

Preparing for Disaster: Backup Scenarios That Work

by Tom Mills

simpliTek Corporation - Technology Consultant to Small Business

Executive Summary

With some simple and inexpensive steps, any small business can protect itself from reasonably foreseeable mistakes, failures and disasters that can render computer information inaccessible.

A good backup strategy is one that:

1. Protects against every day operator error.
2. Protects against hardware or software failure
3. Protects against malicious destruction, theft or natural disaster
4. Provides for a mechanism to be up and running again quickly after any kind of mistake, failure or disaster, with all of the data accessible on a computer that has all the right software loaded.
5. Is convenient enough that you will use it and stick to it.

If you don't have such a strategy in place, ask yourself what you would do to continue your business after one of the occurrences, which, though perhaps unlikely, are certainly reasonably foreseeable.

First and Foremost: Data Security

Losing important data is probably the one of the events that concerns system administrators and business owners the most, and for good reason. The major advantages of using a computer system in the first place are constantly at risk if files are lost, corrupted or sabotaged. There is a saying in the computer business that applies here:

If you only have one copy of something, someday you will have no copies.

Before a strategy can be developed to prevent the loss of important business-critical data, it is important to investigate the possible ways in which data can be lost. The strategy for dealing with each of these possibilities is very different.

Computer data can be lost in one of four basic ways:

- The disk drive on which the data resides can crash or become corrupted, even though the computer itself may or may not continue to function.
 - Data can be accidentally damaged or erased through the error of a well-meaning employee, or through the malfunction of an application software package.
 - Data can be damaged or erased by the deliberate action of an intruder such as a hacker, or by a disgruntled employee.
-

- A disaster can result in the physical destruction or loss of the computer on which the data resides.

When one considers all the possibilities, most data losses come down to one of four events:

- System failure
- Operator error
- Malicious data destruction
- Accidental computer destruction or theft

There is one more scenario that deserves mentioning, and that is **multiple computer destruction**. This can occur as a result of a virus that spreads through the office, or a power surge that takes out multiple computers and causes disk damage. This scenario is similar to the destruction of the entire office through a fire or other disaster, or the theft of one or more machines, but the main consideration here is the fact that data backed up from one machine to another across a network is not necessarily secure even if the likelihood of two disk drives failing at the exact same time is otherwise astronomical.

The strategies for dealing **conveniently** with these modes of destruction are varied. The reason is that if a backup methodology is not convenient, it **will not be used** no matter how critical the need. Thus the challenge is to deal with these failure modes while not impacting day-to-day operations or inflicting an onerous burden on office personnel, which results in the failure of manual processes.

Backup Methodologies

Let's look at ways that data can be backed up, and analyze the strengths and weaknesses of each method.

1. Real time file-by-file copying. A program such as Second Copy can be used to make a backup copy of each file as it is activated or saved. Also, many software applications have this capability built in. This is the best strategy to deal with operator error, because once the operator realizes that a file has been lost or damaged, the backup copy can usually be retrieved easily and quickly. The last copy of the file can be activated from its backup location, and a **Save As . . .** operation can be performed immediately to overwrite the damaged data or recreate the accidentally deleted file in its proper location.

The most important factor in the success of this methodology is to insure that the backup location is on a separate disk drive from the original data store. This will prevent the loss of this backup in the event of a disk failure. Disk drives occasionally fail without the slightest warning.

2. Automated backup. An automated backup routine runs at a set interval, usually every day, and backs up all the data at once. This process is greatly facilitated by having all business critical data stored in a central location such as a file server, instead of being stored on individual workstations. (However, with drives mounted across a network, a central backup routine can still pull files from individual machines and back them up if necessary.) Windows 2000 Professional and Windows XP have a built-in automated backup routine that works fairly well. In addition, there are third party packages such as Retrospect, Backup Now!, Backup My Computer et.al. which offer other functionality. Also, backing up your files to the Internet through a backup service is now a very viable mechanism for automated backup.

There are various modes of automated backup. A **complete** backup backs up each file every day regardless of whether that file was unchanged from yesterday. This mode takes the longest each day, but offers the advantage of the quickest reconstruction of the data set in the event of a failure.

An **incremental** backup looks at each file to see if it has been modified, and only backs up files that have changed since the previous backup. This method is much quicker on a daily basis because only a few files out of a company's entire data store are actually changed every day and thus need to be backed up. For very large companies with many, many employees changing many, many files every day, this is about the only practical methodology. However, upon the need to reconstruct the data, this approach is quite a bit more cumbersome. Reconstruction involves reloading the last complete backup, followed by each incremental backup in the order in which they were made. If any one backup medium is faulty, files that were backed up that day can never be restored and all the work done that day is lost.

For small companies who have a manageable set of data that can be fully backed up in a reasonable time, complete backups are far more convenient. However, it is important to note that the only way a complete backup can get every file is for the backup to be conducted after all files have been closed and stored on the server. The backup can be set to run in the middle of the night, or the last employee to leave can kick off the backup routine manually and wait for it to complete in order to remove the media and leave. As an alternative to this, the backup routine can be set to run at, say, 4:00p each day and back up whatever files it can get to at that time. Any files still open on the server because they are activated on a client computer would not be backed up under this scenario.

The question then becomes: to where does the routine back up the files? Here are some possibilities (assuming that all data is stored on a central server):

Data Still At Risk to the Following Failure Modes:						
If You Back Up Files To:	Disk Failure	Server Destruction	Multiple Computer Destruction	Malicious Destruction	Theft of One or More Computers	Office Destruction
The same disk drive	X	X	X	X	X	X
Another disk of the same server		X	X	X	X	X
Another disk on the network			X	X	?	X
Tape or removable media left in the drive		X	X	X	X	X
Tape or removable media taken out of the drive				X	?	X
Tape or removable media taken out of the office						
Internet Storage						

It is clear from this table that by far the best way to preserve backed up data is to physically remove the backup medium from the office each day. For a local backup process, (tape, CD-R

disk, or removable disk such as a portable USB drive) the backup medium must be taken offsite. Many companies accomplish this through a rolling media swap.

Here's an example:

Data is backed up every day to a tape drive. The office has tapes labeled Monday through Friday. On Monday at the end of the day all data is backed up to the Monday tape. A trusted employee takes the Monday tape home with them but **does not** bring it back to the office the next day. On Tuesday all the data is backed up to the Tuesday tape and that disk is sent home with a trusted employee. The cycle repeats itself but whoever takes home the disk on any given day **keeps it at home** for a minimum of two days before bringing it back to the office. That way at any given time, there is always a data tape off premises that is no more than one day old. As long as the employee who took home the Monday disk brings it back no later than the following Monday morning, this scheme works with 5 tapes.

If there is a natural disaster during the work day, such as a catastrophic fire, all employees can exit the building immediately without giving a thought to backup, because everyone knows that the only data that will be lost is any data that was worked on that day. Data up to the close of business the previous day is preserved at the home of whomever took home the previous day's backup medium.

With the combination of a real-time file copy program to handle operator error and a daily comprehensive backup to a medium that is removed from the site, the scenarios under which data can be lost are reduced to statistically insignificant occurrences.

It should be stressed, however, that there is no such thing as a bulletproof backup policy and methodology. No matter how unlikely a scenario is, as long as it remains even a remote possibility, data can never be totally secure.

That is why business owners have to come to terms with what level of risk they are willing to live with as compared to what level of business disruption for backup purposes they are willing to live with. The total risk will always be related to the level of disruption that the backup method causes. There are things that could be done which could almost guarantee that data would not be lost, but there would instead be a significant loss in productivity as employees spend the time required to safeguard files.

New Technology: Backing Up to the Internet

A new alternative has arisen for comprehensive backup which, though not inexpensive, elegantly handles almost all of the shortcomings of onsite manual or automatic processes. Companies can now subscribe to a **service**, with a monthly fee, which provides the capability to upload data files from their computer(s) to a server elsewhere, accessed through the Internet. When you sign up for a service like this, they provide you with proprietary software that you load on your computer. Then you set up the software to backup your files daily. At the time of the backup, the software starts up, logs into a remote computer across the Internet, and then transfers your data to that server for safekeeping.

Of course, security of that data is of the utmost importance, since now your data is being backed up to a location which is not under your direct control. Most of the companies who provide this service work very hard to make you confident that your data is safe with them. Most **encrypt** the data as it is being transferred up to their servers, so that the data stream cannot be intercepted and compromised by a hacker. Most take stringent steps to safeguard your data from being viewed

or copied at their site. All reputable Internet storage companies replicate your data onto a server distant from their main server, so that if they are subject to a disaster of some kind, they still have a copy of your data safely stored somewhere else.

Most backups to the Internet are incremental backups; that is, they only back up what has changed since the last backup. The reason for this is to minimize the actual amount of data being sent across the Internet. It is pointless to take the time and bandwidth to send up an entire file if that file hasn't changed since the last time it was sent. But these services do something with the data that local backup solutions do not: they **integrate** the changes with the original file so that if you have to restore one file or a whole disk full of files, you end up with the convenience of a full backup instead of having to first load the last full backup and then sequentially load all the incremental backups since that time.

The process that they use for this is intriguing. Let's say that yesterday you created a file called **proposal.doc**. That night, since that is a new file, the entire file would be shipped up to the backup server. Then today, you come in and do some more work on that file. Tonight, when the backup runs, the software looks inside the file to see what you actually changed, and then it uploads only the changes. Then at the remote server, the changes are applied to the original file from yesterday, resulting in a file that is now exactly the same as the what you have currently. It takes a lot less network traffic to ship up only the portions of the file that have changed, and the reassembly at the other end maximizes your convenience if you need to get the file back.

Some online backup systems augment your security by storing multiple versions of each file in order to guard against accidental or malicious destruction. For example, let's say that last night the system backed up your **proposal.doc**, and then today you come in and do something to the file that damages or destroys it, and before you realize it, you save the file with the bad changes. And let's further say that you don't discover until tomorrow that the file is damaged. By that time, the system will have uploaded the bad changes that you made and the latest file will be just as damaged as your copy. **But, if they saved several versions**, then you could simply return to the last version of the file that was undamaged, and restore that version. Backup services vary in the number of versions of each file they save: some save the last 3 versions, some save the last 7 versions, and some so **no previous versions**. It is important to understand this feature as you evaluate various services.

The only real downside to Internet backup is that it is a service and not a product, and thus you have to pay for it forever. Prices range from a low of \$30 per month for a <100MB backup to hundreds of dollars per month for multiple gigabytes of storage. Also, prices vary according to the quality of the security being offered - better companies charge more. Just like with anything else.

The key advantage is: **set it and forget it**. Once the system is set up properly, you are protected as long as you pay your backup bill. You are protected from hardware failures, tapes that never really get anything written to them, tapes that are lost, tapes that are stolen along with the computer that made them, etc. No more shuttling tapes, CDs or portable disks offsite. For some customers, the ease and reliability of online storage far outweighs the cost disadvantage.

Security Against Business Disruption Due to Disk Failure

If you have a backup strategy in place where you are dutifully backing up all of your business-critical data, then in the event of a disk failure, the recovery process would consist of the following:

1. Determine if the disk is salvageable. If so, format the disk (erase it completely) and then rebuild it from your backups. This usually involves reloading the operating system and software from scratch because it is difficult if not impossible to restore an entire computer successfully from even a full backup.
2. If the disk is not salvageable, physically replace the disk with one of the same or greater capacity. (You can guard against delays in the process by keeping a replacement drive in your inventory, ready to switch in, but very few companies actually do this.)
3. Reload the operating system and applications, and then restore the data from the backup medium.

At [simpliTek](#), we estimate this to be approximately an 8 hour process - in other words, by the time you figure out that your drive is corrupted or has failed, get a new drive installed and get the drive back to where you were before the failure, your computer will be down for at least 8 business hours. This is under the ideal circumstances: service personnel readily available, replacement drive readily obtainable, and the backup medium readily restorable.

For some businesses, even 8 hours of downtime could be catastrophic. If you use your computer to run a fast-paced, data-intensive business (such as a tax preparer on April 14th!), even 8 hours of downtime could be fatal to your business.

Fortunately, there is a hardware solution to this which is easy and affordable to install and virtually guarantees that you will be back up and running within minutes of a sudden disk failure. The technology is called **RAID**, and it has nothing to do with insect repellent.

RAID stands for **R**edundant **A**rray of **I**nexpensive **D**isks. There are many ways to implement RAID, but for the purposes of this discussion we are going to focus on a configuration called RAID 1, which involves **disk mirroring**. In this configuration, you outfit your computer with not one but two disks (preferably identical disks) and a special circuit board called a RAID card, which goes inside your computer in one of its expansion slots (most computers have at least one expansion slot open).

What the RAID card does is to take every single disk write or change, and perform that on both disks more or less simultaneously. This results in two disks whose contents are kept perfectly identical, or **mirrored**. This generally does not slow down your computer in any way because the RAID card sees to it that both disks are written to at the same time. One disk is your main, active disk and the other is your standby, or mirrored disk.

Then, if one disk fails, then instead of having to scramble to find a replacement drive, get it installed, and get the data back onto it, you simply go in through the software and switch the disks and make the mirrored disk become the main disk, and you are back up and running in minutes. (With some RAID installations, you don't even have to do that much - the system automatically switches over to the other disk and flashes a message on the screen informing you that "redundancy has been lost", but the computer never goes down, doesn't even have to reboot.) Then, at a time more convenient for you (and with less drop-dead urgency) you can replace the failed drive and get it back into the mirrored configuration.

This probably sounds really expensive and complicated, doesn't it? But it's really very affordable. The entire configuration including the RAID card, the extra disk and the service time required to install it for most computers is in the \$300 to \$400 range. In fact, it is so affordable that we assess

the mission criticality of every customer's computer and suggest this as a hedge against the inevitable downtime associated with a disk failure.

How do you determine if RAID is for you? Simply think about how much it would cost you in lost sales or affected productivity if your computer were to be down for 8 hours. If you would pay \$400 to make sure that never happened to you, then it is a good investment. It is an especially good investment if you have a central computer where everyone in your office stores all their data centrally - if that disk were to go down, it could idle everyone in your office. We at [simpliTek](#) have a RAID 1 configuration on our main computers because it is so easy and inexpensive to implement.

Final note on RAID: this configuration is **not** a comprehensive backup strategy because anything that destroys the entire computer will destroy both disks. RAID 1 should **not** be the essence of your backup strategy, but rather a tool to get you up and running again quickly in the event of a disk failure. Redundancy and backup are **two totally different things** that do not substitute for each other.

Security Against Catastrophic Business Disruption

Let's say that you run an information-related business – that is, the main "product" or service you render to your customers is the processing of data. This would apply to a law office, accountant, realtor, mortgage company, engineer, architect, etc. – the kinds of organizations where the product or service could, under drastic circumstances, be conducted from anywhere and on any computer as long as you could access and process the right information.

If you run such a business and you have developed a workable backup strategy for your data, then even if your business burns to the ground, you still have the work product of your business – the data you have produced or are working on for your customers.

But what good is the data if you don't have a computer with which to operate on it?

Chances are in your business you use a lot of specialized software that has to be installed using license keys or web activation. If all of your computers go up in smoke, even if you have the data safely backed up you could still find yourself in very poor shape because of your inability to use the data you have.

For this reason, it is very prudent to give some thought to a mechanism through which you could have a new computer up and running very quickly with all of your unique and licensed software ready to run, without having to scramble to get in touch with the software vendors, prove that you have a license, obtain replacement disks, etc.

The best way to do this, obviously, would be to build another computer with an identical set of software to that which you use in your business, and simply keep that computer safely stored somewhere else. The presence of even one properly configured computer which is ready to run quickly after a disaster can make the difference between business survival and business failure for companies in the information economy.

Conclusions

No one likes to think about disasters that can threaten a business. But with businesses becoming ever more reliant on computer technology for even the most basic business processes, prudent steps taken to secure valuable business data and a platform on which to access that data can make a huge difference in prospects for business continuity.

With the total investment in all the hardware, software and consulting to implement a good backup strategy falling somewhere in the \$500 to \$2000 range, a small business owner only needs to do a quick mental calculation of how many days revenue that represents, before realizing that investment in a good backup strategy is the best hedge against the unknown that can be made.

Think about it today. **Make the decision to implement it. Get it done.**
