



SmartSync Backup – Efficient NAS-to-NAS Backup

1. Abstract

A common approach to back up NAS data is to run backup software on Windows or UNIX systems and back up NAS via network. The backup time depends on network traffics, NAS performance and tape drive speed. It is also somewhat complicated to restore data when you deal with many tape cassettes, and incremental backups, full backups, cross-tape backups, etc.

The SmartSync Backup runs on the NAStorage servers. It tries to improve backup speed by only transferring the modified data blocks, instead of whole files. At the same time, it always builds full backups online and keeps them online, making it easy for data restoration.

2. The Design Concepts of SmartSync Backup

Currently most NAS backups rely on a backup server running some backup software. This backup solution is common, yet not perfect in some aspects. The SmartSync Backup software can improve NAS backups by incorporating the techniques of remote synchronization and to complement existing backup solutions.

The issues that the SmartSync Backup software tried to deal with are:

- Long backup time

Backup time is affected by many factors – how fast source data can be read from NAS, how busy network traffics are, how many mega-bytes can be written to a tape per minute, etc. IT people always try to improve backup speed and shorten backup window to minimize impact on business operation. They have few choices but keep investing money in high-end hardware, high-speed networks and fast tape drives.

What else can we do except paying a lot in purchasing high-end hardware, software and upgrading network infrastructure?



The SmartSync Backup software tries to improve backup speed from the fundamentals – the algorithm.

Traditional backup software copies the whole files to tapes when making backups even if there are only several bytes changed. In most cases, it wastes much resource since files usually do not undergo drastic changes. Different versions of a file may have 70% in common even if it is modified very often. If we make daily backups of the file, we might transfer the whole file everyday, including the same 70% of that file. Why not just transfer the different 30%?

The idea of SmartSync Backup is to transfer only the modifications since last backup. Backup speed is greatly improved since only parts of files need to be transferred. In addition, it reduces demands on network traffics during backups. The impact on business operation is minimized.

- Difficulties of restoring data

Typically, when data crashes, IT people have to restore the last full backup and some incremental backups in order to re-establish data. The process is complicated. First, you have to locate the data, finding out which tapes hold those files. Second, you have to find those tapes which are placed off-line or even off-site. Then, perform data restoration, twice or more – one for restoring a full backup, others for incremental backups.

The SmartSync Backup keeps all backups online. Even better, it always makes and keeps full backups. IT people only have to perform data restoration once – just to pick and restore one backup version. All tasks can be done online, straightforward and easy.

- Security concerns

Since backup data are sent to the backup server over network, there are some security risks. The SmartSync Backup packs data with strong encryption to prevent any eavesdropping. Anyone who intercepts data packets will not be able to understand them.

Fast backups. Full backups. On-line backups. These are the design concepts of the SmartSync Backup software. The purpose is to shrink the backup window and facilitate data restoration. The enabling technology is the SmartSync algorithm, which utilizes differential block transfers, reducing demands on network bandwidth and largely shortening backup time. It is now possible and feasible to make full backups all the time. MIS people do not have to deal with the complexity of picking tape, finding files and restoring multiple backups. Just choose a backup and restore. Data restoration is simplified.

3. How SmartSync Backup Works

SmartSync Backup performs NAS-to-NAS backup. Two or more NAStorage systems are required, one as the SmartSync backup server, others as backup clients. The SmartSync backup server makes and keeps backups of client data.

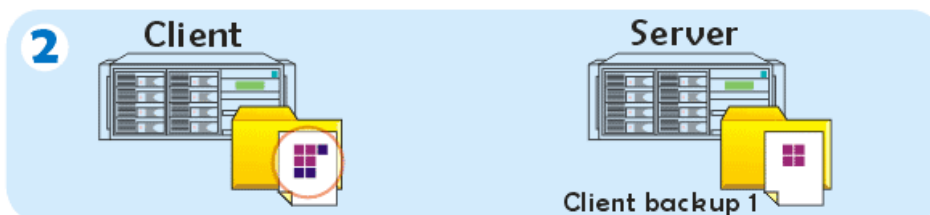
(**Note:** All NAStorage models can be SmartSync backup clients, while NAStorage 8200 or above can be SmartSync backup server.)

Files are composed of data blocks. To back up data, the clients send data blocks to the SmartSync backup server. Let us use some illustrations to see the internal operation of SmartSync Backup and why it has high efficiency.

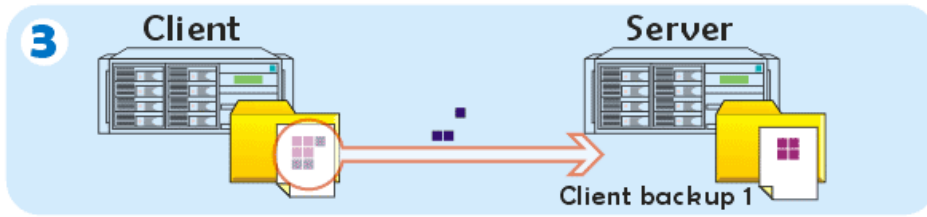
1. Originally, there is a backup of client data in the SmartSync backup server.



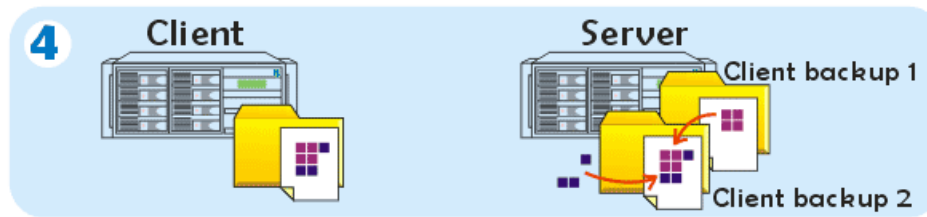
2. Some modifications were made to the client data.



3. A SmartSync backup task starts. It transfers the modified blocks to the SmartSync backup server.



4. The SmartSync backup server reconstructs data and creates a new backup, Client backup 2, based on the received blocks and the Client backup 1.



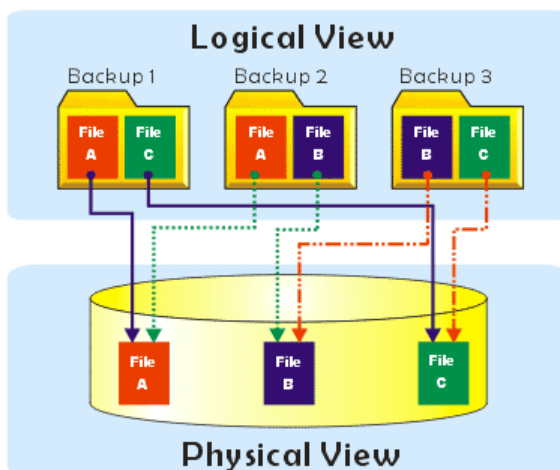
You can see that a new backup is created by transferring several data blocks only.

4. The Idea of Saving Disk Space

- Keeping only one copy of files

Since SmartSync Backup always makes full backups and keeps them online, some might criticize that it takes much disk space. It is not unusual that many files may stay un-modified from backup to backup. It wastes disk space if we keep copies of the same files in different backup versions.

The SmartSync Backup software resolves this issue by using hard links. It only keeps one copy of each file on hard disks. At the file-system level, different backup versions use hard links to point to the files. Please see the following figure.



- Freeing up disk space automatically

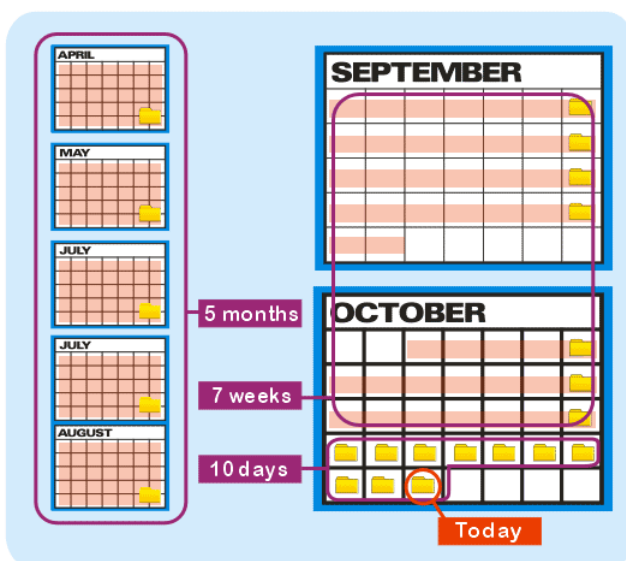
SmartSync Backup keeps all backups in the backup server. Without proper management, the backup server might run out of space unexpectedly. In addition to manually deleting obsolete backups, SmartSync Backup implements a backup deletion mechanism based on a pre-defined policy.

This backup version control is inspired by the GFS (Godfather-Father-Son) tape rotation scheme commonly used on tape backups. We call it advanced GFS media rotation scheme. Instead of re-using tapes as in tape rotation, here it frees hard disk space for future use. When a new backup version is created, it checks and deletes obsolete backups automatically according to the rules.

The rules are described as below. X, Y and Z are user-defined numbers.

- (1) For the last X days, it keeps one backup version each day. If there are two or more backup versions on one day, only the newest backup will remain. Others will be deleted.
- (2) For the Y weeks prior to the X days, it keeps one backup per week – the newest backup in each week.
- (3) For the Z months prior to the Y weeks, it keeps one backup per month – the newest backup in each month.

If X=10, Y=7, Z=5, backup versions which remain on the backup server will be:





5. SmartSync Backup vs. Tape Backup

- Unit costs of backup media

Tape backups used to claim low cost per GB. However, hard disks are getting more attractive in terms of prices. The prices below are listed on <http://shopper.cnet.com>, surveyed in Oct. 2003.

	SuperDLT tape 110GB raw capacity	LTO tape 100GB raw capacity	Maxtor DiamondMax Plus 9 200GB, 7200rpm
Prices	around US\$100	around US\$50	around US\$200
Cost per GB	US\$0.9	US\$0.5	US\$1

The unit cost of HDDs is almost the same as that of SuperDLT tapes. Considering the maintenance and management costs, tapes are not always of the lowest costs.

- Backup time

As described above, the SmartSync Backup software only sends modified blocks of files during backups. It saves much backup time because there are usually partial changes to files.

In case of the first-time backup, the SmartSync Backup software has to send all data for a baseline backup. In a real world test, SmartSync can transfer data at 14MB/second, about 840MB/minute. For reference, a Quantum SuperDLT tape drive transfers data at 11MB/second, or 660MB/minute.

- Online, offline and offsite data backups

SmartSync Backup keeps all backups online, while tapes are always off-line. You cannot open files on tape directly. You must restore them in order to read files.

However, there are situations where you must use tapes. When you want to keep backup media offsite to prevent some human errors or natural accidents which cause permanent data loss or system damages, tape backups are the best choice. SmartSync Backup cannot help.



- Concurrent backups of multiple clients

Tape media allow sequential access only. It is not possible to write multiple backups to one tape media at the same time. SmartSync Backup uses hard disks as backup media, which allow concurrent access. Multiple SmartSync backup tasks can be performed at the same time. In fact, SmartSync backup servers allow up to 8 concurrent backups to one sync point.

6. SmartSync Backup Benefits

The SmartSync backup differs from traditional tape backups in many ways. It does not have to replace tape backups, but to complement them. In cases where you prefer online backup data, SmartSync Backup is the choice. The benefits include:

- Full backups with less backup time than incremental backups
- Fast backup with differential block transfers
- All backups are online, making it easy to restore
- All communication and data transfers are encrypted and secure
- Backup version control