed on IBM NAS storage, or a general purpose server, running Windows 2000 or the NAS variant, Windows Powered OS.

Mirroring

Mirroring is the process of transmitting user-specified data from the source to the target so that an identical copy of data exists on the target. When Double-Take initially performs mirroring, it copies all of the selected data including file attributes and permissions. Double-Take mirrors the data to initially synchronize the production, i.e. *Source* and secondary, i.e. *Target* servers. A variance of this option is called a *partial mirror* that allows the servers to synchronize based on individual file block checking and reduces network traffic.

If subsequent mirroring operations are necessary, Double-Take can mirror specific files or blocks of changed data within files. By mirroring only files that have changed, network administrators can expedite the mirroring of data on the source and target machines.

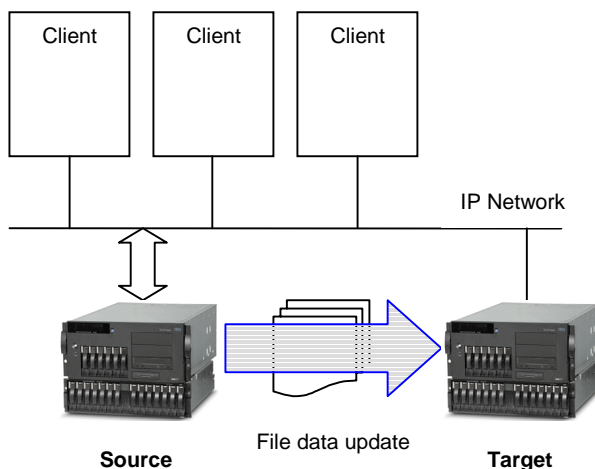


Replication

Replication is the real-time transmission of file changes. Double-Take's replication process operates at the file system level and is able to track file changes independently from the file's related application. In terms of network resources and time, replicating changes is a more efficient method of maintaining a real-time copy of data than copying an entire file that has changed.

Unlike mirroring which is complete when all of the files have been transmitted to the target, replication captures the changes as they are written to the source and sends them to the target continuously. Replication keeps the target up-to-date and synchronized with the source.

Replication is based on individual byte-level changes and is real-time. With replication, it is possible to replicate files open by other applications like SQL or Exchange. Replication keeps the target data current with the source.

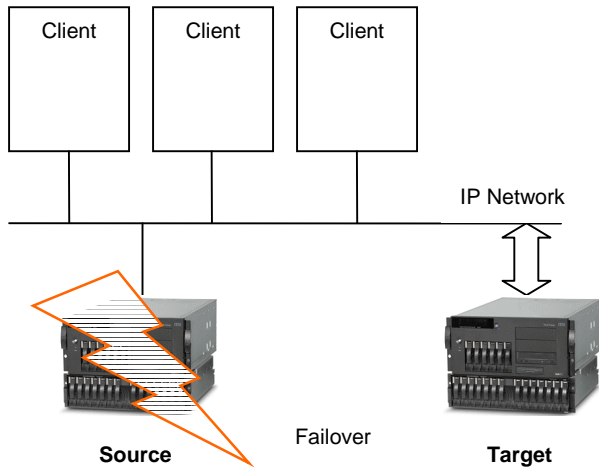


Failover

Failover is a process in which a target machine stands in for a failed source machine. As a result, user and application requests that are directed to the failed source machine are routed to the

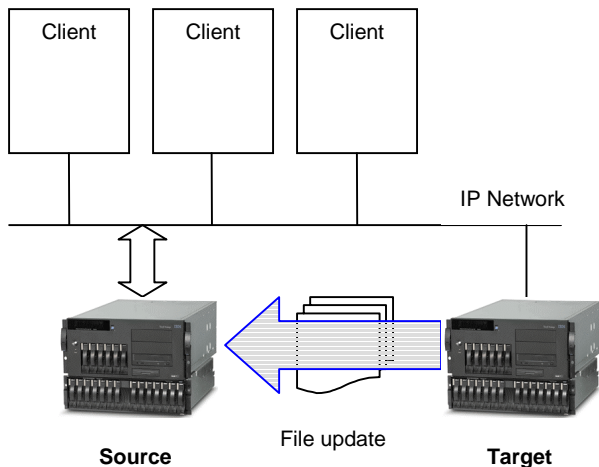
target machine. Double-Take monitors the status of machines by tracking network requests and responses exchanged between source and target machines. When a monitored machine misses a user-defined number of requests, Double-Take assumes that the machine has failed. Double-Take then prompts the network administrator to initiate failover, or, if configured, it occurs automatically.

When used in data replication capabilities, failover routes user and application requests with minimal disruption and little or no data loss. In some cases, failover may be used without data replication to ensure high availability on a machine that only provides processing services, such as a web server.



Restoration

Restoration provides an easy method for copying replicated data from the target back to its original location on the source. The process only requires selecting the source, target, and the appropriate replication set. Restoration can be used if the source data is lost due to a disk crash or when the most up-to-date data exists on the target due to failover.



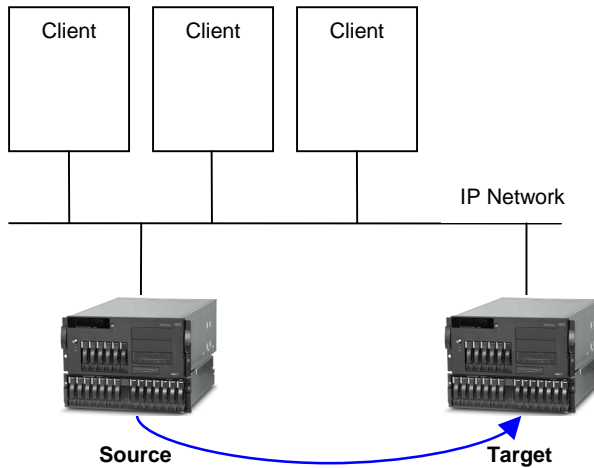
At the time of a source machine failure, the Double-Take target will contain the same data as the source. If the failover capabilities are employed, then users can continue updating data on the target machine while the problems on the source are resolved. Because of the continued updates on the target, when the source machine is ready to come back online, the two machines will no longer contain the same data. Restoration is the process of copying the up-to-date data from the target back to the original source or new source location when bringing the source back online.

Data Replication Scenarios

Considering the flexibility of Double-Take, many scenarios can be developed to meet a diversity of business needs. These scenarios may be deployed independently or in combination to suit more complex situations.

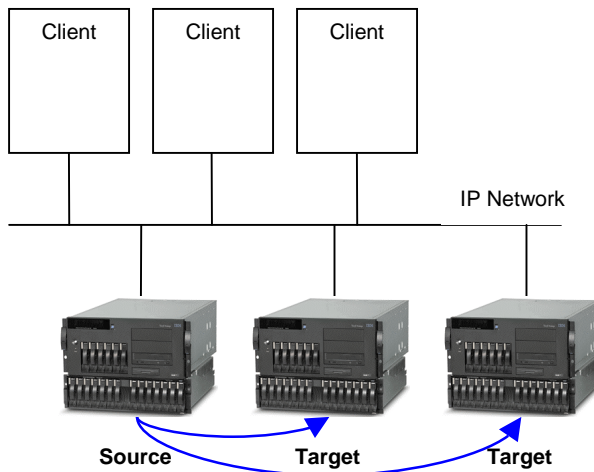
One-to-One

A one-to-one scenario contains one target and one source. It can be active/passive where the target is dedicated to supporting one source machine or active/active where both machines are actively replicating. This scenario is appropriate for failover, offsite disaster recovery, and critical data backup, such as Exchange, SQL, and web servers.



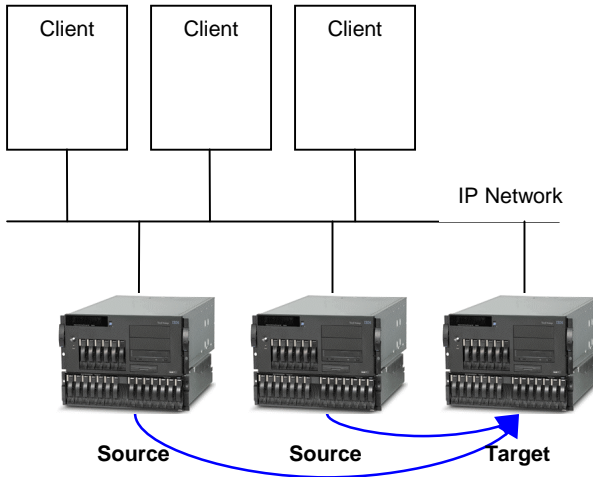
One-to-Many

One source machine sends data to multiple target machines. The target machines may or may not be accessible by one another. This scenario provides offsite disaster recovery, redundant backups, and data distribution. For example, all data can be replicated to a local target machine and a subset of the data can be separately replicated to an offsite disaster recovery machine.



Many-to One

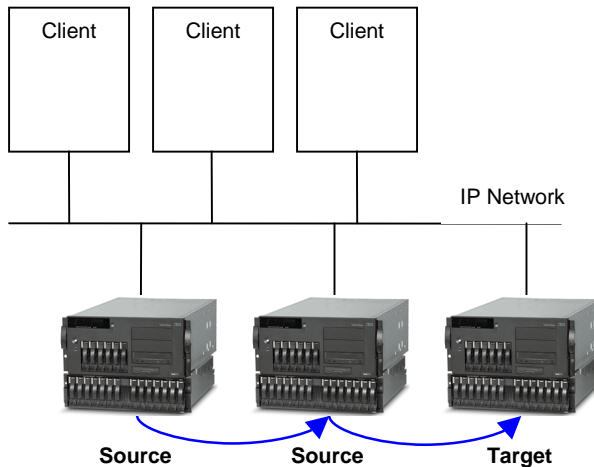
Many source machines are protected by one target machine. This scenario is appropriate for offsite disaster recovery. It is also an excellent choice for providing centralized tape backup because it spreads the cost of one target machine among many source machines.



In the case of Microsoft SQL Server, for example, in a one-to-one configuration, the SQL database should be replicated to the same location on the target so that failover is automatic. In a many-to-one configuration, each SQL data store will need to be replicated to a unique location and then renamed to the corresponding SQL directory on the source before failover occurs.

Chained

A source machine sends replicated data to a target machine, which in turn acts as a source machine, sending selected data to a final target machine, which is often offsite. This is a convenient approach for integrating local high availability with offsite disaster recovery.



Solutions

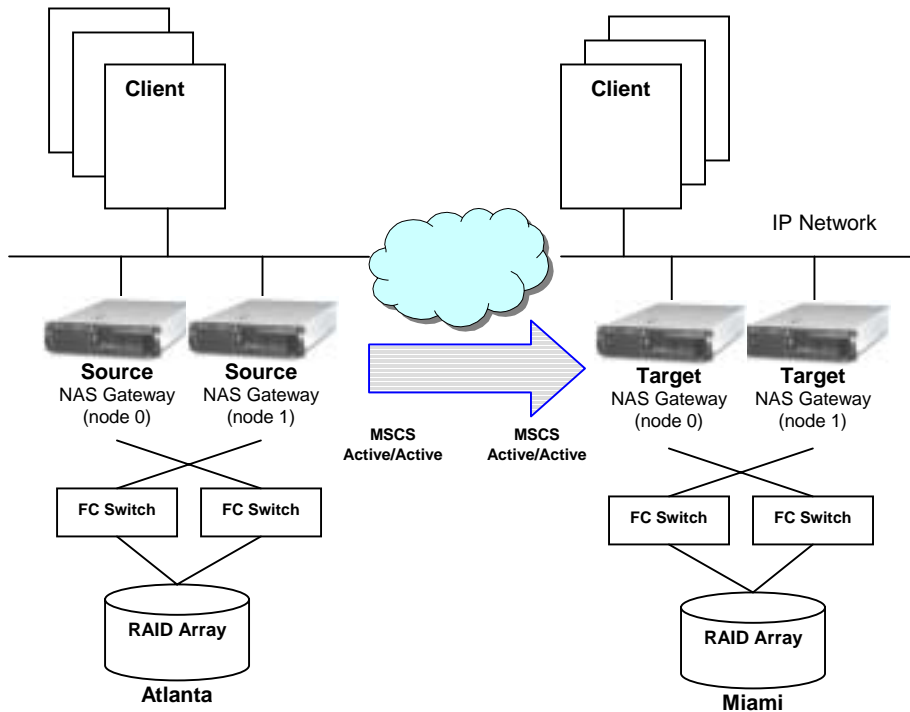
The following are some applications that, when deployed with NSI Double-Take and IBM NAS appliances or SAN storage products deliver robust, cost efficient solutions for availability, storage consolidation, business continuity and enhanced backup.

High Availability

Clustering products for the Windows platform improves both availability and manageability to reduce planned and unplanned downtime. Deploying clustering improves the availability of applications by providing failover capability in the event of a hardware or software error. Microsoft Cluster Server (MSCS) supports 2-node failover on Windows 2000 Advanced Server and 4-node failover with Windows Datacenter Server.

Double-Take provides a reliable local or remote replication solution for Microsoft clusters. Commonly, MSCS is implemented with both nodes in close physical proximity. In the event of a site disaster, the cluster could be lost. Setting-up with Double-Take a local or remote failover from one cluster to another, protection is available from a site disaster, cluster failure and even routine cluster maintenance. Double-Take supports Active/Active, as well as Active/Passive configurations.

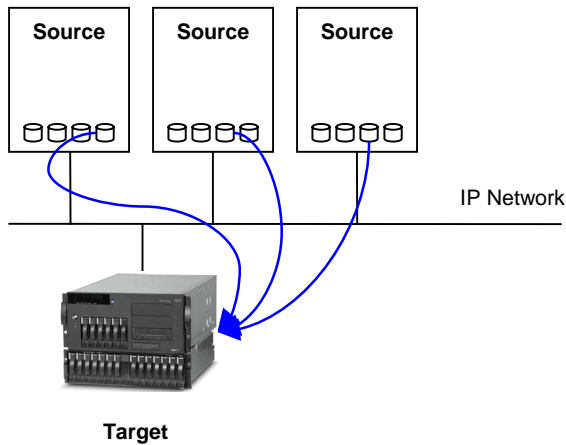
The following diagram illustrates a high availability MSCS configuration with Double-Take installed on clustered IBM NAS gateways.



Storage Consolidation

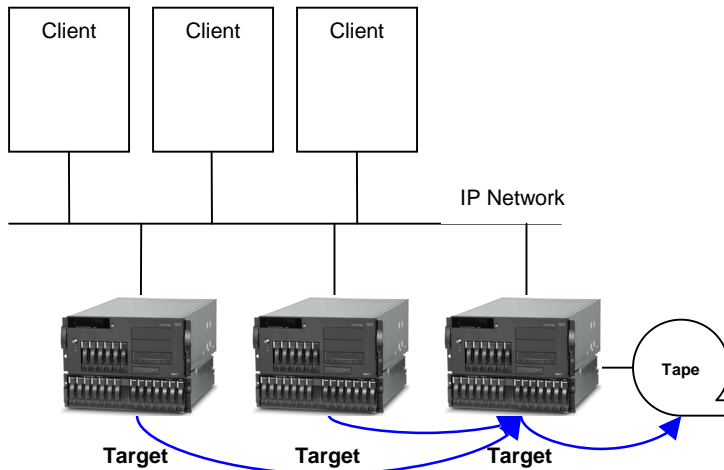
To keep the growing amounts of essential data available, but still keep storage affordable, businesses are looking for ways to better utilize existing resources. To deal with these problems, they are turning to storage consolidation to increase efficiency and save valuable resources. Most businesses have many servers from different vendors, often running different operating systems. Managing these islands of computing in a rapid growth of stored data is difficult to manage and costly. Consolidating storage via storage networking allows bringing together the storage into a centralized, easy-to-manage combined data pool. NAS and Storage Area Networks (SAN) are the established types of storage networks relevant to storage consolidation

Double-Take can be used to mirror and replicate the data to the NAS appliance. Once the NAS device is synchronized with the old machines, the cutover period between the old and the new storage devices is reduced.



Business Continuity

Restoring data from tape is a common practice when recovering after a disaster. The time to become operational again can be considerable, even with fast tape drives or libraries. With Double-Take, the data is always live and accessible on the target. The common practice for a tape solution is to backup the data nightly. This could potentially allow a window of lost data of an entire business day. Because Double-Take is a real-time replication engine, the data is up-to-date the moment that the production data is lost.



Enhanced Backup

Tape backup can introduce significant overhead to a production server, seriously impacting performance. While the impact of periodic full system backups may be obvious, even nightly incremental backups must “walk the file system” to examine every file system object and then read any file that has changed in its entirety for backup. Performing this process across a network adds additional overhead as this entire process must happen “across the wire”.

Double-Take can offload the burden of periodic tape backup from multiple production servers to a dedicated backup system and enable centralized backup, thus potentially reducing management cost and improving reliability.

Double-Take eliminates the need for a backup agent on the production server because of its interaction with the file system and ability to access files even if they are open. On the target, if a file is being backed up via tape, the changes to that file will be queued automatically and applied once the file is released by the backup process.

Summary

Using Double-Take with your NAS storage solution can help you overcome many of the business needs of today as well as overcome the storage needs for the future. Double-Take allows a company to utilize their NAS storage more efficiently by empowering you to consolidate data, ignore backup windows, and increase server performance by offloading backup responsibilities.



© International Business Machines Corporation 2002

IBM America

Department CNPA

PO Box 12195

Research Triangle Park NC 27709

U.S.A.

Printed in the United States

09/02

All rights reserved

IBM, the IBM logo, DB2®, DB2 Universal Database™, eServer logo, Enterprise Storage Server, FlashCopy, ThinkPad, TotalStorage, and xSeries are trademarks or registered trademarks of International Business Machines Corporation.

Microsoft, Windows, and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both. Tivoli is a registered trademark of Tivoli Systems, Inc. UNIX is a registered trademark of The Open Group in the United States and other countries.

Other product names are trademarks or registered trademarks of their respective companies.

This material has not been submitted to any formal IBM test and is published AS IS. It has not been the subject of rigorous review. IBM assumes no responsibility for its accuracy or completeness.

The use of this information or the implementation of any of these techniques is a customer responsibility and depends upon the customer's ability to evaluate and integrate them into the customer's operational environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

References in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part the materials for this IBM product and use of those Web sites is at your own risk.

The authors have utilized their professional expertise in preparing this report. However, neither International Business Machines Corporation nor the authors make any representation or warranties with respect to the contents of this report. Rather, this report is provided on an AS IS basis, without warranty, express or implied, INCLUDING A FULL DISCLAIMER OF THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE.