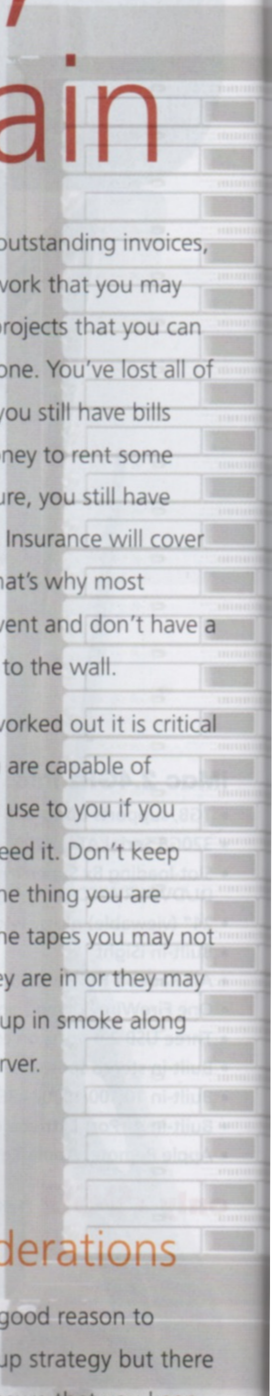
Bill Gallop

Under the Hood

Backup, backup, then backup again



A wise person once said that there are only two certainties in life, Death and Taxes. For those of us who spend much of our lives using computers there is another certainty to add, and that is Data Loss. If you haven't lost data yet then don't worry, you'll join those of us who have sooner or later. The Law of Sod is the main thing to abide by when planning for the worst.

Protect your data, it's not just part of your business it IS your business

Backup never seems important until you have lost data. Backup hardware can seem expensive and if all goes well you will never see the benefits so it can appear to be dead money. That's all well and good until something goes wrong and that workstation breaks down, you have a theft or worse still there is a fire and all your equipment is destroyed. Even something simple like a gas main failing can prevent you accessing your data.

Data is the lifeblood of any company, you can replace computers, you can move to another office and you can hire new people but if you lose access to your data the company will, in most cases, fail. Think for a moment what it would be like to start your company again from scratch. You've lost all records of your customers and suppliers, you have no idea of any outstanding invoices, no idea of outstanding orders, no work that you may have had in progress, no previous projects that you can rework so everything has to be redone. You've lost all of that but you still have staff to pay, you still have bills coming in, you have to find the money to rent some office space, equipment and furniture, you still have customers expecting you to deliver. Insurance will cover some things but not loss of data, that's why most businesses that have a disastrous event and don't have a proper backup solution in place go to the wall.

Once you have a backup strategy worked out it is critical that you test it and ensure that you are capable of restoring from it. A backup is of no use to you if you cannot get at the data when you need it. Don't keep your backup in the same place as the thing you are backing up. If you need to access the tapes you may not have access to the building that they are in or they may have gone up in smoke along with the server.

> "Even if you have a RAID attached to your server you still need to back it up"

Legal considerations

Planning for a disaster is one very good reason to ensure that you have a good backup strategy but there are also legal reasons for you to ensure that you keep certain data.

There are a number of regulatory frameworks that require that companies retain certain information for long periods of time. More and more business is done

electronically. Traditionally retaining data for regulatory reasons meant archiving paper copies of contracts, invoices, personnel records and so on but in many cases today there is no paper version. You may have emails between you and a customer that may constitute a contract.

Somehow you need to be able to manage all of that data and once you start storing data electronically you then need to consider how you retrieve it should you receive a request under the Freedom of Information Act or the Data Protection Act. The National Archives has estimated that perhaps 25% of all emails could be defined as documents that should be retained unaltered.

RAID is not a backup!

I'll get this one off my chest straight away, RAID is a great way of protecting yourself against equipment failure. If a hard drive in a RAID5 array fails you will not loose your data and all you have to do is replace that drive in order to restore your protection. So far, so good. What RAID does not protect you against is the user who inadvertently deletes a file and suddenly needs it again; the fire that burns down the building or the auditor who needs to see the customer approval that somebody thought they didn't need anymore. Even if you have a RAID attached to your server you still need to back it up.

How much data do you have?

Consider what you want to back up and what you will need if you wanted to start over again. If you don't have centralised storage and don't use a workflow with version control there is a good chance that you have duplicates of files scattered across your network. A reasonable starting point would be to tot up the sizes of the user's home folders on each of your Macs and add in any servers or NAS devices that you may use. Don't just think about how much data you need to back up today, consider how much you will need to be able to store in 18 months, 3 years or whatever period you would write off your investment in backup over.

How much changes each day?

It is important to remember that this means any file that has changed within that time and not just the amount of new files that are created, e.g. an Entourage mail database file can be up to 2GB in size and any time that Entourage is used this file will be changed and so will require backing up in it's entirety. Since a user is likely to be accessing their mail on a daily basis Entourage will cause up to 2GB of data to be backed up every single day for each user if you are backing up each user's home folder.

Remote users

Don't forget your remote or mobile users. Laptops are very attractive to thieves so as well as securing the data on them you should ensure that you have a backup of it. If you use Mobile Homes you will always have a copy of the data on the server, where it can be backed up, but if you don't there is another option. EMC Insignia's Retrospect has a feature called Backup Server, which constantly monitors the network looking for computers that have not been backed up recently. If the computer appears on the network, for example when a mobile user is in the office or when a remote user connects to your VPN, the Backup Server will notice that it hasn't been backed up recently and will start to back it up as soon as possible.

How long do you have to back it up?

To be safe you should only backup data when it is not being used, effectively that means when all of your users have gone home for the day. If you never close you may have to take several "snapshots" of your data at various points during the day.

Choosing a backup technology

Knowing how much data you have, how much changes each day and how long you have to run your backup you can start to make decisions about what sort of backup technology to use.

When looking at a tape technology always be careful to check what the capacity is. All tape formats list two sizes, a native capacity and a compressed one. The native capacity is the only one that you can guarantee. The compressed capacity, between 2x and 2.6x the native one depending on tape format, is an estimate of how much data could be stored on the tape but there is no guarantee that this amount will ever be achieved. You should always plan on just using the native capacity, treat the compressed capacity as a bonus that you can run into if you have to but you should not rely on it. Some types of data are a lot less compressible than others and some cannot be compressed at all.

How do the most common tape technologies cope with backing up 1TB of data?

Technology	Capacity (GB)	Speed (GB/hr)	N° Tapes	Duration (hrs)
AIT-3	100	42	10	23.8
AIT-4	200	84	5	11.9
DAT-72	36	10.8	28	92.6
VXA-2	80	21.6	13	46.3
VXA-320	160	43.2	7	23.1
DLT VS160	80	28.8	13	34.7
DLT V4	160	36	7	27.8
DLT S4	800	216	2	4.6
LTO-2	200	93.6	5	10.7
LTO-3	400	245	3	4.1

Speed & capacity figures are for native performance only. Duration is for write to tape phase only and does not include verification or tape loading. Speed based on SCSI connection.

Once you tot up all the data that you need to backup you may well find that you have a good bit over 1TB of data to backup. Indeed none of the solutions listed above would be capable of backing up a fully populated 10.5TB Xserve RAID, configured as two RAID 5 arrays, in less than 36hrs.

That leaves you with three basic choices; either write data to more than one tape drive at a time, don't backup everything every day or adopt what is known as a disk-to-disk-to-tape backup.

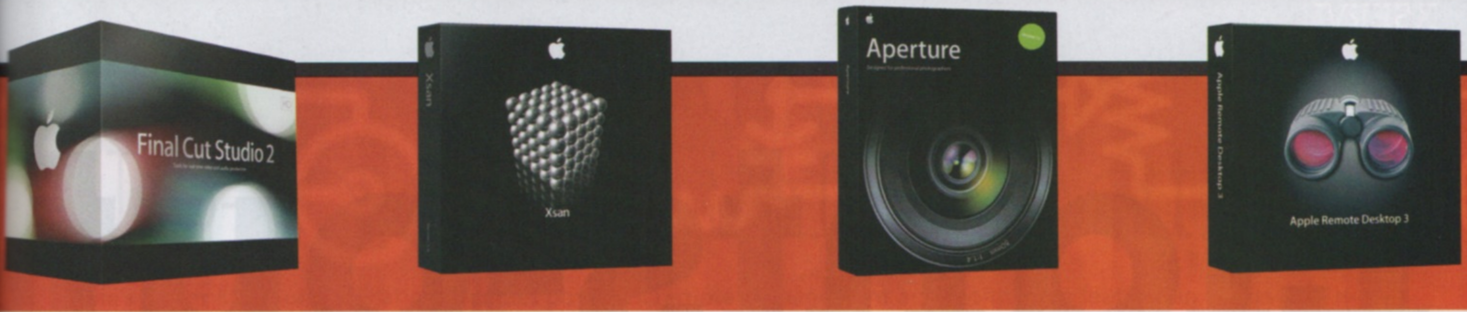
Using more than one tape drive is quite common in enterprise scenarios but there is one major stumbling block on the Mac and that is that Retrospect, the most commonly used backup software for OS X, is not capable of backing up to more than one tape drive at a time. If you want to use more than one tape drive you will have to use alternative software such as NetVault from BakBone.

With a disk-to-disk-to-tape strategy you first back up to some fast storage such as a RAID array that you can get all of your data onto within your window of opportunity. Once you have taken that snapshot of your data you then have the rest of the day to be able to back that snapshot up to tape.

NetVault from Bakbone has a very powerful feature that allows you to back up to disk as what is known as a Virtual Disk Library that allows you to back up to large and fast disk storage that is seen by NetVault as a tape library, once the backup has been written to disk it can then be automatically duplicated to tape giving you the speed of a backup to disk but the safety of off-site storage.

Incremental or differential

For most users there is no need to backup all of their data on a daily basis, as most



of it will be unchanged from day to day. You can take advantage of downtimes such as the weekend to do a full backup and then only backup the things that have changed each day. There are two different ways of doing this, incremental and differential backups.

With an incremental backup we only backup the data that has changed since the last time that we did a backup. With a differential backup we only backup the data that has changed since we last did a full backup.

The longer the gap between your differential backup and the last full backup the longer the backup will take, as more data will have changed. Where differential backups score highly is when you come to restore your data, as all you will need are the full backup and the most recent differential backup. With an incremental backup you will need the full backup and every incremental backup. This is not only slower than restoring from a differential backup but it is more risky since you could require a lot more tapes, and the more tapes that you require the higher the chances that one of those tapes will develop a fault. If you have a faulty tape that is in the middle of a group required for restoring from an incremental backup you may not be able to restore the data that you need.

"For most users there is no need to backup all of their data on a daily basis"

The time taken to perform a backup is not your real issue, the important thing is how long it takes you to recover from an incident, how long you have to stop being productive.

A point to note is that Retrospect for Macintosh neither supports differential backups nor does it allow you to write its IncrementalPLUS backup to a different tape than the one it wrote the full backup to. This means that a single faulty tape presents a much greater risk for a Retrospect backup set than it would do for one that was done using differential backups.

Grandfather-Father-Son

A common backup scheme is known as Grandfather-Father-Son whereby once every 4 weeks you do a full backup (the Grandfather), reusing those tapes each year, once a week you do a full backup (the Father), reusing those tapes once every 4 weeks, and each day you do an incremental backup (the son), reusing those tapes once a week. Using this scheme you can restore a system to an exact day within the past week, to within a week for the past 4 weeks and to within a month for the past year. You may need more granularity and so a more complex scheme will be required.

In order to perform a Grandfather-Father-Son tape rotation you will require 14 periodical sets of full backup tapes, 3 weekly sets of full backup tapes and 4 daily differential backup sets of tapes to cover any changes that are made to files. If all of your data will fit on a single tape that still means you will need 21 tapes. For most businesses a single tape is not adequate and it is quite normal to have a full backup that spans over 5 tapes with differential backups fitting on a single tape, which takes our tape requirements up to 89 tapes for an annual backup rotation.

It is also sensible to make copies of each of the weekly and monthly tapes and to keep those in a separate location to the masters as tapes can be affected by environmental conditions.