

Only you can see this message



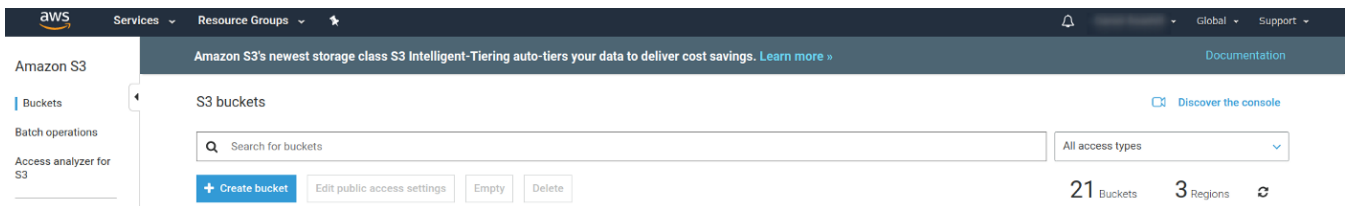
This story's distribution setting is on. [Learn more](#)

How To: Save Gsuite and Gmail Storage Space by Archiving Old Emails and Files To AWS S3



Daniel Rosehill

Feb 4 · 7 min read



As well as taking backups of my desktop and cloud services, another really cool tech practice I'm into* is **periodically cleaning up my inbox and cloud-hosted filesystem.**

My inbox seems to be getting progressively busier and more cluttered these days, so I'm trying to up the frequency of these clean-outs to once every three months rather than my hitherto annual swoop.

Using Google Takeouts to Export Gmail Labels to AWS S3



I've just closed up a long project with a client.

So I thought that this was a good opportunity to demonstrate — and document — my current methodology.

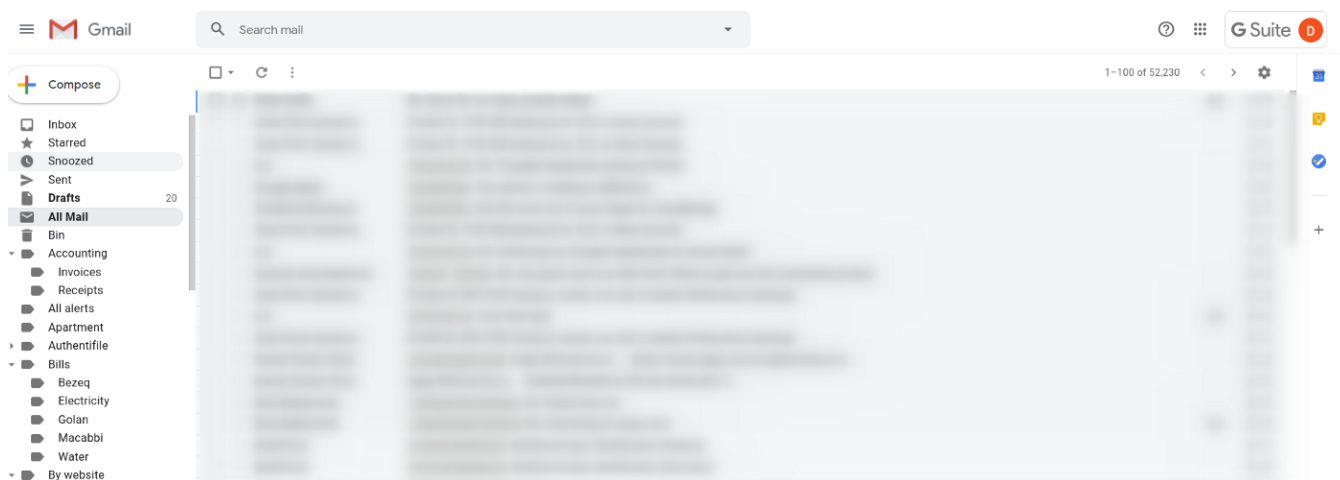
Here's how I do it.

*Sarcasm

. . .

1: Batch All Projects / Clients Into Gmail Labels — And Set Filters to Direct Inbound Mail

The first step is to make sure that every unique project — or client you're working with — has its own label in **G Suite**.

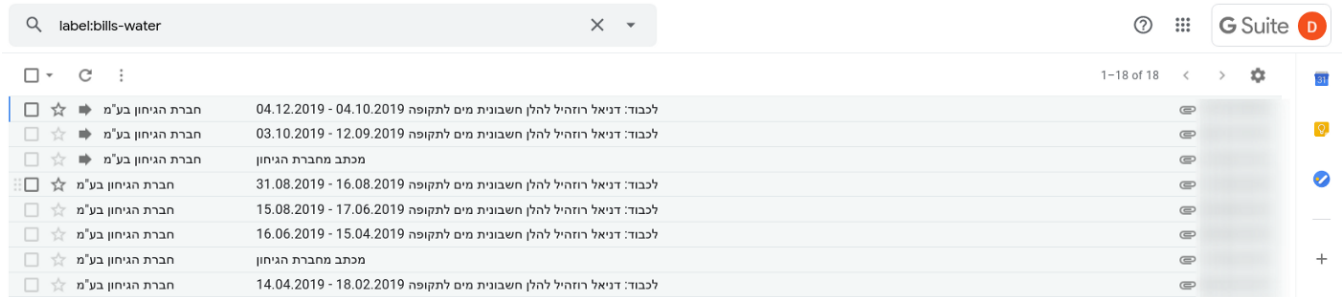


I'm pretty insane about this and have set up labels — and corresponding rules — for just about every conceivable purpose.

E.g.:

I have a label for my utilities (“Bills”) and under that sub-labels for power, my cellphone plan, my water bill, and my healthcare provider.

I have a label for my IFTTT “it’s going to rain today” notifications and another for correspondence from my accountant.



So, if you’re also a freelancer / small business owner, set up labels for your clients. (I have mine sub-divided into current clients, past clients, and prospective clients).

Next, I **create Gmail filters to automatically route inbound messages to the correct folder.**

Generally, the easiest way to do this is to filter by sending address, but you can also:

- **Filter by unique text strings in the subject line.** The “exact match” Boolean operator is useful for this purpose.
- **Filter by text strings in the body text.**

HaGihon delivers water to my apartment in Jerusalem. And they send their e-bills from the domain `printernet.co.il`.

So, I create a filter to route all email from that address into the Gmail label I created for water bills.

The asterisks (*) symbol denotes a wildcard — so email from any other address at that domain will also route to the folder:

From `*@printernet.co.il`

To

Subject

Includes the words

includes the words

Doesn't have

Size greater than MB

☐ Has attachment ☐ Don't include chats

Create filter Search

from:(*@printernet.co.il)

← When a message arrives that matches this search:

- ☒ Skip the Inbox (Archive it)
- ☐ Mark as read
- ☐ Star it
- ☒ Apply the label: Bills/Water
- ☐ Forward it to: Choose an address. [add forwarding address](#)
- ☐ Delete it
- ☒ Never send it to Spam
- ☐ Send template: Choose template...
- ☐ Always mark it as important
- ☐ Never mark it as important
- ☐ Categorise as: Choose category...
- ☒ Also apply filter to **18** matching messages.

Note: filter will not be applied to old conversations in Spam or Bin

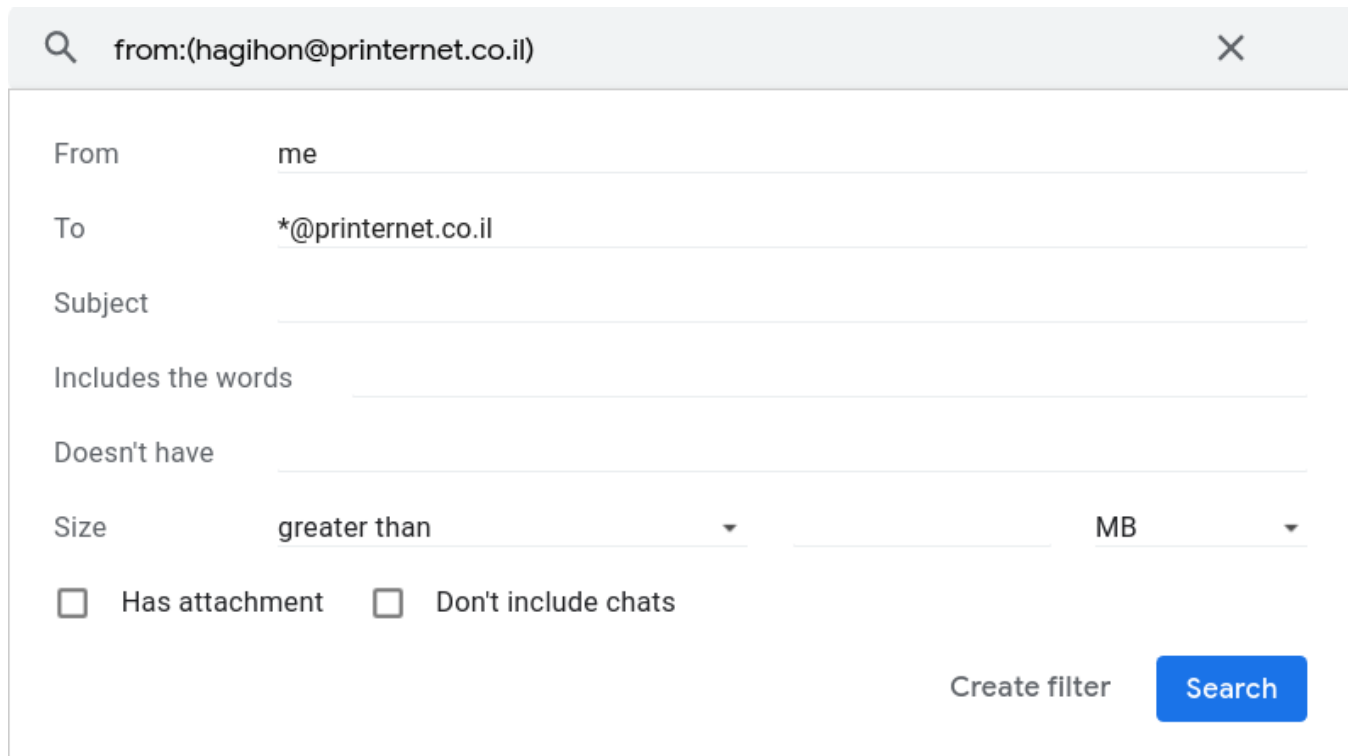
[Learn more](#) Create filter

I typically also tick the “Never send it to spam” box — because I know it’s not spam and I need to see it — so I can “whitelist” the domain and prevent it getting caught up in my spam folder.

And I tick the “also apply filter to X matching messages” box in order to get all existing messages from that sender into the folder.

If this were a client, then I would also want to capture all the outbound (sent email) in the label I just created, again making sure that I apply a wildcard operator.

So I would create a filter in the opposite direction — just swap ‘me’ for your email:

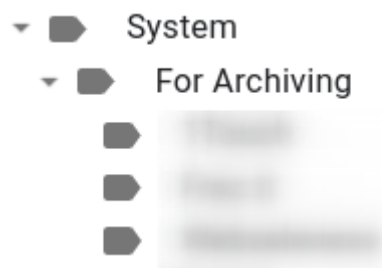
A screenshot of the Gmail 'Create a filter' dialog. At the top, a search bar contains 'from:(hagihon@printernet.co.il)' with a magnifying glass icon on the left and a close 'X' icon on the right. Below the search bar are several input fields: 'From' with the value 'me', 'To' with the value '*@printernet.co.il', 'Subject' (empty), 'Includes the words' (empty), and 'Doesn't have' (empty). There are also dropdown menus for 'Size' (set to 'greater than') and 'MB' (set to 'MB'). At the bottom left, there are two checkboxes: 'Has attachment' and 'Don't include chats', both of which are unchecked. At the bottom right, there are two buttons: 'Create filter' and a blue 'Search' button.

. . .

2: Create A Label for Archiving And Nest Old Labels There

Next, I set up label called ‘For Archiving’.

I’ve nested this under another label called ‘System’.

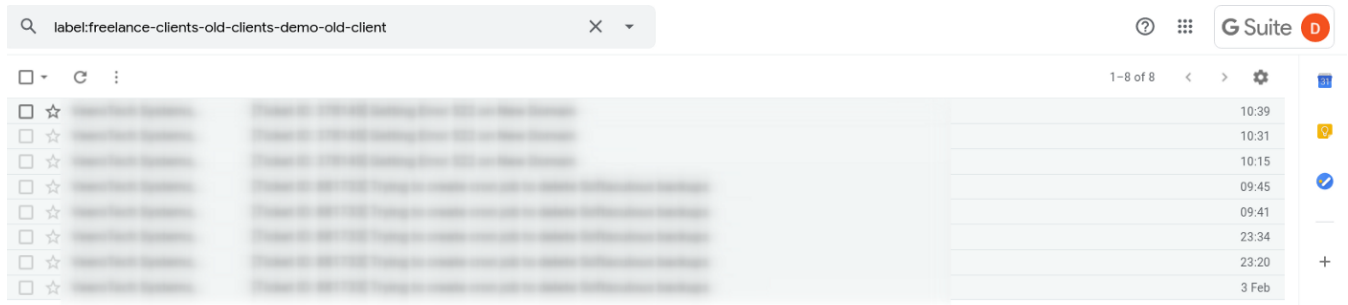


Let’s say I wrap up a project with a client.

I don’t really need or want to have all the back and forth with my point of contact cluttering up my inbox.

So I'll archive that label in my S3 bucket.

For the purpose of this demo, I batched a few emails I don't need under a label called 'Demo Old Client'.



I then moved the label under 'For Archiving'.

Edit label

Label name:

Demo Old Client

☒ Nest label under:

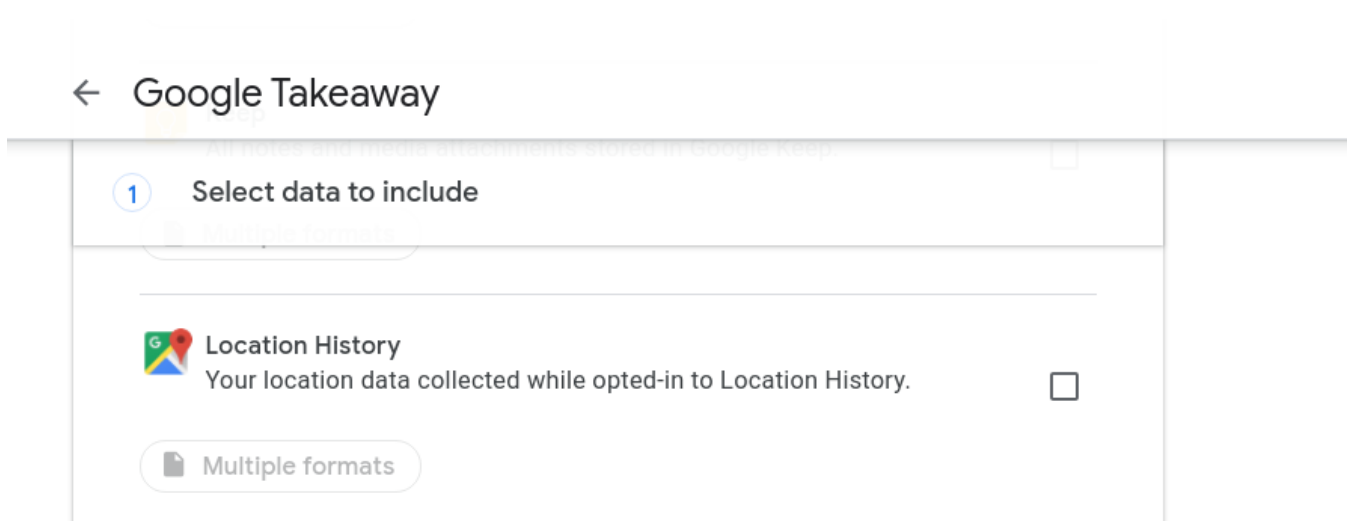
For Archiving

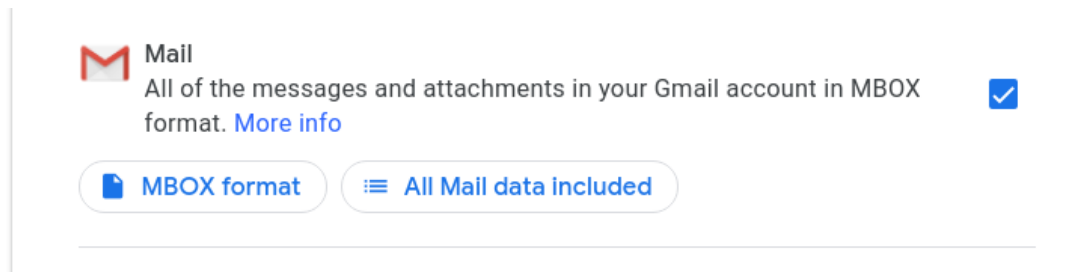
Cancel

Save

Now I want to initiate a Google Takeaway — but only capture the label that I want to archive.

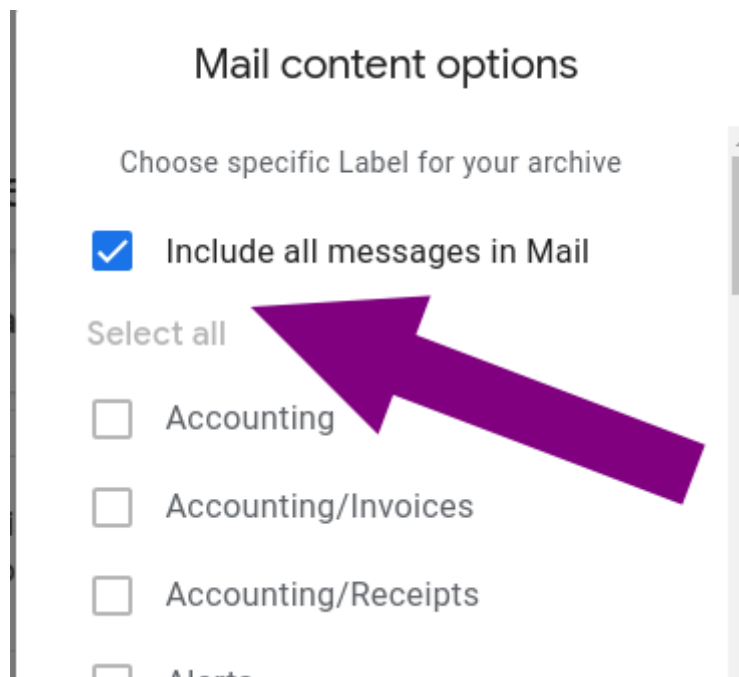
Firstly, deselect all GSuite services other than Mail.





Click on 'All Mail data included'.

And unselect "Include all messages in mail".

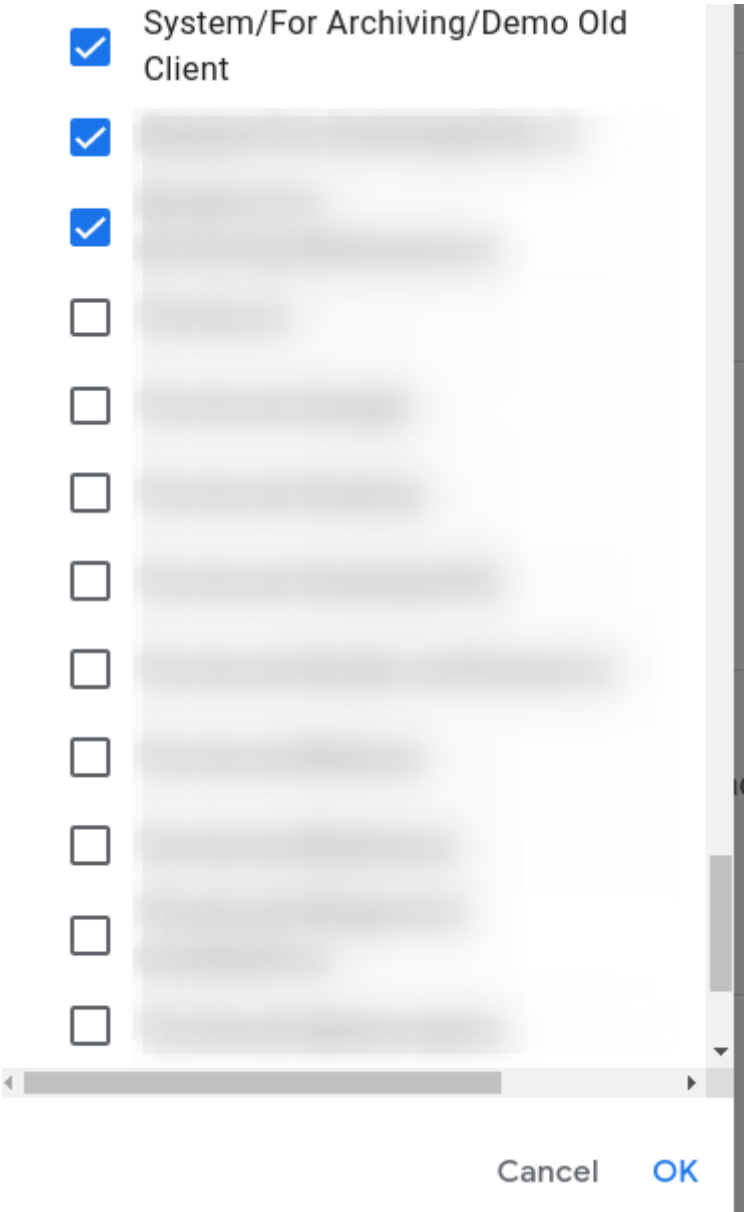


(FYI: despite this, the UI will still automatically select some folders that you probably don't want to delete. So be sure to deselect these if you want to keep them.)

- ☒ Sent
- ☐ Snoozed
- ☐ Spam
- ☒ Starred

Watch out. Labels like 'Sent' and 'Starred' are still automatically selected!

Just tick the old clients that you want to archive and click **OK** :



Unless all these messages come with large attachments, email is pretty light. So the download link shouldn't take more than a few seconds to generate.

The few folders I decided to export came to just 8 MB:

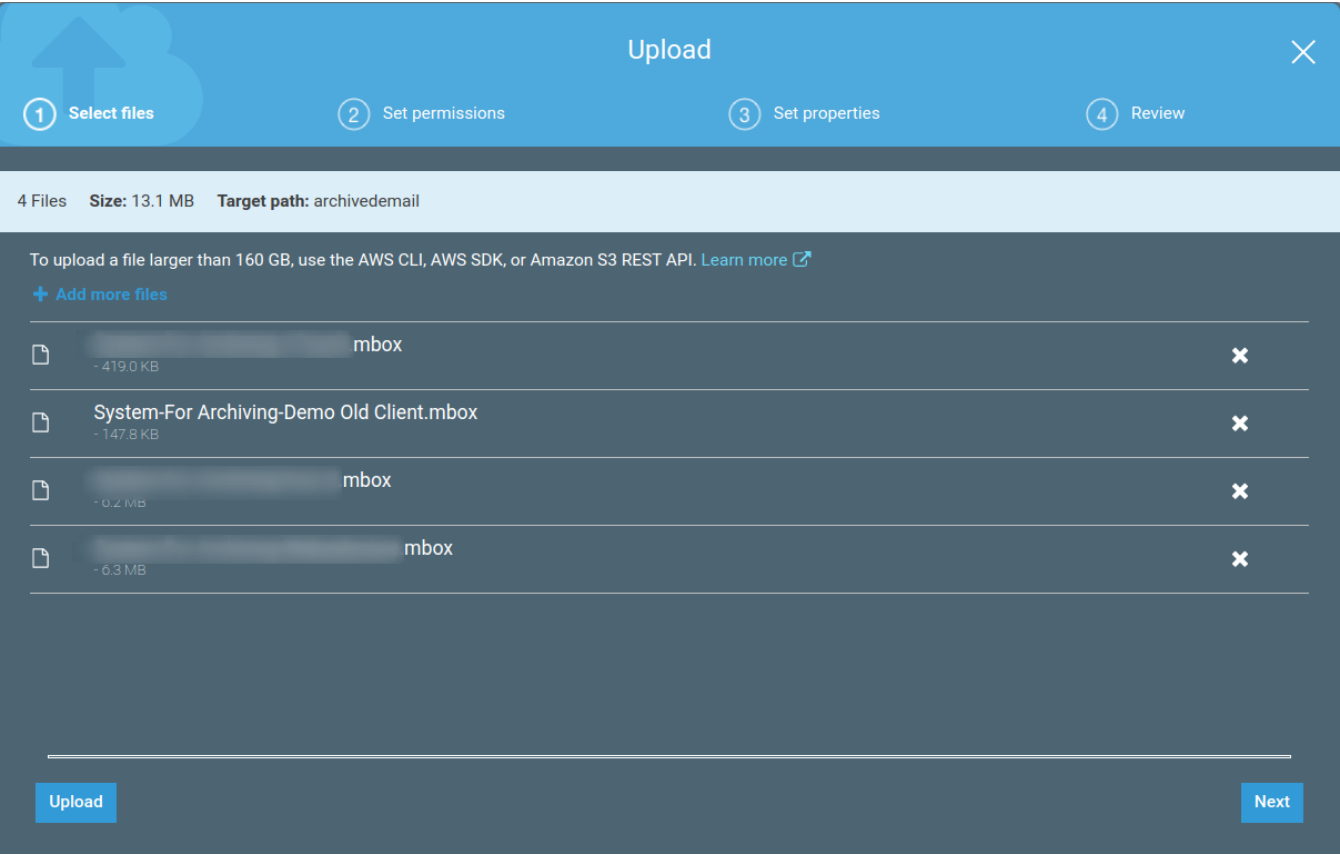
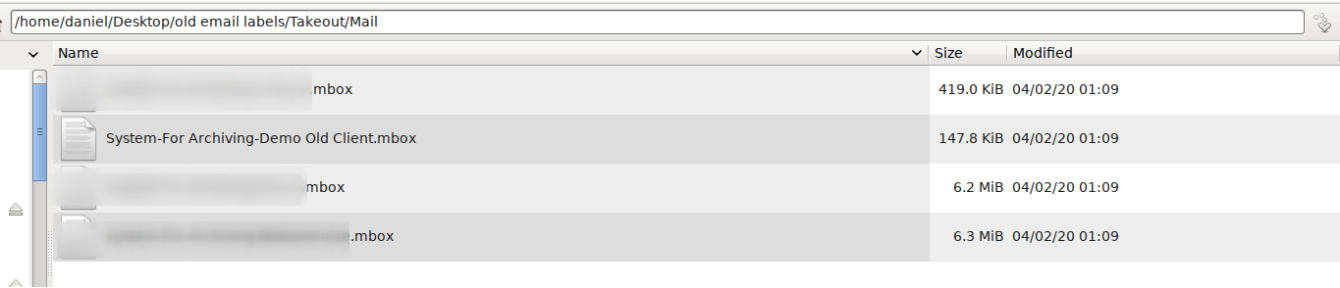
Export	Created on	Available until	Details
Mail 8 MB	4 February 2020	11 February 2020	Download ▼

Create new export

In AWS S3, I created a bucket called archivedemail.

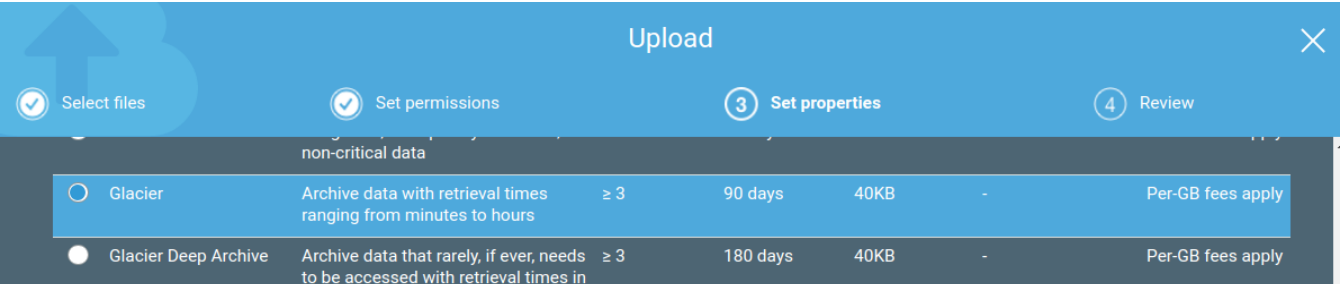
After downloading the Takeaway, you’ll want to unzip the archive and navigate a few folders into the directory to find the actual label archives.

These end with the file extension .mbox.



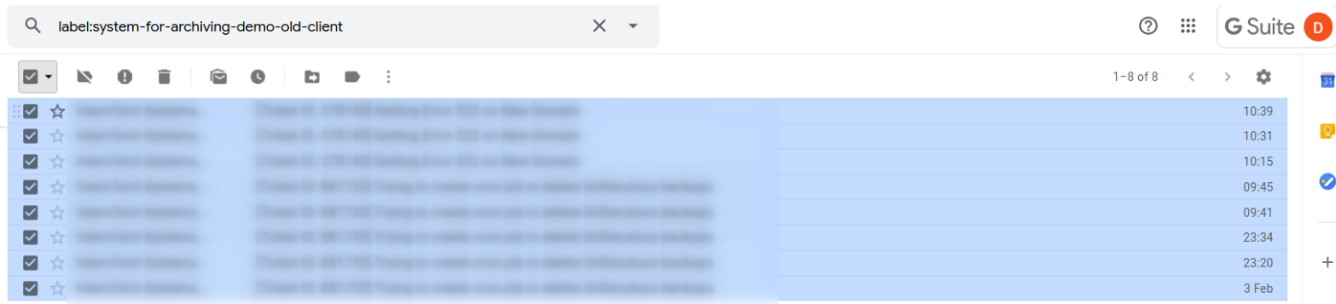
And that’s basically it.

I then select **Glacier** as the storage class because it’s unlikely that I will ever need to access the correspondence again:

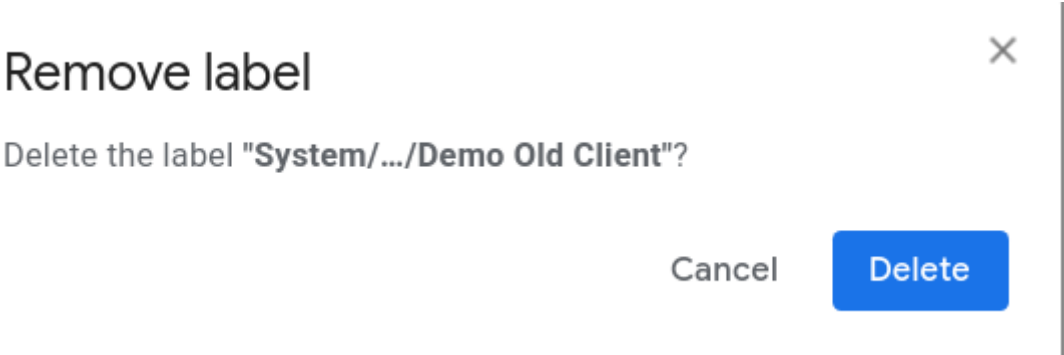


hours				
<div><div></div><div>Reduced Redundancy (Not recommended)</div></div>	Frequently accessed, non-critical data	≥ 3	-	-

Now that the emails are safely archived, I can delete all the messages from the labels in **G Suite**.



And finally, I'll delete the labels themselves:

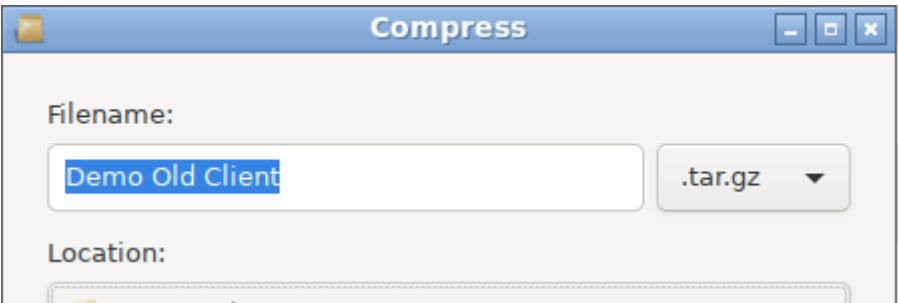


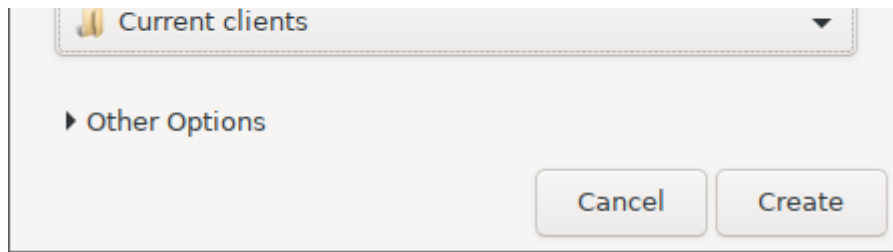
. . .

2: (Manual Method) To Do The Same Thing With Files

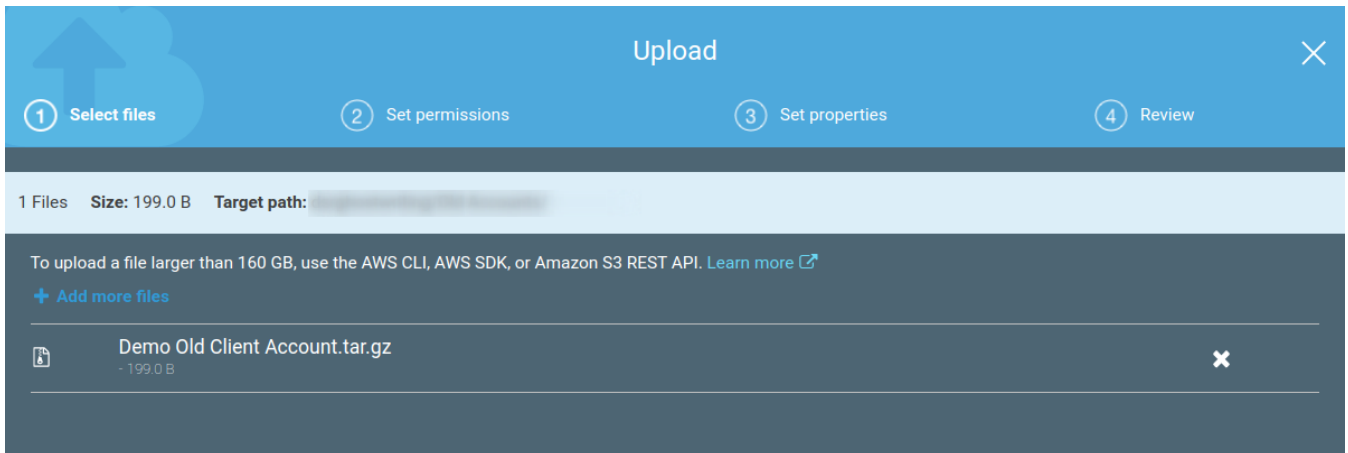
I'll repeat the same process with files.

I don't use Google Drive as my main cloud storage system but this process should work the same if you use Drive File Stream to mount your Google storage to the local filesystem.





Again, I'll push the file to an S3 bucket using AWS Console and set the storage class to something like Glacier.

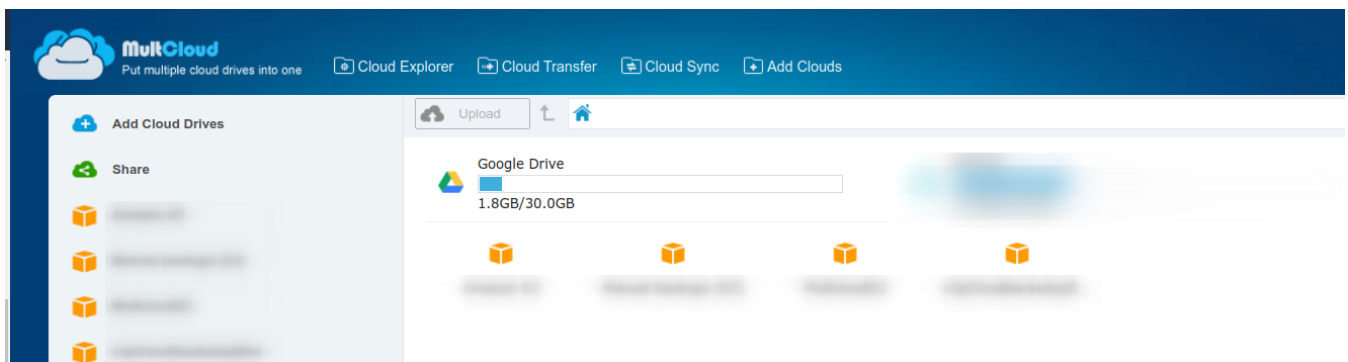


...

3: (Automatic Method) To Do The Same Thing With Files

Finally, I use a service called [MultCloud.com](https://multcloud.com) to automatically sync files I don't need to an S3 bucket.

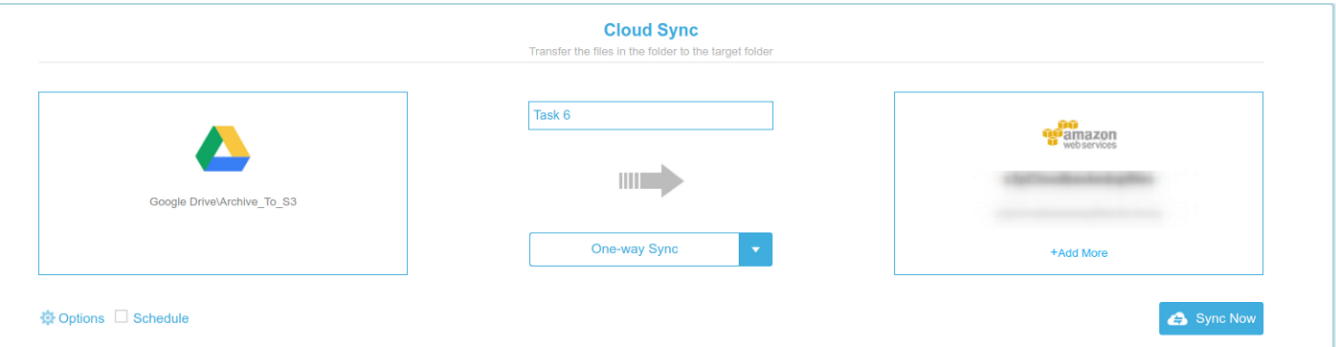
In order to do this, I'll attach my S3 bucket and cloud storage via Multcloud:



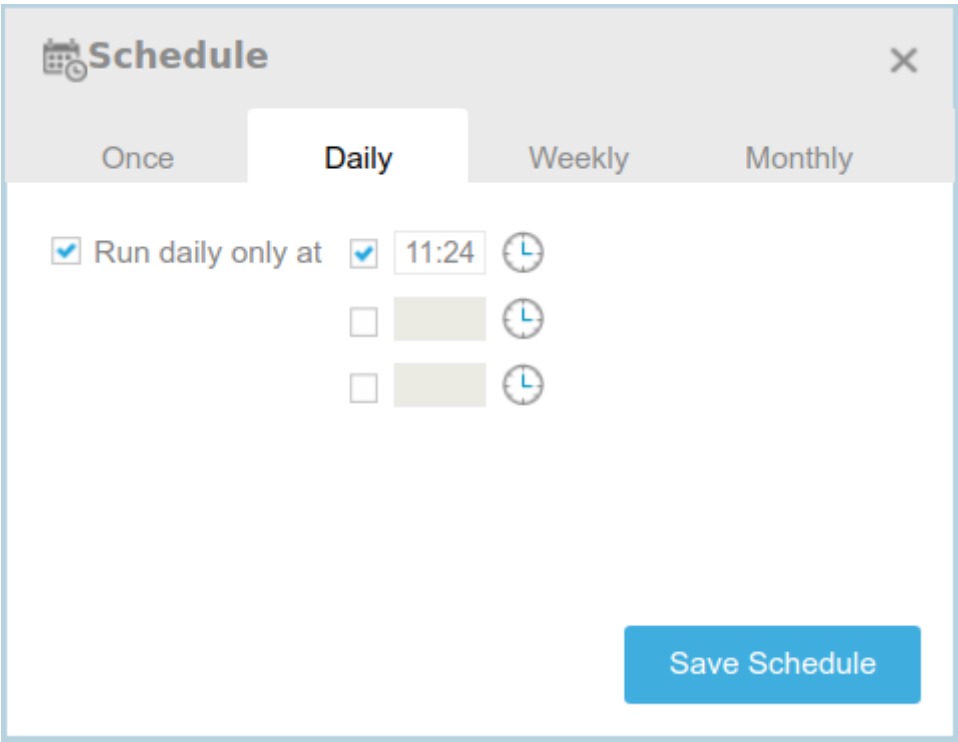
Then, I'll create a folder in my primary cloud storage called S3_Autoarchiver:



And then I'll create a sync between the “autoarchive” folder and my S3 bucket:



And then put this on a daily schedule:

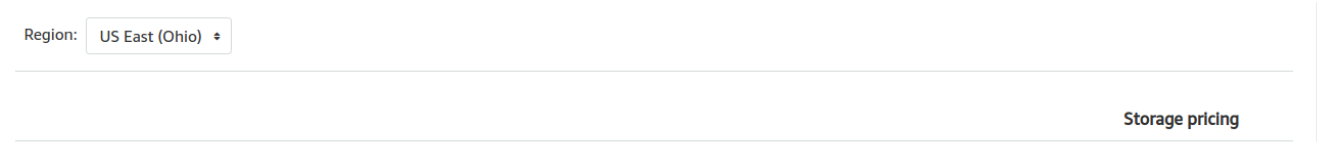


. . .

Use S3 To Keep Your Main Cloud Storage Uncluttered

AWS S3 is (almost) as affordable as object cloud storage gets.

At the time of writing, it's just \$0.023/GB for the first 50 TB of storage in the US East region with free data transfer in:



S3 Standard - General purpose storage for any type of data, typically used for frequently accessed data	
First 50 TB / Month	\$0.023 per GB
Next 450 TB / Month	\$0.022 per GB
Over 500 TB / Month	\$0.021 per GB

I’ve found periodically archiving all unneeded files and emails to it from Gsuite and my other cloud services a great way to keep within storage limits — and keep clutter and old projects from distracting me.

If you a main cloud storage system and have set up some storage space in AWS then you should give it a shot too!

[Gmail](#) [AWS](#) [S3](#) [Email](#) [G Suite](#)

[About](#) [Help](#) [Legal](#)

Get the Medium app

