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 the event 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.

Before a strategy can be developed to prevent the loss of important business-critical data, it is important to understand 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 several 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.
- Data can be destroyed by malware such as a virus or spyware
- A computer can be stolen, or a disaster can result in the physical destruction or loss of the machine.

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

- System failure or disk failure
- Operator error
- Malicious data destruction by a disgruntled employee
- Data destruction by malware or by a hacker
- Destruction or theft of the computer(s)

Backup Methodologies

The various modes and mechanisms of backup offer varying degrees of effectiveness in guarding against these potential losses.

Today most companies use one of the following modes of backup:

- Backing up files to a different folder on the same disk drive
- Backing up data files to a different internal disk in the same computer
- Backing up data files to another computer's disk over the network
- Backing up data files to an external disk drive (usually a USB drive)
- Backing up data files to an Internet backup service such as Carbonite, Mozy or many others

The most frequently used method today is backing up to an external medium such as a disk drive or a flash drive. What is actually done to safeguard this medium becomes a new variable in the effectiveness of the backup. Some companies connect an external disk drive to a computer, leave it resting on top of the computer and never move it. Some companies secure the disk within cable length of the computer, by bolting it to something immovable or placing it in a safe enclosure with the cable still attached. This type of handling of external disks protects against certain threats but not against others. For example, if someone breaks into your office to steal your computer, they may or may not also steal your backup drive. If it is bolted to a table or stored inside a safe, the thief will probably not bother with stealing it.

On the other hand, if there is a fire or flood that destroys the computer, chances are it will destroy the external drive as well as the computer. Some people get a false sense of security from placing their external disk in a fire safe. However, a fire safe is designed to keep the contents below 400º, which is the temperature at which paper starts to burn up. However, by 400º the external disk drive (or any other computer medium for that matter, including CDs, magnetic tapes, flash drives, etc.) has long since burned to a crisp. The only way to protect computer media is with a **media safe**, designed to keep the temperature down to 125º. These safes are much more expensive and the time rating is always much shorter – usually 30 minutes or less. Companies who expect even a media safe to protect them from a bad fire are taking a real risk.

	You are Protected			
If You Back Up Files To:	**Disk Failure or Server Destruction**	**Operator Error**	**Malicious Destruction**	**Theft or Destruction of Computers**
The same disk drive	✗	✓	✗	✗
Another internal disk of the same server	✗	✓	✗	✗
A disk in another computer on the network	✓	✓	✗	✗
External Disk Drive left onsite	✓	✗	✗	✗
External Disk Drive taken offsite	✓	✓	✓	✓
Internet Storage	✓	✓	✓	✓

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

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. Windows XP and Windows 7 both have a built-in automated backup routine that works fairly well. (The built in backup software with Vista has so many limitations as to be virtually unusable.) 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 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):

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.

Here's an example:

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 whoever 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 elegantly handles almost all of the shortcomings of onsite manual or automatic processes. You can now subscribe to a **service**, with a monthly fee, which provides the capability to upload data files from your 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, and then the software backs up your files daily or even continuously.

Of course, security of that data is of the upmost 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 right on your own computer before it is 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.

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.** Then today you come in and do some more work on that file. When the backup runs next, 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 file you have currently. It takes a lot less network traffic to ship up only the portions of the file that have changed.

This service sounds perfect, so what if any are the downsides? There are actually some risks and issues that you should know about.

The first downside to Internet backup is that it is a service and not a product, and thus you have to pay for it forever. Prices have come down dramatically over the years – both MozyHome and Carbonite say they will back up an unlimited amount of data for about $55 per year. However, this number seems suspiciously low, so the long term financial viability of these companies has yet to be established.

Second, the amount of network "bandwidth" that these services will allocate to each user is extremely limited. When your file changes are being copied up to their "servers in the cloud", the files take a lot longer to transfer than your high speed Internet is capable of. The reason is that on their end, if they could take your files at the same rate your Internet provider is capable of sending them, and they were doing this across thousands or even millions of users, the capacity of their Internet connection would have to be **huge.** When it comes to Internet capacity or bandwidth, "huge" translates directly into "expensive".

So, they throttle every user's file uploads drastically. To you, this is really no big deal because you don't change that much data every day, and over a couple of hours in the middle of the night all your file changes can be copied up to their server easily. If you have a lot of data, the very first backup may take a really long time, maybe even weeks to complete. (That fact is never mentioned in the brochure!) But what is worse, if you have a catastrophic failure and you need to get all of your data back, at the normal bandwidth that they allocate to you, it also could take weeks to get all your data downloaded. If your disk drive dies and you need to get your 500GB of data downloaded onto a new drive, how would you like to wait three weeks to get it all back?

Some of the more enterprise quality online backup providers will solve this problem by quickly downloading all your data onto a disk drive and then overnighting the drive to you so that you can get your files back online quickly. But don't expect Carbonite or Mozy to do that for you for $55 per year.

Third, when you transfer your files to an online backup service, you have to be comfortable with the fact that you have really **lost control** over what happens to your data. Hopefully you can trust the

provider to take good care of your files, encrypt them, back them up on their end, and put into place policies and procedures that safeguard your data. However, if you look at possible worst case scenarios, some are very worrisome. A disgruntled or unethical employee could possibly access your data and use it in some way against your interests. Or, probably worse, a government agency like the IRS or even the FBI could come after your data for real or bogus reasons. **The best encryption scheme on the planet is no match for a subpoena.** Perhaps that is paranoid thinking, but one thing is very clear: there is a lot greater risk of something like that happening if you backup your files online than if you manage your backups yourself. If you want to keep your data out of someone's hands, you can always hide or destroy an external disk drive; you sacrifice that option when you backup online.

Finally, when the industry first started pushing online backups, the thinking was that you could "set it and forget it". At last – a backup scheme you could just turn on, it would always work and you didn't have to monitor it. Such thinking turned out to be highly **optimistic**. SimpliTek has seen many, many instances where online backups have simply stopped working, and the user was totally unaware. Here are just **some** of the things we have seen that have caused online backups to stop working:

- The subscription ran out, the auto-renewal failed because the credit card expired, and all the email warnings about the impending subscription expiration got stripped out by a spam filter
- Data on the computer was moved to a new location and the folders that need to be backed up were not updated with the service. So, the backup service was left backing up nothing and the folders that needed backing up were ignored. The same thing can happen if you create a new folder and start storing data in it – depending on how you do it, that folder probably does NOT automatically get added to the backup.
- The backup software decided (incorrectly, as it turned out) that the computer was no longer on the Internet. No Internet connection = no backup. Mozy is famous for this: there is actually a setting in Mozy where you tell it to try to back up even if it cannot detect a network connection! If the software decides you have a real or imagined network problem, we have seen the software stop working even though there was **nothing** wrong with the network at all.
- Passwords can get corrupted. All online backups are conducted based on a working login and password on the computer that matches a login and password on the backup server. But if the password gets corrupted either on the computer or on the server, then the software can perceive that you are no longer "authorized" to backup, and the backup fails. It's a simple matter to refresh the passwords on each end, **if** you know you have a problem.

So, the bottom line is that online backups solve some problems nicely, create other problems, and in the end are really no more reliable than local backups. The one thing they do best is to get your data offsite so that you don't have to worry any more about fire/flood/theft. But online backups have to be checked and monitored, just like disk-based backups. You can "set it" but you better not "forget it".

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 replaced 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.**