Backing up Data in Personal Computing

Purpose: With the rapid accumulation of personal files in computing, the need to find an effective strategy to backup data is corresponding more urgent. With the advent of better hard-drives and cloud storage, one would welcome a strategy that fits in the balance between ease of finding files and the ease of implementing a backup strategy. Our purpose is not to find a definitive solution (if that even is possible) but to explore the components of such a strategy and reflect on their merits, leading to a suggested backup strategy

Preface

Ultimately back up data is time-bound in one way or other; hard-drives suffer failures as they get older and although cloud storage appears faultless, the user may fail to pay the fees, the provider may fail in the competitive space and even file types can become obsolete. All backup strategies need a value system eg. priority/non-priority, urgent/ non-urgent etc. We will try to constraint values to a maximum of 5 levels but actually aim for 4 as to make recall more instantaneous. Bandwidth performance and cost is still an issue in Uganda, and with that said, we see the best strategy is one that uses both hard-drives and cloud storage.

Outline of the Components involved in Backing Up

(1) **Time Period:** such as 5yr, Yearly, Quarterly: Essentially all files are time-bound which is the very foundation of the concept of "Archiving". It's importance lies in the idea of closure wherein archived files remain unaltered. Thus in a storage strategy, archives are stored once and need no updating, eliminating further action and used as means of reducing the storage involved with current folders and files that are new or frequently changed.

(2) **Priority:** Storage is much more simplified and efficient, if one can incorporate a scale of priority. Personal/family identification information will, for example, be required to be always found as opposed to an interesting news snippet that's worth keeping for a few days prior to be discarded. Having a method to either tag and/or store on a priority basis makes for good file management. Here below is a suggested framework:

a) Level 1: "A" – for top priority items, possibly secured in encrypted folders, to be permanently available – no time limit for their disposal. Examples: Passport scans, a password-locked file with banking info,

b) Level 2: "B" – Information that needs to be kept for some length of time. For example: home utility bills (3-4 years?), car maintenance records, subscriptions, supplier info, etc.

c) Level 3: "C" – Information covering a specific period but where the information could still be useful thereafter: such as summer pool schedules, price lists, invitations. Such information would be discarded after a year (or held no more than 2 years)

d) Level 4: "D" – The very opposite to top-priority, these items have a very short life (less than 30 days) in the storage structure. Examples: An advert for an item that one may purchase within the week, a screenshot of some interesting info.

It's worth mentioning here that the concept of ownership often goes in tandem with priority. Original/personal work will often have no copy elsewhere whereas work from the public-domain may be easily replaced, thus links, not storage, may suffice in these cases. Institutional/group information also made well be stored securely elsewhere. One then could apply priority values on what should be stored in personal folders and for how long.

(3) **Versions:** The evolution of documents and other works (graphical, musical, etc)

Often with project files, there is the issue of versions in the evolution of the work – usually done in a group setting. The issue is commonly addressed with tagging file labels with extra information, such as V001or ver01a. While this needs to be considered, this is a wide topic that can't be adequately covered here.

(4) **Location**

It would helpful to visualize the general storage structure within a computer in a graphical sense that gives richer meaning to each level. Below is an example

a) Level 1: "Warehouse" – These being the overall folders (often found as the default) such as "Pictures", "Documents", "Music" etc. that are meaning in the way that they describe data areas where great variety and large amounts of file will be found corresponding to such concepts. New "Warehouses" folders are obviously created as per one's needs.

b) Level 2: "Shelf" – like those large shelving found in "warehouses", they accommodate a more specialized concept or a specific sub-topic, giving the folders, labels such as in "Pictures": "MyFamily" separate from say, "MyArtwork", yet still broad enough to embrace a wide range of topics within each of these sub-topics. Moreover, this may possibly but likely <u>not</u> be a time related such as "MyFamily2019".

c) Level 3: "Box" – This needs to be distinct from level 2 in a more specialized way and to be definitely time-tagged. Following the above example, one could "MyFamily-Summer2019" or "MyArtwork2019" – justified by the amount of different shooting sessions where each session would be a sub-folder at level 4.

d) Level 4: "File-Folder" – The final folder level where folders have labels that are time-stamped (eg. 190620 Birthday Peter 5th - for a photo session that 20 June 2019) and the subject briefly stated. The folder would have only files and avoid further subfolders.

Suggested Strategy

Of the above components, one will approach particularly the location levels 3 & 4 with the most attention, having the goal of ultimately archiving these folders therein. For level 3, it is good to find the time-period best suited for archiving: be it every 5 years, yearly etc. This time-period may differ for a level 1 "warehouse" folder to another.

As for level 4, there similarly could be differing labelling strategies for different topics. Picture folders at level 4 (file-folders) as mentioned, could simply start with a date (YYMMDD year first since files are sorted alphabetically) followed with a short description. Folders for correspondence , the labelling could be a short code for the source (eg.'Hom' for home), a date in the form of YYMMDD, a target (optional) and a short description. Here's an example: Hom190620 Water Utility receipt 23USD.pdf

Moreover, one would visualize storage of having three distinct areas:

1. **Archived Folders:** files accessible but to remain unchanged. These folders contain, in part, the label "Archive". These folders have no need to be on the computer hard-drive but to be moved and copied both to a portable hard-drive and to a cloud-storage. If changes are required with any file therein, a copy is drawn out into the Active Folders (point b.), initially via the Current Folder (see point c.), and the file, now changed, is not to be returned to its initial archive but will be part of another more current archive.

2. **Active Folders:** This is within the computer existing file structure for personal files: that has the usual headings of "My Documents", "My Pictures" etc. Folder labels are without the tag "Archive". These folders periodically (weekly, monthly or quarterly) are copied both to a portable hard-drive as well as to a cloud store. Don't synchronize these folders with the cloud-storage as this will engage much bandwidth and processing power.

3. **Current Folder:** The folder where all newly acquired, recently changed or freshly created files (and folders) are first stored. This folder is synchronized with the mirrored folder in the cloud store. Once a week, month or quarter, the Current Folder is dated and copy made. From the copy all files are sorted into the appropriate folders in the Active Folders. When the dated Current Folders reached three or four in number, the oldest is then deleted. Lastly, it's also an ideal place to have a folder called "Scrapbook" or simply "CB" (for clipboard) to hold temporarily files which are interesting momentarily but will not be considered for storage.

Having these three distinct storage areas will ease the task of backing up. The archived folders having not been changed, there is no need for these folders to remain on the computer as they can be accessed online. Archiving files reduces the content of the Active Folders to only to a few Gb's. That done, the backup task would not be of long duration. Given that the Current Folder is synchronized with its equilvalent cloud folder and that it's earlier dated ones are also copied both to a portable hard-drive and to a cloud store, then any mistaken deletion or a badly saved file can be readily found and restored.

Storage Options

One needs to estimate the kind of capacity expected to contain one's files during a sufficient time period – often one looks at a 5-10 year time-frame. Having assessed the rate of growth of data then the choice for storage becomes clearer. Given the current state of bandwidth in Uganda, one would consider the hard-drive as the "primary master" and the cloud-storage as the "secondary master" in the view of restoring files.

(1) **The Hard-drive:** Obviously one would wish to get a very good quality unit that has both sufficient capacity for the period of its use (one should consider replacing a hard-drive used for critical file storage every five years.) The downside: The hard-drive are prone to failure and possibly, misuse or theft. Care toward its handling and its security needs to be considered.

(2) **Cloud-storage**: Most cloud storage providers provide means to grow storage according to need. Thus one would take a conservative view on the initial choice of storage options as well as to consider the cost benefit of making annual payment instead of the monthly one. Importantly, the cloud-storage should have the benefit of an efficient search ability as well as ease of use. Currently, Google Drive and OneDrive from MS have these benefits and moreover attractive thresholds of free storage before charging reasonable rate above that threshold. The downside: cloud-storage requires good and cheaply available bandwidth, which in many countries is still a challenge. Here in Uganda, Roke Telkom currently offers unmetered bandwidth after business hours (7pm-7am) but its bandwidth performance varies widely. Zuku Cable has also very attractive offers but obviously serves only those who are close to its cable network.

Write-up by,

John Anglin, 3 August 2019

Notes:

(1) OneDrive – Default Action: On the OneDrive folder – PC-based, the file is always synchronized (as the default). However, one can set the environment to **Files On-Demand** (via the online "settings") where one can choose: File-On-Demand. "When you open an online-only file, it downloads to your device and becomes a locally available file. **If you need more space, you can change the file back to online only. Just right-click the file and select "Free up space."**

https://support.office.com/en-us/article/save-disk-space-with-onedrive-files-on-demand-for-windows-10-0e6860d3-d9f3-4971-b321-7092438fb38e?ui=en-US&rs=en-US&ad=US

(2) **File-on-Demand [OneDrive]**: When placing a folder with a significant amount of files directly to the Cloud store for the purpose of saving PC storage space (using the File-On-Demand option), the **upload will be incomplete if there is an interruption with the connection**. After the interruption, the upload doesn't resume at the breakpoint. **As a solution**; It appears that one needs to first place the folder on the PC-based OneDrive folder in synchronized mode. After the synchronization is complete, one switches (at the Cloud directory) to File-On-Demand which should automatically erase the PC-based copy.

(3) **Folder Size:** Google Drive and OneDrive do not give folder size on the cloud. For OneDrive and where the folder has be set to File-On-Demand, the PC-based folder gives the size in terms of memory taken up by the links. The PC-based folder for Google Drive shows only the synced folders.

(4) **TrashCan:** OneDrive (work account) gives 93 days whereas Google Drive files can spend **30 days** in the Trash and **25 days** in this post-Trash, still-recoverable state, which only provides 55 total days of data protection.

(5) Google Drive: Having the Google Drive folder on the USB device:

a) By the way, the problem is solved: I run a file sync software, that mirrors my folders placed on my USB-drive to my Google Drive/My Drive web-based. Nothing is placed on my C-drive. I can access that Google Drive from any other device I need. https://support.google.com/drive/thread/4857764?msgid=8725083

b) Google Drive Location Change: https://support.google.com/drive/thread/2153532?msgid=2162844