



SmartSync – NAS-to-NAS Data Replication

7/23 Henry Ho

1. Abstract

Data replication has become a common feature among NAS systems. It provides a cost-effective and efficient implementation of remote data redundancy. There are different designs of implementing NAS data replication. Among them are volume-based and file-based approaches. Ingrasys designed the data replication feature for NAStorage – SmartSync – by balancing between the two approaches. It replicates data file by file, while deploying differential block transfers during file replication. It is a balance between flexibility and efficiency.

2. What Is Data Replication?

Data replication, in short, is to copy data from one site to another site over network. All files, including directory hierarchy, under the specified source location will be replicated to the target site. All data replication should take place automatically, without human intervention, to provide data redundancy.

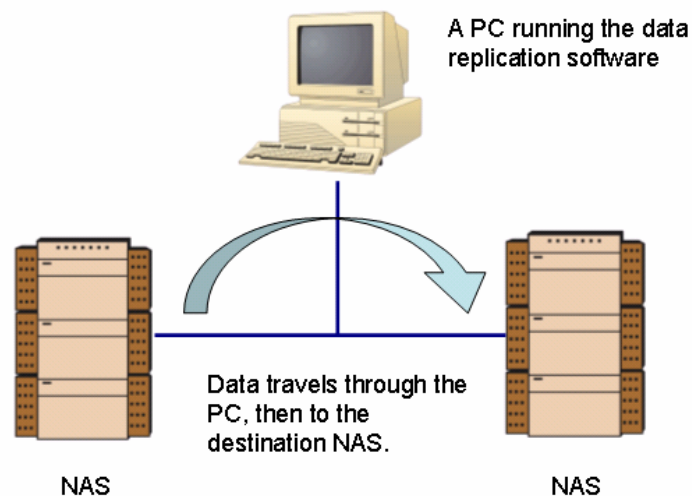
3. Why Data Replication?

- **Data recovery** – to replicate data to a mirror site for data redundancy. Once the master site fails, the mirror site can be in service.
- **Data distribution** – to distribute design files or e-learning materials to one or more remote sites regularly.
- **Tapeless backup** – to complement existing tape backup solutions, providing on-line access to data backups and facilitating data recovery.
- **Data migration** – data, along with the associated ACLs, can be transferred with SmartSync. It simplifies data migration jobs.
- **Load sharing** – clients can have instant access to those replicated data instantly, reducing the load of the primary site.

4. Different Approaches of Implementing Data Replication

- Host-based replication

The host-based approach is to run some data replication software on a PC, monitoring changes in the source NAS and replicating data to the target NAS. This method has been commonly used in situations where there are only small amounts of data.



Pros:

Since the replication software does not reside in the storage systems, you can replicate data between different NAS systems from any vendors.

Cons:

Firstly, it is inefficient. All data transfers between NAS systems must go through the PC server. It has big impact in efficiency and increases network traffic a lot.

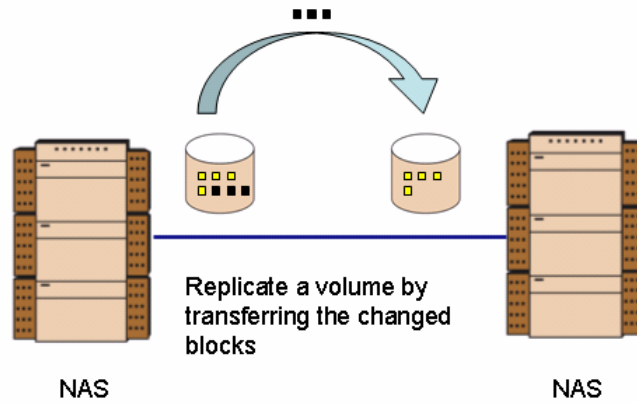
Moreover, it is OS-dependent. If you run the data replication software on a Windows server, basically it replicates data to and from Windows shares. The NAS systems must have Windows shares in them.

- Volume-based replication

This approach is to replicate a whole volume with the built-in replication software in the NAS systems. Most storage systems that use this approach



will log the blocks ever changed since last replication. It only transfers the ever-changed blocks during replication.



Pros:

The whole volume is replicated. It means that the whole file-system is preserved. All directory hierarchy, file attributes and security settings are all replicated.

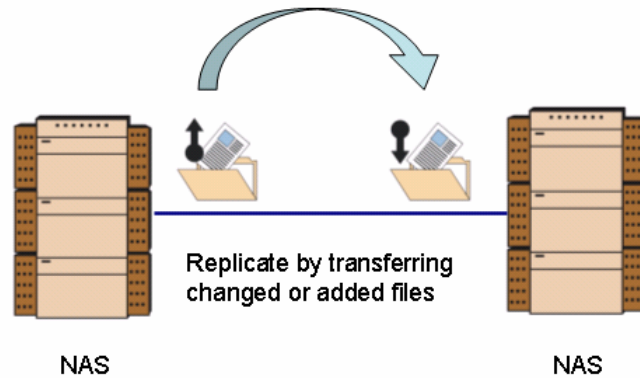
Moreover, by logging and transferring only those blocks ever changed since last replication, replication efficiency is greatly improved.

Cons:

It is less flexible in that it must copy the whole volume. You cannot specify a folder to replicate.

- File-based replication – Windows File Replication Service (FRS)

With Windows DFS (Distributed File-System) enabled, you will be able to use its replication services, called FRS. FRS supports continuous replication and replication schedules. The former will monitor any file changes in the replica set and copy the whole files whenever they are changed or added. The latter can do the replication on predefined schedules so that it can take place during off-hours.



Pros:

FRS tightly integrates into the Windows Server infrastructure. Using the FRS, you can have the features like unified namespace and security management of Active Directory. You can also increase data availability by incorporating the Windows clustering feature.

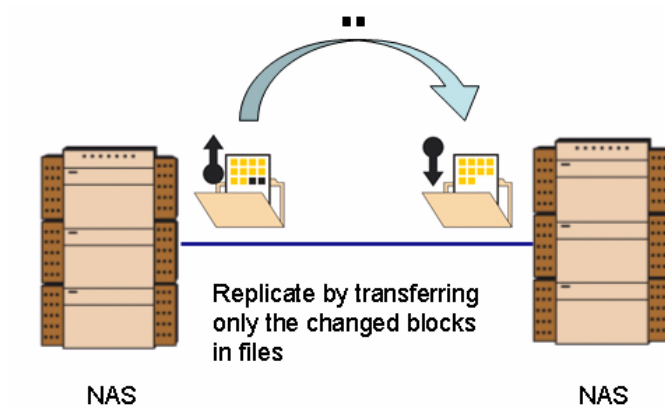
Cons:

FRS will transfer the whole file even if there is only one byte changed. It is extremely inefficient when files are big and frequently updated.

Next, Windows FRS only works on Windows platforms. You need at least one Windows Server hosting a DFS tree. It only replicates data across Windows systems, including Windows-powered NAS.

- File-based replication with differential block transfers – SmartSync

The data replication function of NAStorage, SmartSync, tries to balance between the volume-based and the file-based approaches. You do not have to replicate a whole volume. Instead, you can specify a folder to replicate. During replication, it checks and finds out the different blocks between the source file and the target file. It only transfers those different blocks to bring data into synchronization. You do not have to transfer the whole file if only one byte is changed. Only the different blocks are transferred.



Pros:

Since SmartSync only transfers the different blocks, network traffic is minimized. It is efficient. At the same time, you do not have to replicate the whole volume every time. You can replicate only those data you want. It is flexible.

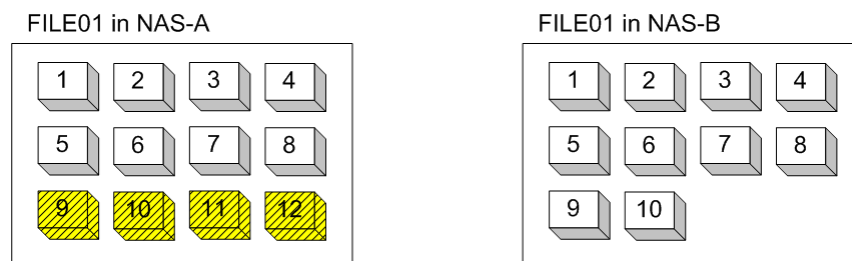
Cons:

It needs memory when it is finding different blocks between files. NASStorage requires at least 256MB of memory to run SmartSync tasks.

5. What Does It Mean By “Differential Block Transfer”?

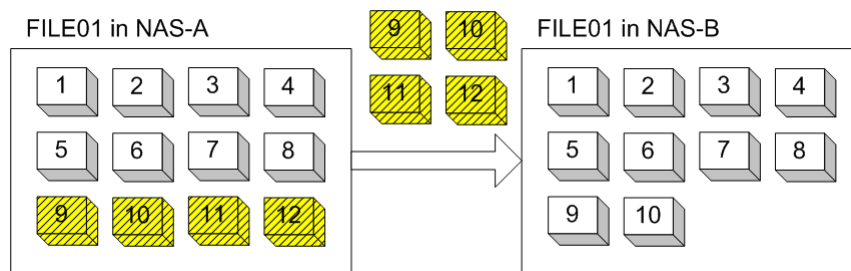
Let us explain by an example. Suppose that FILE01 was modified since last replication. The FILE01 in NAS-A (the source) is different from NAS-B (the target) by several blocks. To synchronize the two files, it will transfer those different blocks from NAS-A to NAS-B. The operations are as the figures below.

(1) Before replication, the two files are different but similar. Suppose that only 9th, 10th, 11th and 12th blocks are different.

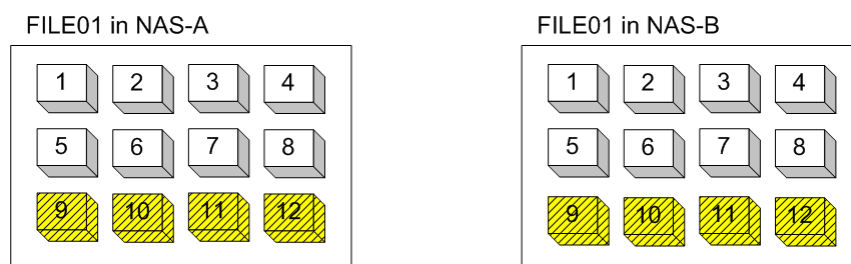




(2) During replication, SmartSync transfers the different blocks to NAS-B.



(3) After replication, FILE01 in NAS-B is made identical to FILE01 in NAS-A.



With this algorithm, SmartSync has the flexibility of file-based replication and the efficiency of block-level transfers. It can replicate only portions of data, but it does not cause so much network traffic as normal file-based replication.

6. More About SmartSync

- **Fully Utilize the Scarce Network Bandwidth**

Data replication should do more than simply copying data from one site to another. It must take into consideration that network bandwidth is scarce resource and the efficiency of bringing data into replication. SmartSync has designed with the features for effectively utilizing network bandwidth.

- ◆ **“Block-level” incremental file transfers**

SmartSync only transfers the differential blocks in files, thus largely reducing demands on network bandwidth. It checks files block-by-block and only sends those blocks which are required to bring files into synchronization.



- ◆ **Compress data streams on the fly**

SmartSync can compress data stream during replication. Files are compressed on the master site, transferred to and decompressed on the mirror site. For certain file-types, it can trim down data size and minimize network bandwidth usage. Administrators can decide whether to turn on the compression option.

- ◆ **Bandwidth control**

To allocate the network resource, SmartSync also allows administrators to specify the maximum network bandwidth. Administrators can assign different bandwidth limits for different SmartSync tasks, depending on applications and priorities.

To further assist in resource management, the SmartSync software displays all running and idle tasks and their progress. Administrators can watch task status and abort some tasks for others to run faster.

- **Enhanced Security**

- ◆ **Secure connection**

SmartSync jobs can be invoked between NAStorage systems over LAN or WAN. If data are to be sent over WAN, security becomes a critical issue. SmartSync encrypts data stream during replication. No data alteration or eavesdropping is possible. It ensures that data are transferred in a secure manner.

- ◆ **Replicate security settings along with files**

Data replicated to a mirror site might be exposed in security risks if the security settings (access control lists, to be more specific) are not replicated and applied. SmartSync can transfer all the related security information together with data. The same access control (file/folder-level security) can apply on the mirror site just the same as on the master site.



7. Key Features of SmartSync

- Configure and monitor all SmartSync tasks on web-based UI
- Instant access to replicated data
- Set sync points at any volumes or folders
- Only transfers the differences in files, with efficient network utilization
- Optional compressed data link
- All data transfers are encrypted, preventing risks of being eavesdropped
- The administrator can view task progress and abort a task at any time
- Security information (ACL) can also be replicated
- Provides bandwidth control