Cloud Services and Backup

John Steele

© John Steele Aug 2021

What we will be covering

- Why backup your data
- What is “The Cloud”
  - Relevance to cloud as backup strategy (rather than just storage) in the cloud
  - Cloud backup options
- Local backup options
  - Do I need to automate backup?
  - External USB drives
  - Local network storage (NAS)
- Backup software
- Note that this talk is based in Windows solutions
  - The principles also apply to MAC users

© John Steele Aug 2021

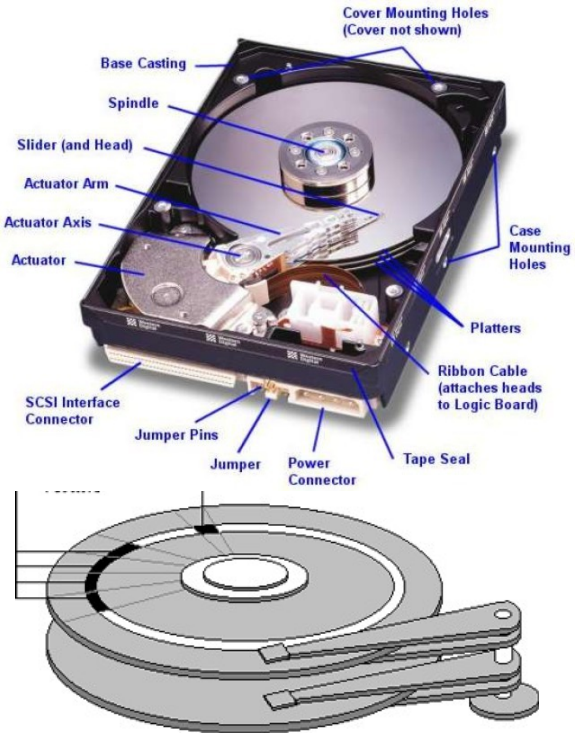
What is stored on your device

- Types of data
  - Operating System e.g. Windows
    - Can normally be reinstalled if necessary
  - Applications/Programs
    - can usually be reinstalled so may not need to be backed up
  - Temporary files
    - do not usually care about temporary files or files in recycle bin
  - Important data files that you do not have any other copies of e.g.
    - Documents you have written or received
    - Emails

© John Steele Aug 2021

Inside a standard disk drive – main storage medium

- One or more double sided platters with heads on top and bottom of each platter
- If not been used for a while, spin up disk (if necessary) and wait for it to come up to speed
  - Disk rotates at 5400 or 7200 rpm (or faster in servers)
- Wait for Read/Write heads move to the specific track required. All heads move together.
- Wait for required sector(s) to reach head
- Read one or more sectors



© John Steele Aug 2021

What can go wrong with rotating disk drives - HDD

- Apart from human error (by accidentally deleting data) any rotating disk drives can fail
  - Ether gradually or abruptly without warning
- They have a number of “platters” spinning at typically 5400 or 7200 rpm for home use
  - They can now store an enormous quantity of data up to 10 terabytes or more but 1 Terabyte or less is more usual in a home computer
- Each platter (often glass) has a magnetic coating on both sides formatted as concentric “tracks” that can be written to and read by “heads” that are moved across the surface of the platters
  - To avoid wear on the platter and the heads they “fly” at about 3 nanometres above the surface – about the same as a DNA molecule (according to Wikipedia link)
    - If the heads were to “crash” into the surface the disk would be destroyed
  - They are mechanical devices and will wear out
- Disks can, and do, fail, and often suddenly, and your data would then be lost forever!
- A sharp shock especially when spinning can cause a head crash and hence immediate drive failure

© John Steele Aug 2021

Solid State Drives - SSD

- SSD drives are all electronic – no moving parts
  - More physically robust
  - They do have a limited number of Read/Write cycles and, while they are designed to even out the “wear”, they will ultimately fail
- SSD drives are currently far more expensive at larger sizes than rotating disks
  - But getting cheaper as time goes by
- 1 Terabyte SSD drives are available for laptops, are amazingly fast, and are being widely used in mid to high range laptops
  - 256 Gigabyte and 512 Gigabyte are more common

© John Steele Aug 2021

Why do we need to backup our data

- Computers are electronic devices
  - Electronic devices can fail, often suddenly, data becomes inaccessible and possibly corrupted
  - Disk drives were all mechanical devices until Solid State Drives (SSD) became available
    - Spinning disks can fail (bearings) and have moving R/W heads that can crash into the disk surface
  - Computers, tablets and phones can be lost (especially laptops) or stolen
  - There could be a disaster (fire or flood) or dropped from any height (laptops) e.g. a desk
- You are not infallible – human error aka “finger trouble”
  - You might delete something by mistake and only realise later
- You may be infected with a ransomware virus making your data inaccessible
- Your data (well at least some of it) is presumably important to you
  - What would be the impact on YOU if your important data was lost permanently?
- BUT - Can you rely on remembering to backup your data before it is too late
  - Oh – I will take a backup tomorrow…

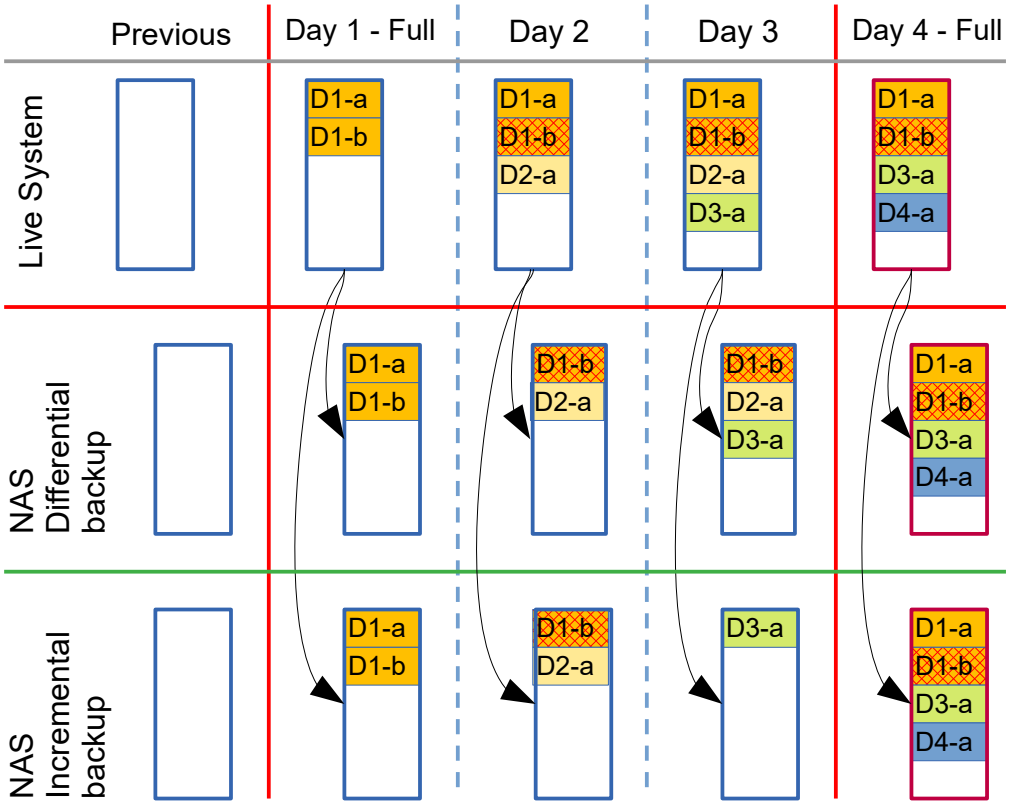
© John Steele Aug 2021

Backup strategies

- Full backup
  - Backup ALL data each time it is run
    - Could overwrite previous backups but then you only have one copy
    - Could create a new copy each time up to a limit but most of the data is the same in each backup
- Differential backup
  - Backup all files that have changed since last full backup
  - Previous copies are not overwritten
- Incremental backup
  - Backup only the files that have changed since last backup (full or previous incremental backup)
  - Previous copies are not overwritten

© John Steele Aug 2021

Comparison of Incremental and Differential backup



- Comparing Incremental and Differential backup strategies
  - Day 1 two files created, D1-a and D-b
  - Day 2 One file, D1 b, edited but saved with same name, file D2-a added
  - Day 3 New file D3-a added
  - Day 4 – New file D4-a added - Full backup again
- Incremental takes less space
  - Only changed files saved
- Differential is quicker for restore
  - You only need to recover the previous full backup and most recent differential backup

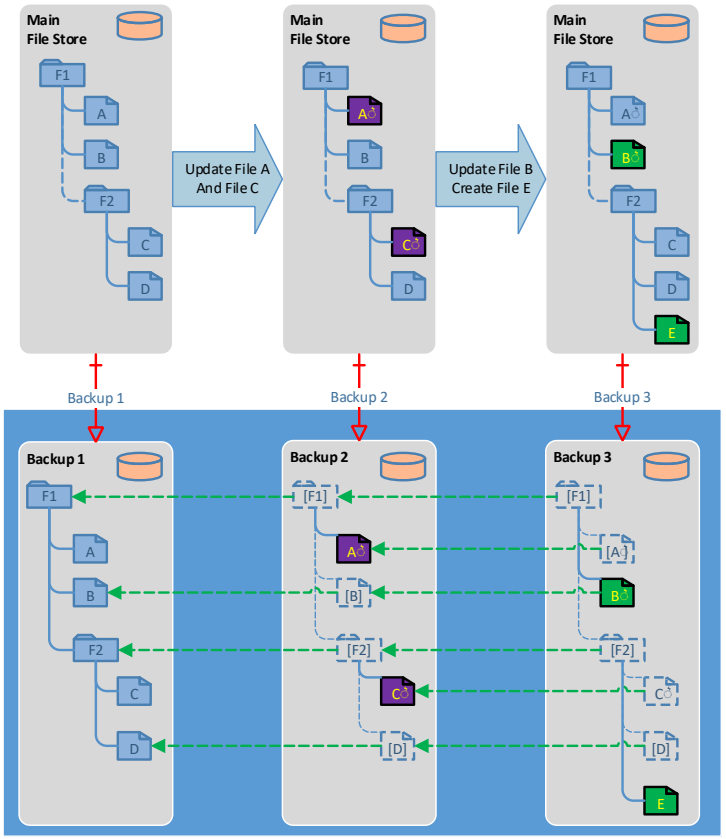
© John Steele Aug 2021

File versioning as an alternative to daily backup

- Cloud services and some local back solutions support File Versions
  - When you make a change to a file
    - The previous copy becomes an old Version
    - The new file becomes the current Version and is the only one directly visible in the File Explorer
- With this method there will be a method of recovering an older version and making it current
- These systems usually have a configurable system wide limit of the number of versions retained for each file as they each use up valuable storage space
  - It is then YOUR choice as to how many versions are retained

© John Steele Aug 2021

MAC Specific - Time Machine



- Slide from a previous backup presentation showing that MAC keeps a record of the current state with a pointer back to earlier times
- Avoids duplicating data
- Windows can now do something like this as well using File versions
  - Unknown to me is what happens if a file is modified frequently – does it retain all versions forever?
  - If the latter I am concerned about how much space to allow over extended periods

© John Steele Aug 2021

Backup – types of data

- Operating System and programs you install
  - These do not change very often and, if lost, can be reinstalled
  - Backup is therefore a convenience rather than a necessity
    - New computers usually come with a means of recovery to the initial installed state
    - You can take an “Image” of a computer that allows you to go back to a known good state
      - This is not normally the best approach for regular user data backups
- Your Data - Personal to you and will be unique, some could be important
  - Documents you created or have received
  - Emails you created or received
  - Without backup your data could be lost forever
    - Only you can judge what will be important!

© John Steele Aug 2021

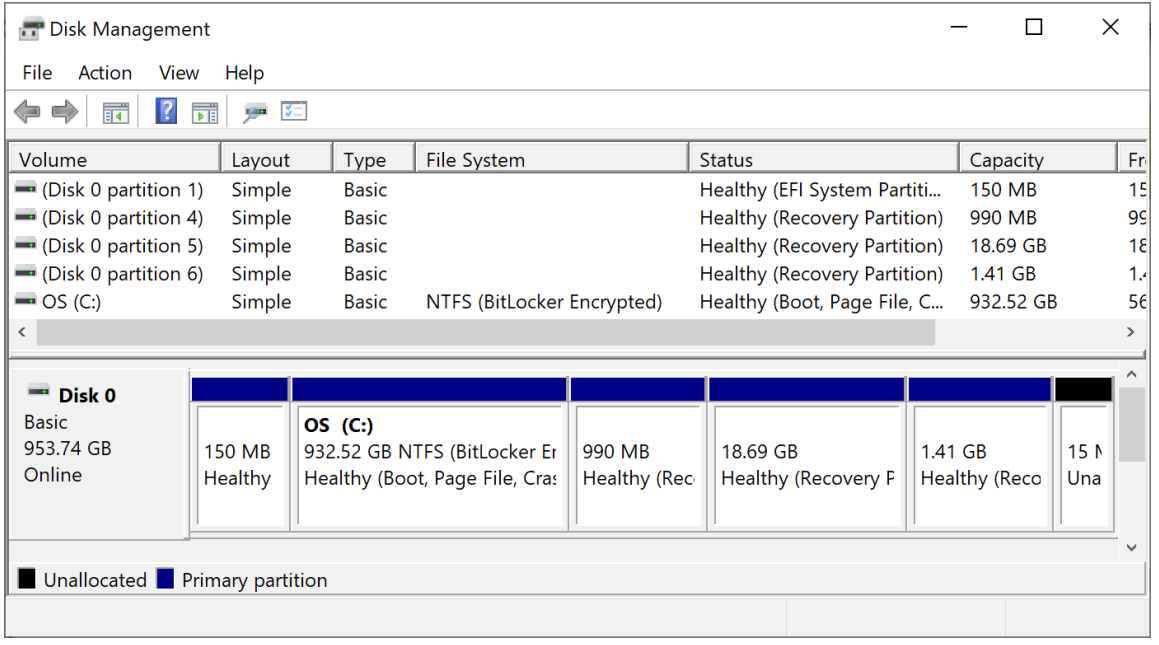
Understanding Disk Partitions

- Modern disks are usually split into separate functional areas called partitions
  - This provides isolation between critical areas of the Operating System
- Some Operating Systems segregate Operating System from User Data into separate partitions
  - Windows does not encourage this while Unix/Linux derived systems do provide this separation. This includes MACs.
    - Although it can be done I have now given up separating my user data from the OS program data on Windows into different partitions
  - Windows still has some “hidden” partitions however which are important
    - Do not touch them unless you really know what you are doing

© John Steele Aug 2021

Windows disk partitions

- Here is my partition table
- Windows assigns a Drive Letter to accessible partitions
  - C is always the main one having reserved A and B for Floppy Disk drives
  - A DVD drive will be allocated a letter if you have one or if you plug one in
  - USB drives will also be allocated a letter when you plug them in



© John Steele Aug 2021

What risks are you protecting against

- “Finger trouble”: You accidentally delete some important file
- Hardware failure: A hardware failure makes your data inaccessible e.g. head crash
- Ransomware: Encrypts all of your data
- Domestic disaster: Fire in the home, or theft from your home or when away
- Loss of typically laptop when away from home: Leave it behind at a bar
- Etc.

© John Steele Aug 2021

Types of Backup – Disk/Partition Image

- An image backup will take a complete copy of your entire disk (all partitions) or selectively one or more partitions
  - With Windows the System Partition is shown as the “C” drive
    - If you partition your disk the user data can be separated into its own Partition and you can allocate it a drive letter e.g. U
      - Partitioning is more complex and has implications for Solid State Drives.
      - I have not partitioned my new computer which has an SSD
- There are several disk imaging programs e.g. Macrium Reflect free edition
  - Takes a full (compressed) image onto a local USB (typically) disk or NAS
  - Subsequent backups can copy just changes to that image (incremental backup)
  - Typically requires a disk space of about 30% to 50% of the source disk image

© John Steele Aug 2021

Restoring from an Image

- Typically by design Image Restore will overwrite the whole of your disk partition or the whole drive so use with care!
- You MUST have previously created a bootable recovery media containing the recovery program so that you can restore an image to a new, or overwriting the existing, disk
  - Recovery media = bootable DVD or bootable USB
    - Macrium can create this for you
    - It does NOT have to be created on your computer
  - You should test that you can boot from this before you need it!
- Image backup/restore is intended to recover from a disk failure and assist replacement
  - Macrium can recover individual files from an image
  - **But do I not recommend Image Backup/Restore as your only backup**

© John Steele Aug 2021

Data backup – Options for Individual files

- Where to backup to:
  - A separate area of your existing disk drive
    - Better than nothing
  - Cloud – off site
    - Most secure against physical damage as data is off-site **if you trust the cloud provider**
      - Data is often stored in plaintext on their servers
    - Can provide a defence against ransomware – versions are usually retained and versions cannot be directly accessed
  - Plugin USB HDD
    - You need to remember to plug it in!
    - Immunity from ransomware if NOT plugged in and you know your are infected
  - Local dedicated file store (NAS – Network Attached Storage)
    - Can be effective as a defence against ransomware by careful configuration of backup accounts
- I personally use both Cloud and NAS

© John Steele Aug 2021

Cloud Storage – Typical features

- Typically an area of your current storage that is replicated to a remote server “in the cloud”
- Needs an account with the cloud provider – even for free cloud services
- Storage size is limited to typically between 2 Gbytes and 5 Gbytes for a free service, chargeable for more storage
- Data is encrypted “in transit” to the storage provider
- Data is typically NOT encrypted on the cloud storage and hence can be viewed by the cloud provider system administrators or by a hacker of their servers
  - TeamDrive is an exception
- Will usually keep several “versions” of each file. Ransomware can only encrypt the most recent copy. Previous versions can be retrieved

© John Steele Aug 2021

Some Cloud Storage providers – Others are available

- All of the following can share data between multiple devices, not just Windows
  - OneDrive – Microsoft: 5 Gbytes
    - If you have Windows you probably already have this installed. It may not have been activated and needs a Microsoft on-line Account (Outlook.com or Hotmail)
  - DropBox: 2 Gbytes free
    - Can be installed. Often comes pre-installed on a new computer
  - TeamDrive: 2 Gbytes free
    - Can be installed. Be careful to avoid registering for a commercial license it is difficult to revert and needs TeamDrive admins to do it
    - **Encrypts all data on your computer before being uploaded**
      - Data is inaccessible to any cloud host administrator or TeamDrive staff (although they can see the file names)
    - Based in European Union – German company

© John Steele Aug 2021

Network Attached Storage

- This is the “always on” local storage and is the “Rolls Royce” solution
- Contains one or more Hard Disk Drives to act as a backup server to your local network
  - Drives should be designed for NAS use – they are powered on permanently (but usually not spinning unless needed)
  - SSD not recommended here!
- Most reliable if if is complemented by an Uninterruptible Power Supply (UPS) that will provide power if you have a power cut – can potentially corrupt data
  - Note that UPS is not intended to power the device for along time. Just to cover a short outage (e.g. 15 minutes) and then shut the NAS device down in a controlled manner so that there is no corruption

© John Steele Aug 2021

NAS configurations – Part 1

- NAS comes in a number of configurations which may be confusing
  - They are usually supplied WITHOUT disk drives which must be bought separately
  - They typically connect via a wired Ethernet cable to your router and can be accessed by any device on your local network
  - Some can be configured to act as a private cloud server and many other tricks
- Single drive
  - Most basic and lowest cost
  - Disk drives can fail so your backup solution is at risk if the single drive fails
- Two drives
  - The two drives are configured in a special way so that data is recorded to BOTH drives. This is called Mirrored drives and is also called RAID 1
    - If you have 2 * 1 Tbyte drives you only have 1 Tbyte of storage
  - I would recommend this as the minimum preferred entry point due to its reliance

© John Steele Aug 2021

NAS configurations continued

- NAS storage can also have more than two bays and usually the jump is to 4
  - There are several configurations available but the most basic recommended one is called RAID 5
  - This would usually have a minimum of 4 disks (although it could work with three
    - Data is distributed across all of the drives but with 4 drives the data is distributed across three of the drives and the fourth is a validation block called a parity
    - You can therefore use 75% of the total capacity of the four drives and you are protected against failure of one drive
    - If a drive failure is detected you are notified and inserting a new blank drive will automatically be added to the drive pool repairing the redundancy
  - There are other permutations – ask for advice before investing!

© John Steele Aug 2021

NAS usage

- Whichever NAS configuration you select it can be configured to appear as a Network Drive in Windows.
- It can then be used like any other drive but with one proviso
  - You must configure the NAS with the user account credentials that you are using
    - These should match your Windows account
    - Note that if you have two (or more) Windows accounts they must ALL have accounts created on the NAS if you need them to be able to access
    - This only has to be done once
  - As a security hint it is possible (and recommended) that for backup you should create a special Windows account called (for example) Backup
    - Allow this account to have full read/write access on the NAS
      - The Backup account can backup all Windows user accounts if given Read access permission to those accounts
    - User accounts have Read Only access to their backup data (so that it can be restored)
      - Ransomware needs to have write access to encrypt your data in the backup. It cannot write to the NAS unless it can also gain access to the Backup account credentials
    - Ask for help/advice if you are intending to follow this route! It needs t be configured carefully but only once!

© John Steele Aug 2021

Some points to note about cloud services and Windows

- All the listed providers have “Overlays” which can cause Windows Explorer to indicate the state of the file in the cloud e.g. locked by another user or being uploaded
- **There is a generic Windows implementation problem** – Microsoft have only allowed for 15 such overlays within the Windows design and grabs 4 of these for its own purposes. This is a long standing but not widely publicised limitation. Any overlays beyond 15 are still stored – but ignored
- The most recently installed cloud program pushes all of the other overlays down the stack so they are not operational – here is a partial list, there are many other cloud services!
  - TeamDrive requires 8 overlays
  - OneDrive requires 7 overlays
  - DropBox requires 10 overlays
- Updates to any of the cloud services, e.g. OneDrive, pushes its overlays to the top of the stack again!
- **Fixing the overlays for your preferred cloud provider needs some editing with a program called Regedit. I do not normally recommend its use unless you are an “Expert”**

© John Steele Aug 2021

Backup storage solutions - Documents

- It is likely that Cloud storage will not provide all of the backup you need. Its main benefits are
  - Off-site protecting against fire or theft and Its versioning protects against ransomware
  - Backups are automatic
  - It will have quite limited storage in the free version
- The “cheap” solution is to use an External USB drive to keep the the files you want to backup
  - The advantage of this solution is that the USB drive is not normally plugged in (or should not be) so it does protect to some extent against ransomware
  - The disadvantage is that you have to remember to plug it in and run the backup software
- The most comprehensive solution is a dedicated storage server e,g, NAS just for this purpose
  - The advantage of this solution is that backups can be scheduled using the right backup software and just happen e.g. daily without any user action
  - It can be configured to protect against ransomware but that is more complex

© John Steele Aug 2021

Backup strategies

- There could be situations where you want to keep more than one backup copy of a document
  - e.g. a complex document that evolves over time. You may have accidentally deleted a page some days/weeks ago and only just noticed. With multiple version backups you can go back and retrieve the older version
- This is solved by keeping multiple historic versions but how many?
  - Simple version management is typically handled by the Cloud providers
    - You can configure limits as to how many versions to keep
  - Local solutions can do this or perhaps have a strategy of making a full backup periodically and in-between only backup files that have changed I.e. Incremental backup
    - Differential backup is faster to restore but file restore is, or should be, a rare event. I recommend using Incremental backup

© John Steele Aug 2021

Backup software

- Windows backup to external drive or NAS
  - Windows can be configured to automatically backup changes to files
  - It only backs up to an external drive. I used my NAS to test it. I could find no way of limiting the amount of storage used – it just grew.
- Other Software – Cobian is a good choice for external drive or NAS
  - Cobian requires one or more “tasks” to be created that defines the policy
    - A typical task would be to take a full backup every 28 days to back up a configured set of files. It does not need to be all of them and you can have multiple tasks.
    - In-between full backups it will back up any files daily that have been changed or added since the last full backup. This is an incremental backup.
    - It can backup open files using the Windows Volume Shadow Copy service

© John Steele Aug 2021

Backup issues - Email

- There is a specific problem backing up email if you use Microsoft Outlook or the default installation of Thunderbird. I like to take a daily backup of my email.
- Outlook (and the default Thunderbird configuration) typically puts all of the emails into a single file so the only way of backing this up is to copy the whole PST file. It is feasible to separate archive data into one or more OST files which can partially alleviate the situation
  - This did not happen with Microsoft LiveMail (which is what I have used for years) as each email was stored in a separate file
  - My single email folder would be a single file of about 6 Gbyte. Incremental backups of part of the data are not possible and storing that for the three months I have as my backup cycle would consume VAST amounts of storage approx 90 * 6 Gybtes = 540 Gbytes or 2/3 of my current NAS storage just for email for one computer.

© John Steele Aug 2021

Email backup continued - Duplicati

- I have switched to Thunderbird for my new computer and am EXPERIMENTING with a backup program called Duplicati which uses a different method of storage that is far more efficient for Outlook (and Thunderbird) storage files
- I have since discovered that Thunderbird can now store emails as separate files. Technically this is still experimental but reports are that it is reliable and has been so for some years. While I have configured Thunderbird to the single message per file format on my wife's laptop I have not yet converted my own mailboxes to the new storage method so I am still using Duplicati. Duplicati also has the advantage of backup speed over Cobian. Cobian takes about 3 hours to backup my full email, including my archive storage folders, once a month Duplicati takes about two minutes as all backups after the first are effectively incremental
- I cannot yet confirm the reliability of Duplicati but will support any member who wants to try it! I currently think that I will continue to use it even after I convert my Thunderbird mailboxes to the individual file format

© John Steele Aug 2021

Demos

- MAC:Time Machine – Dave Harrop
- Macrium configuration – Image backup
- Cobian configuration – Incremental or Differential file backup
- Duplicati configuration – Experimental backup (differential on steroids!)

© John Steele Aug 2021