

Backup with synchronization/ replication

Peer-to-peer synchronization and replication software can augment and simplify existing data backup and retrieval systems.

BY PAUL MARSALA
May, 2001

According to International Data Corp. (IDC), demand for storage capacity will increase more than tenfold from 2000 to 2003. Expanding storage space at this accelerated pace is expensive. Implementation of a backup process that preserves investment in existing infrastructure while providing scalability, reliability, and high performance can provide significant cost savings and operating efficiencies.

This article introduces Local Area Backup Software (LABS), peer-to-peer synchronization and replication software that leverages existing networked PCs and servers. The software manages file backup and synchronization routines. Hardware upgrades such as the addition of network-attached storage (NAS) can increase flexibility and capacity but are not required. (For an example of how the software works, see the sidebar on p. 54.)

Many data backup systems only protect information stored on servers, and much of this information is collected daily or less frequently. Work done during the previous 24 hours-or longer-is at risk of permanent loss, and important data on user PCs and workstations may not be backed up at all.

Synchronization/replication

Synchronization and replication software automatically synchronizes and replicates files and folders among laptops, desktops, servers, and any shared storage resource on the LAN.

Synchronization is the updating of any existing files or addition of new files in a target folder using a source folder as the guide. Upon completion, the target folder contains all of the files in the source folder, with matching content and time stamps.

Replication is nearly identical to synchronization, except files in the target folder that do not have a matching source file will be deleted.

Synchronization and replication procedures generally include

- Automated launch upon logon or system start;
- Creating source folders on participating PCs;
- Establishing target folders on file servers, NAS devices, or other attached storage; and
- Scanning target and source files and folders to identify differences, and synchronization or replication of source and target by acting on differences.

The synchronization and replication software prevents loss of data by checking file dates and times to ensure newer information is not overwritten.

Peer-to-peer software

While providing the functionality described above, peer-to-peer synchronization and replication software (LABS) conducts live comparisons between the target and source lists that reside on different machines. Unlike other systems that rely either on client-server or database-driven ("store or forward") architectures, peer-to-peer synchronization and replication software operates in isolation from and has no impact on other applications or on database and client-server logic and processes. Data is not stored in a local database and no installation of server software is required.

LABS operates the same on servers as on workstations. While synchronization and replication can take place from server to server to server, it can also be established among servers, desktops, laptops, and NAS devices (e.g., server to desktop or desktop to laptop).

LABS software must be installed on each participating device. A common profile can be configured for all who utilize users' logon names or assigned computer names as part of the master path (e.g., \\Server\DriveLetter\Users\%UserName%). This common profile can reside on the primary storage resource. A login script can be run from such a resource and be accessed by all users when they log on, thereby reducing the need to administer each installation.

Command line, logon script, system batch file, detailed shortcut, and shell command execution are available. Some LABS software can be set up to run as an NT service.

Creating backup sets

Files in source data directories can be monitored according to schedule and/or in real-time. When a file is modified, added, or deleted, the change is propagated to the storage resource either within seconds of the change, if monitored in real-time, or within seconds after the requested schedule.

Backup routines vary to match individual requirements. Administrators and users with rights access can modify the scope of the backup. In many installations, users will have no access to the file synchronization and replication software or the backup processes.

Upon initial user logon, for example, a complete scan of selected source folders could be compared to target folders to ensure all changes have been backed up. This will initiate synchronization that may be necessitated if, for example, the local PC was used while the network was unavailable or a laptop was used offline. Subsequent backups could be only of file changes (incremental).

LABS does not eliminate traditional backup solutions. For example, it works with traditional methods for periodic backup to removable media (e.g., tape) for off-site storage. While tape backup and restoration are burdensome and time-consuming, such limitations are mitigated when tapes are used in conjunction with LABS solely for the extra protection afforded by off-site storage.

Electronic vaulting

This backup approach also works with "electronic vaulting," in which both current and archived data can be maintained in a safe and remote environment such as an Internet data center.

Electronic vaulting replaces the multi-day practice of trucking backup tapes to safe centers and bringing them back to restore data in case of disaster.

Data can be replicated to a target on the LAN/WAN and then subsequently replicated to either a remote site or to servers at a rented location such as an Internet data center or a disaster-recovery hot site. This strategy ensures data is always available and current in at least two physical locations. Data updates and recoveries

can transmit across private networks or the Internet using Virtual Private Network (VPN) technology, which adds encryption and security, thereby creating a private connection over public networks.

Paul Marsala is president of Peer Software (www.peersoftware.com) in Hauppauge, NY.

Potential benefits

- Automated or unattended backup in real-time from any client PC on the network
 - Easy-to-use, network-based file backup of user data
 - Minimal network overhead requiring only file storage capacity on storage resources
 - Data kept "inside the firewall"
 - Data moved either real-time or incrementally throughout the day
 - Continuous on-site access to all stored data from original directory structures
 - Continuous access to data if additional storage resource is unavailable
 - Central availability of data from numerous machines based on logon name
 - Continuous availability of network bandwidth
 - Leveraging of existing network security infrastructure
-

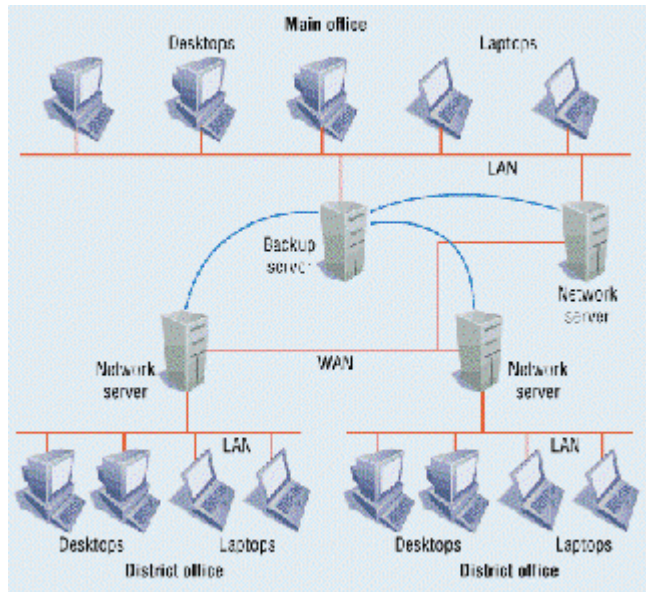
A practical example

Consider a company with a main office and two district offices. The company wants to synchronize all user data every time a user creates or changes a file or directory on PCs. The data must be available from different locations on each LAN because many employees work from a laptop and desktop within each office. The company also requires that newer files on the target be brought back to the source folders on PCs. In addition, the company wants to centralize distribution of application patches and upgrades to users at initial login at each office.

The primary goal is to automate, standardize, and centralize backup among the offices.

One server in each office provides file sharing and is connected to the WAN. The main office also has a backup server with removable media.

Synchronization and replication software is installed on all participating devices. A login script copies source-target synchronization/replication instructions from each network server to each PC on the LAN, and the synchronization/replication profile is launched from each PC. (If the server goes down, the synchronization/replication software can continue to run locally.)



Selected source folders on local machines are scanned on desktops and laptops upon login to create an accurate synchronization on the target (in this case, the local network server). Subsequently, real-time bidirectional synchronization is persistently deployed throughout the day, so that changes in data are continually moved between the source and the target as changed files are saved and closed. Any item in either the source or the target folders that has a creation date newer than the last time synchronization was run will be added to the corresponding source or target folder.

Users are assigned security permissions to the user data directory on the target, thereby making such data available on all machines throughout the LAN, even if an employee's PC fails. In addition, application patches and upgrades are distributed from the network servers to the PCs in each office through replication of the source directory on the server on which these files are initially collected to target directories throughout the LAN.

This solution enables persistent local backup of changed files in which PC data is recoverable on demand without the traditional time lag that threatens the loss of the most recent data.

This approach can also be a key component in enabling efficient centralized backups over a WAN. Using the same peer-to-peer synchronization and replication software, the synchronized directories on each network server—which were the targets from local PC source directories—can be replicated as source directories to target directories on the WAN backup server. Following an initial scan to synchronize user data directories on the local network servers with directories on the central backup server, changes in the data directories will be replicated to the backup server in real-time. Data is immediately recoverable in case of a machine failure (e.g., if a local network server is unavailable). Periodically, backup data is copied to removable media on the central server and moved off-site. System-wide backup is more automated, reliable, and persistent.

Synchronization/replication software features and options

- Simple setup, modification and administration, including "silent" enterprise-wide installations
- Pre- and post-process scripting integration
- Fault tolerance with real-time bidirectional synchronization between storage resources
- FTP site file synchronization that supports proxy servers or firewalls

- Instant administrative and user reports, including e-mail notification of synchronization/replication activity
- File inclusion or exclusion within synchronization according to standard DOS-type wild cards, sub-string comparisons, file attribute settings, or date and size
- Revisioning, which enables management of additional backup copies of files, giving users the ability to step back to previous saves of a document
- Immediate restoration of accidentally deleted files
- Bandwidth throttling in conjunction with buffering
- Scheduled, manual, end-to-end, and real-time methods of deployment in varying combinations. (Real-time deployment scans source folders for changes and synchronizes specified target folders as changes occur.)
- Real-time file monitoring and updating for real-time stable state detection and synchronization of individual changes to files. Only changes in source files are acted upon. There generally is no scanning of the entire folder in which the change took place, thereby reducing traffic.
- Hundreds of source-target folder combinations (filters) can be created and stored (in profiles). While the synchronization of large numbers of folder combinations occurs at once, each combination can have a unique set of rules (e.g., exceptions) and method of deployment (e.g., real-time).
- Multi-threaded filters, allowing filters to run simultaneously through separate threads
- Quick synchronization features that provide easy and persistent on-the-fly synchronization of a work folder through Windows Explorer context menus