

Speed versus Accuracy in Backup and Restore

By Don Chouinard

The purpose of backing up business-critical data is to enable an accurate restore when data loss occurs. Backups have to run as quickly as possible in order to protect the multiple gigabytes of data that reside on typical networks of servers, desktops, and notebooks. Ideally, backup software would offer both fast backups and 100% accurate restores.

Unfortunately, backup administrators have had to compromise between speed and accuracy because traditional backup software forces a choice between slow backups with accurate restores or fast backups with inaccurate restores.

Full Backups: Accuracy without Speed

With traditional backup software, time-consuming full backups are the only way to ensure completely accurate restores. A full backup copies data from all of the computers selected for backup onto the backup media. Restores place all of the data from the backup media onto the selected computers.

However, full backups take too long to run because large amounts of information have to be transferred. Often computers that are low on the backup list are not backed up if there is not enough time to copy their data. To correct this problem, backup administrators must carefully review backup logs each day and write new scripts to backup the specific files, folders, and computers that were missed during the previous backup window.

Some computers might still remain unprotected despite these additional steps. As a result, organizations often settle for protecting only a fraction of their computers rather than dealing with the complexity and manual effort involved in trying to back up too much data in too little time.

Incremental and Differential Backups: Speed without Accuracy

Typically, incremental or differential backups are used as a concession to the time constraints that exist in most businesses. Fast backups are achieved by running an initial full backup followed by periodic incremental or differential backups.

Incremental backups contain only files that have changed since the previous full or incremental backup. This creates a series of incremental backup media sets, each containing the information that changed since the previous backup.

Differential backups contain only files that have changed since the initial full backup. Because the amount of changed data increases over time, differential backups consume increasingly more time and media, until it becomes more efficient to run another full backup and begin the differential backup cycle again.

Because incremental and differential backups copy only files that are new or have changed, they are faster than full backups and consume less media space. They can run on a regular schedule within a reasonable amount of time, allowing a company to safeguard the most recent information on all computers.

Although incremental and differential backups are faster than full backups, restoring from them is more complex and the restore is always flawed when restoring more than a single file. During a restore, the files and folders from the full backup are returned to the hard drive, followed by data from the series of incremental backups or from the latest differential backup. This method of performing restores is flawed because the process restores all the contents of the full backup plus all the contents of the required incremental or differential backups, including previously deleted, moved, or renamed files and folders.

How flawed restores occur

To understand how flawed restores occur, consider the following example of a backup cycle that begins with a full backup and continues with daily incremental or differential backups.

Figure 1 File changes during a typical backup schedule.

Day 1	Day 2	Day 3	Day 4	Day 5
File A	File A renamed to File F	File F	File F	File F deleted
File B	File B deleted			
File C	File C renamed to File G	File G	File G	File G
File D	File D moved to new location	File D (new location)	File D deleted	
File E	File E	File E	File E	File E

Figure 2 Flawed restore to Day 3 for the computer files shown in Figure 1.

Files from Day 1 FULL backup		Files from Day 3 INCREMENTAL / DIFFERENTIAL backup		Hard Drive after a restore to Day 3
File A		File F		File A – wrong
File B	+		=	File F
File C		File G		File B – wrong
File D		File D (new location)		File C – wrong
File E				File G
				File D – wrong
				File D (new location)
				File E

After the restore, the user discovers that File B is back, even though it was deleted. File A and File C are back, even though they were renamed. And File D appears both in its original location and in its new location. The same kinds of problems can occur for restored folders as well.

Without an accurate restore, the time saved by running incremental or differential backups is clearly not a benefit, because it is difficult or even impossible to clean up the hard drives after the flawed restore.

Progressive Backup: Both Speed and Accuracy

Another type of backup strategy, Progressive Backup, allows for speedy incremental backups, but uses a different technique to provide accurate restores that do not place unwanted deleted, moved, or renamed files and folders back onto a drive.

Before a daily backup is performed, Progressive Backups take a list of all of the files and folders on the hard drives being backed up. When a restore takes place, that list is used as a guide so that only the correct files and folders are restored onto the hard drives. The result is an accurate restore. No unwanted files or folders that were previously deleted, renamed, or moved are mistakenly placed onto the hard drives.

Figure 3 File changes during a typical backup schedule.

Day 1	Day 2	Day 3	Day 4	Day 5
File A	File A renamed to File F	File F	File F	File F deleted
File B	File B deleted			
File C	File C renamed to File G	File G	File G	File G
File D	File D moved to new location	File D (new location)	File D deleted	
File E	File E	File E	File E	File E

Figure 4 A 100% accurate restore to Day 3 for the computer files in Figure 3.

Required files from Day 1 FULL backup						
File E						

+

Required files from Day 2 & Day 3 INCREMENTAL backups
File F
File G
File D (new location)

=

Hard Drive after a restore to Day 3
File F
File G
File D (new location)
File E

Progressive Backup provides the speed and flexibility of daily incremental or differential backups while providing the restore accuracy of full backups. It also enables fast backups that can run more frequently and protect all business-critical data on the selected computers during limited backup windows. It allows businesses to protect more computers, keeping information complete and up-to-date without requiring complicated backup schedules.

A progressive backup can automatically recognize identical files on all of the computers on the network. It then transfers only one copy of each unique file to the backup media, saving considerable space and time. For example, if 20 desktop PCs running Windows XP are being completely backed up, only one copy of identical operating system files will be moved to the backup media. If identical copies of a 20 MB PowerPoint presentation exist on 10 different computers, only one copy of the presentation will be backed up.

Speed and Accuracy

A fundamental problem with backup and restore involves compromising between fast backups and accurate restores. Full backups provide accurate restores, but take too long to run. Incremental or differential backups save time, but result in flawed restores that mistakenly retrieve files and folders that have been previously deleted, renamed, or moved. As a result, traditional backup software forces you to choose between speed and accuracy.

With Progressive Backups, there is no compromise between speed and accuracy. Combining speed and accuracy provides a range of benefits, including complete protection for data, confidence in the quality of the restores, and minimal administrative headaches for an IT department.

Progressive Backup delivers the best qualities of incremental, differential, and full backups – without sharing their significant problems. Various versions of software that uses Progressive Backup technology are tailored for small and medium businesses, departments within larger corporations, or for home and small offices.

About 1100 words

Don Chouinard *is the director of product management at Dantz Development Corp. (Orinda, Calif.)* He can be reached at donc@dantz.com