AWS Glacier for Large Data Backups

Aug 4, 2023

After releasing last week's episode, in which I reported my findings on the iDrive service, a friend replied in the comments and mentioned that there might be a way to send data in via offline backup to Amazon Glacier, which I was unaware of. I had looked into AWS Storage but had never heard of Glacier, so I went to AWS and poked around and found that they have a service called Snowball, via which you can request a device to be shipped to you. You fill it with your data and send it back to Amazon. They will copy the data to a Glacier vault that you specify, so before we move on, a huge thanks to Catalin for letting me know about Glacier and ultimately pointing me in the right direction.

CLICK HERE TO LISTEN
(Requires an Internet Connection)

This led me to look further into the Glacier option. Based on this, I'm developing a new backup strategy, so I'll share my findings today as an update to the previous episode, as this is very relevant and, without a doubt, the way to go regarding backing up data online as an offsite backup. Some areas are not so user-friendly, but tools are available to make this a viable option, which I'll share today.

I won't go into all the costs, but I looked through the Glacier pricing and found it very reasonable. Very inexpensive to store data, and although it's a little costly to restore data, it's not extortionate. Besides, I would only need to restore data as part of a disaster recovery process, so I won't mind paying to get my precious data back.

Another critical element to the success of this new strategy is that I also found that Synology, the maker of my NAS, provides a plugin to backup and restore data from Glacier. I had to create an API Key and Secret in my AWS account and drop that into the Synology Glacier app as I create each new backup job, and it automatically connects to my AWS account and

creates a new Glacier vault for each job.

One issue here is that the vaults are created with a cryptic name rather than the name you give to the backup job, which is probably to ensure that you don't lose track of the vault if you change its readable name in the Synology NAS, so if you do ever need to restore something, you'll have to do a few additional checks to find the proper vault.

I also found an application for Mac OS called <u>Freeze</u>, which provides a relatively nice-looking user interface for Glacier so that you can both add files or look for files inside your vaults. One of the downsides to Glacier is that your backup data is buried pretty deep, so viewing what you have inside a vault is slow, taking over four hours. You also have to wait 24 hours after adding your data before it's possible to view it, but again, this shouldn't be necessary very often. It's probably most frustrating as you are setting things up and trying to check your backups.

Without a doubt, the most impressive thing I've found compared to how the iDrive backup was running is that I can stop a backup task, at say 100GB, of a 132GB backup job, and when I restart it, it goes straight back to the point at which I stopped it, and data starts flowing into the Glacier vault immediately. Having to wait days for iDrive to catch up was incredibly frustrating, and the fact that it never completed an initial backup comparing the data I mailed to them with the same data on my Synology NAS made it simply unusable for the amount of data I have.

Now, my dilemma is if I want to automatically continue to incrementally modify my data, as I occasionally revisit old photographs, I have to perform the original backup via the Synology NAS Glacier plugin. I can't see any way to send in some data via a Snowball and then link that back to the NAS Backups. It may be possible to simply create a job,

which creates a Glacier vault even before you start adding data, then have the AWS team put data into the vault you specify, but I haven't been able to confirm this online, and Amazon wants me to pay them for support to answer this simple question, so I'm currently planning not to do that, unless I get stung by my Internet providers, as I'll explain.

My current situation is, that I have everything until the end of 2022 offsite in my iDrive, because I sent it to them on a hard drive, but I have already stopped the iDrive online backup so that I can use all of my daily bandwidth rations to upload my latest data into the Glacier vault.

I started my backups a few days ago, thinking that I would be happy if I could back up all of my new data from this year over the network to AWS Glacier, and then I will have an offsite backup, albeit split into two locations, with iDrive and Glacier. As a test, a few days ago, I started to upload my Finals for 2023, and I think I could limit the bandwidth used by limiting the maximum outbound traffic for each Ethernet port to 300KB/s. That should cap my uploads to around 25GB per day, as I'm only allowed 30GB before my provider starts to shout at me. I have also connected my two 10GBPs ports on the Synology NAS to my 5G router and limited both of those ports to 200KB/s.

I've calculated that I'm currently uploading 84GB per day, so if the traffic has been successfully split between providers, I should still be under the 30GB limit for my primary wired network provider. I don't have an amount for the limit on my 5G router, as they state that if you upload "a lot," they will restrict bandwidth. I guess I'll find out over the next few days if I annoy either provider.

If I can continuously upload 84GB of data daily, I can upload 1TB in 12 days. My entire archive of 22TB would take 262 days or 38 weeks, which is still a long time, but if I prioritize what I backup and

the order in which I perform the backups, it's not as daunting a proposition as it originally was, and I can see an end goal.

If I proceed with such a long-term backup strategy, I must carefully set the optimal order for my backup schedules. From my tests, only one active backup can be running, as it places all other backups on hold. If, for example, I was to create a task that backed up all of my Finals, which is a hair over 1TB, it would stop any other scheduled backup from running for 12 days.

To overcome this, I have included my 2023 Finals folder and a few early-year folders because they don't contain much data. The earlier years finished uploading in just a few minutes, and the 2023 folder took about a day and a half, until midnight last night. Anticipating that job finishing last night, I set my 2023 Photo Originals backup to start at midnight, and when I checked this morning, it had kicked off automatically as planned. If my providers don't shout at me, this upload will finish in 30 days, and I'll have everything important to me in the cloud.

As you can see from this screenshot of my Glacier backups, I have set the Home folder for my user to run daily at 2 am. This won't start until after my Photo Original for 2023 backup finishes, but I want to ensure that the next important task starts as soon as it can. The Home folder on my NAS holds a folder that I use, like Dropbox, to synchronize files between all of my devices via the Synology NAS.

As an aside, to help redeem some of the cost of the NAS, I've canceled a few subscriptions, such as Dropbox, and now use the NAS. You'll also see a Joplin job there, my replacement for Bear, which was my less expensive alternative to Evernote, which I canceled a few years ago.

As recent folders finish uploading, I will set my older Photo Originals backups to run once a week, as I don't update old

archives often, and I'll have them check for updates on different days and at different times. I never do more than edit a few images in old archives, so the incremental backups should finish quickly.

I thought about putting all previous years into one vault, but I figured that would make it more time-consuming if I needed to view my vaults' contents or restore data from them. Hopefully, this will never be necessary, but a business continuation plan that doesn't provide a way to restore data after a catastrophe isn't much of a plan.

So, to wrap this up, I'd say that using Glacier in this way will probably be the best option now that I'm aware of the service. I can recommend it to people that have some decent technical skills, as it does require that you set up a few components, but it's not difficult to set up if you are running a Synology NAS, as the kind folks at Synology have done most of the heavy lifting for us with their Glacier add-on.

If I change my strategy or maybe order a Snowball to transfer my data, I'll update you in a future episode. This depends on whether or not I get another letter from my provider over the next few days telling me that I'm uploading an unfathomably large amount of data again. If I can get away with my current settings, I might just let this run and look forward to being fully backed up on one service by next spring. I may have to pay for a second year with iDrive to hold my data until the end of 2022, but that depends on how much of it is still not in Glacier by the time I have to renew. If I'm just a few months away, I'll cancel my iDrive account rather than pay for another year.

Once again, a huge thank you to Catalin for pointing me in the right direction with this! I appreciate it, as I do the rest of you who tried to help. Seeing how many people jumped to my aid when I was not doing so well with my

previous strategy was heart-warming. Thank you all!

And before we finish, I've now got a date of September 21 to have surgery on my left eye to remove the lens, which has developed a cataract, and replace it with a man-made lens. My right eye is only a few months behind my left, so I may have the right lens replaced by the end of the year, but we decided to leave it in place a little longer while I can still see with it. If it follows my left eye as

quickly as it was, I won't be able to see with my right eye by around November, although hopefully, my new left eye lens will augment my vision once that is in place. I'll monitor the situation and let you know how it goes.
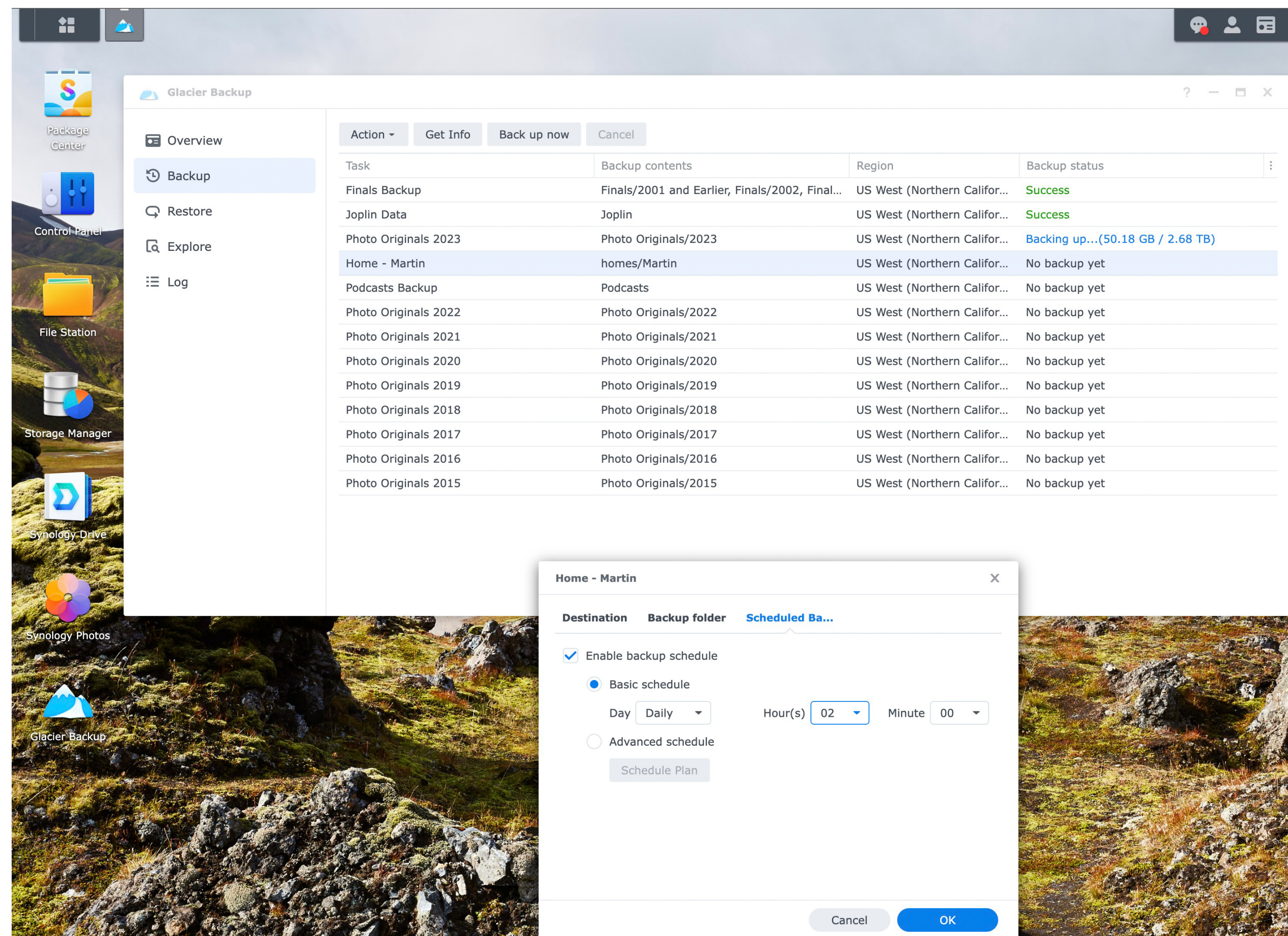
Martin Bailey Photography K.K.

This article was first published on Aug 4, 2023

See You Online!

Check out our Tours & Workshops, eBooks and fine art prints and other products and services, or listen to the audio of our articles on our website!

https://martinbaileyphotography.com