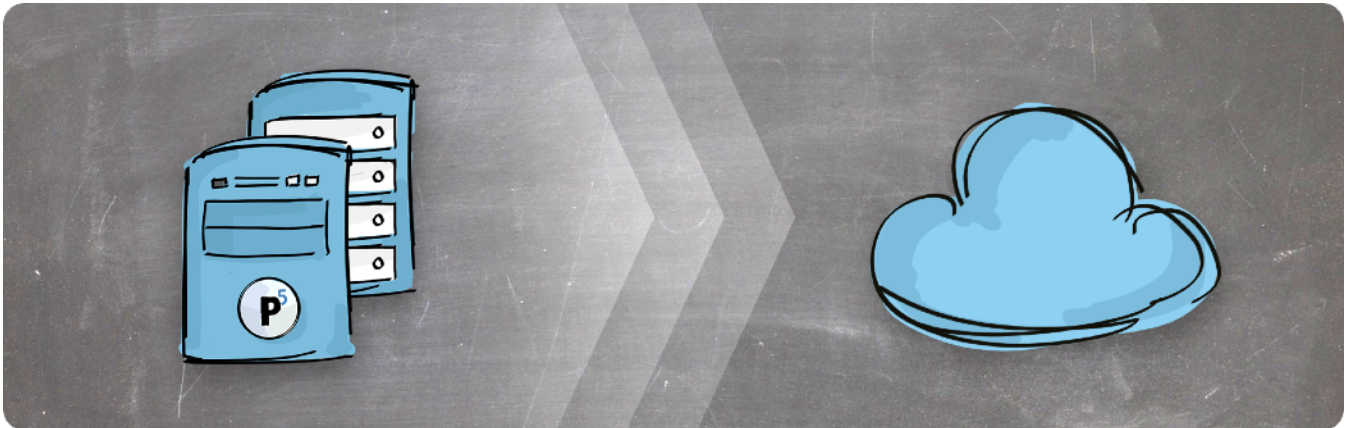# ARCHIWARE

Archiware P5: Technical Guide

P5 & cloud storage

Set up and best practise guide

With the release of Archiware P5 version 5.5, backup and archive to cloud storage are supported. There are various options and providers for cloud storage, this article aims to explain how this works and how to set it up.

P5's cloud storage support is based around a change to the format used for disk based volumes. This allows a backup or archive to first save data to disk-based volumes and then have those volumes copied to cloud storage. The local disk-based volumes can be optionally deleted automatically, or kept as a separate local copy to speed restores.

More guides like this one available here:

p5.archiware.com/manuals

P5 Backup/Archive are configured in the same way, regardless of whether tape, disk or cloud storage is being used. This article will use the P5 Backup module as an example, the same cloud storage configuration cal also be used for a P5 Archive workflow. We will assume that the basics of configuring a P5 backup plan is understood and we'll concentrate on how cloud storage is configured within the product.

Which cloud storage services are currently supported?

As of version 5.6.4 (May 2019):

- Amazon S3
- Generic S3 (allows third-parties such as Wasabi)
- Amazon Glacier
- Backblaze B2

What are the differences between those different services?

Amazon S3 is charged per GB per month and is a good general purpose option and probably the most popular commercial cloud storage product at the time of writing. Amazon S3 data is redundantly stored across more than one facility in a chosen geographical region. See https://aws.amazon.com/s3/faqs/. There's lot to read about S3, start by Googling for the AWS S3 pricing page to see what it will cost. Note that pricing differs depending upon the region (geographic) into which you wish to have your data stored.

Amazon S3 uses it's own https protocol and organisation of data as objects. There are open source implementations of the same protocol allowing you to build your own S3 storage box or use someone else's - e.g. https://minio.io/.

The Generic S3 option allows P5 to use this storage as well as the commercial S3 offering from Amazon.

Backblaze B2 is an S3 competitor and offers similar functionality, it's provides somewhat less redundancy and is based only in the US at the time of writing. It is considerably cheaper however, and comes from a company who have been providing their own cloud backup service for many years, and understand engineering cloud storage very well.

Finally, **Amazon Glacier** comes from the same stable as S3 but has some unique attributes which make it attractive for certain archive scenarios. Glacier is considerably (around one fifth in Jan 2018) cheaper than regular S3, costs are kept low by providing varying retrieval speeds, Glacier provides three options for access to data, from a few minutes to several hours. Recovery of files is performed by making a request and waiting for it to be executed. P5 will handle this process for you, but if you're likely to require your data back quickly, this is not the best choice.

How P5 volumes are able to work with cloud storage

Let's start looking at how P5 can utilise these modern cloud storage products. Before we look at this in detail though, let's first understand how disk-based volumes have changed in P5.5.

A disk volume is a container for data, just like an LTO-tape, that can be managed in P5. In versions prior to 5.5, a disk volume was a single container file, that resided on a filesystem local to the P5 installation. Multiple volumes are generally used together, so that, over time, a single volume can be recycled and re-used. Treating the total amount of storage in a granular fashion, recycling older data volumes as we continue to write new backups. Working in rotation through these disk volumes - just like a stack of physical LTO tape volumes.

The important change with P5.5 is that a single volume is no longer stored as a single file but instead, as a folder. Inside this folder you'll find a number of smaller files, called 'chunks', arranged inside a folder structure that P5 will manage for you. Don't concern yourself with what is contained within the volumes folder, there are no user-serviceable parts within!

These chunks allow smaller parts of the whole volume to be copied up to cloud storage, or downloaded, without the entire volume needing to be stored/retrieved as a whole. For example, you might configure your disk library to provide volumes that are 250GB in size, but the entire 250GB of storage need not be uploaded to cloud storage in one operation - that's simply not feasible, even with the fastest WAN links.

The size of these chunks of data within the volume can be configured when the pool defining the cloud storage. The default size is 128MB. Therefore, with our example 250GB volumes, almost 2,000 separate chunks will be required.

This change to the volume storage format on disk means that, as a job is running and writing to a volume, each time a chunk is 'filled' (in our example 128MB of data written), the chunk can be uploaded to your cloud storage and optionally be deleted from disk. The job will continue to read data from the source (e.g. machine being backed up), and write to chunk files within the volumes folder until it's complete. When the job finishes, one or more chunks will have been written to on local disk, and completed chunks uploaded to cloud storage and optionally deleted from disk.

It therefore follows that, when a file is to be restored from this cloud backup or archive, P5 will first look on disk to see if the required volumes chunks are available locally (in the case where the chunks are not deleted from local storage). If they are not available locally, the specific chunks required to perform the restore will be downloaded from the cloud, written locally and then the data will be restored from them back to disk.
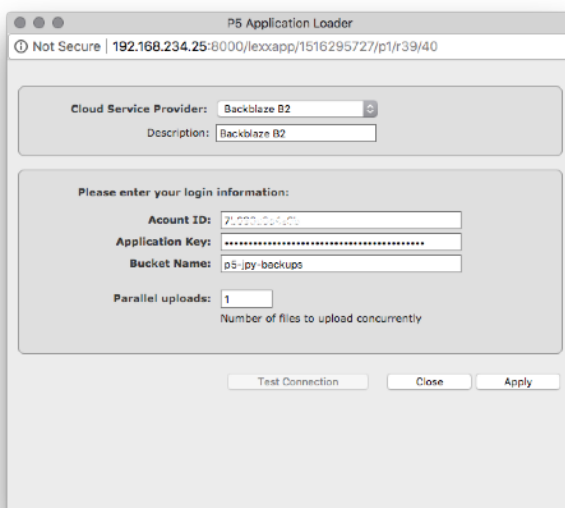
How to set up cloud Backup/Archive

As is the tradition with P5, there is a wizard interface that will configure this for you in a single step. This is accessed by visiting 'Cloud Storage' and then clicking on 'Add Cloud Storage' below. See screenshot below.
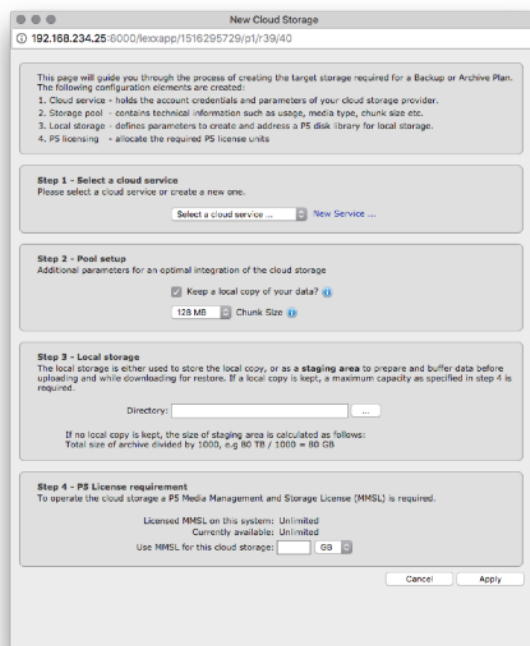
To understand what that wizard does and have the knowledge to allow you to tweak what it has done for you, it's best to understand how to manually set up what the wizard does for you - e.g. not use the wizard. Therefore I will explain how to set up the three components required to configure cloud services below.

1. Set up a cloud service

With either Backup or Archive selected in P5's main tab interface, depending on which you're using, click on 'Cloud Service' and then 'New Service' at the bottom of the browser window.

From here you'll be choosing which cloud storage provider (from S3, Glacier of Blackblaze) you wish to work with. You'll then enter the credentials required to allow P5 to access that service.

For example, with Amazon's S3 you'll enter:

- **Access Key ID and Secret Access Key** - Your credentials provided by S3
- **Bucket Name** - The container within S3 that you wish to use - you create this in S3
- **Parallel Uploads** - How many separate uploads P5 should simultaneously allow. More uploads can better utilize your WAN connection.

The set up of other services will be similar. Please familiarise yourself with the cloud storage service so that you're able to generate credentials and create 'buckets'. E.g. for Amazon S3, you'll need to create yourself an account and provide a credit card for payment. Both S3 and Backblaze offer a generous amount of free GB's of storage for free - plenty for testing with.

Once you've entered all the details, save them with the 'Apply' button and test that they work by clicking on 'Test'.

The web interfaces provided by the various cloud storage products allow you to browse the storage to see what P5 has written there and manually delete data if necessary. GUI tools for your OS of choice are available for browsing S3 buckets.

2. Create a pool for your cloud volumes

Now that we have somewhere for our 'chunks' to be stored, we create a pool to allow volumes to be created and organised locally. We connect this pool to our cloud service so that the volumes 'chunks' will be uploaded to the cloud service and optionally deleted from local storage.
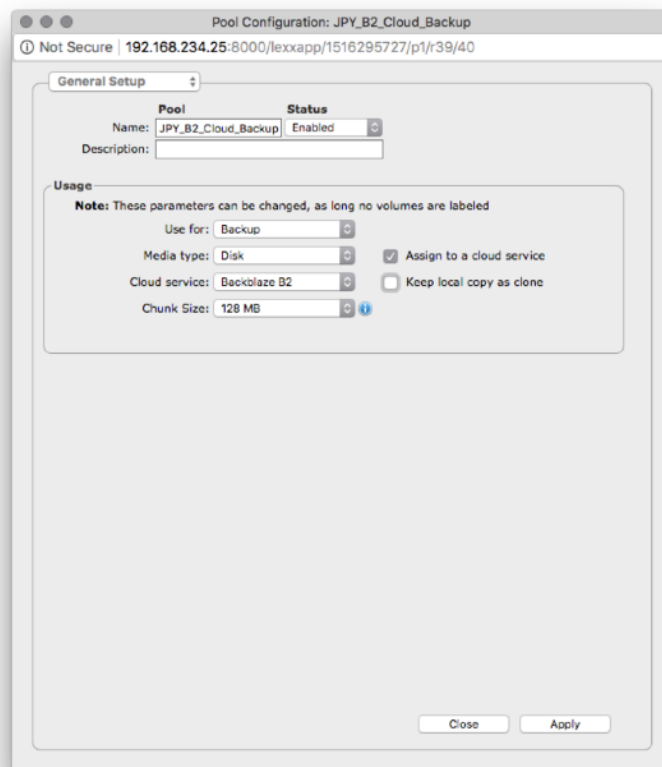
Each volume you later create for a given cloud backup/archive must reside within a pool. You can create more than one pool to divide volumes into different groups for different tasks.

Visit 'Pools' and click 'New' to create a new pool - see example to the right. Provide a name for your pool, specify if it will be used for Backup or Archive and click the button to assign this pool to a cloud service. Now select the cloud service that you created and tested in step 1 above.

Now select a different chunk size from the default of 128MB if you wish. Note that there's a help button here which states the following regards this option:

> *For data management reasons, disk space is partitioned into volumes. Each volume consists of chunk files. This optimises the data traffic between the local and the cloud storage. Increasing the chunk size is useful if it saves time or money during uploads or downloads. For example, the standard restore from Amazon Glacier takes a long time to respond to each request and retrieve a chunk, so it is better to use larger chunks.*

The default size is recommended for everything except Glacier where there is an overhead for retrieving each chunk, in this case you may wish to increase the size from 128MB upwards. In almost all cases - leave this at the default. I'll write more here in future, or a separate article, about this option.

Finally, if you wish for your backup/archive to exist only on the cloud storage, un-check the 'keep local copy as clone' box. As explained above, this causes each chunk to be removed from local storage after it has been successfully uploaded to your cloud service.

3. Create Disk Library to allow volumes to be created

A disk library is always required to write volumes to disk - even if you're not using cloud storage. Therefore you may already have a disk library in use in your installation.

For the purposes of writing to disk for cloud storage, we recommend creating a new disk library. It isn't essential but does keep things separated for ease of understanding and management.

Therefore, under the 'Storage Manager' area of P5's configuration, click on 'New Disk Storage'. Remember that, depending on your choices above, this area on disk might only be used for storing your volumes chunks temporarily. If you've elected not to keep a local copy, you'll not need a lot of storage to accommodate this disk library - one of the benefits of using cloud storage!

Provide the local directory for volume storage and the total amount of cloud storage space you wish to address. Cloud storage is effectively unlimited, you pay for what you use. The limit you set here will be the total amount of cloud storage space P5 will be able to address. If you're using this for backups, the process of recycling and reusing older volumes will be used, as with tape, to limit the amount of storage space used.

For archive, choose a limit large enough for accommodate your storage needs for the life of the archive. Note that the amount of space you specify here is a future limit. The space will not be pre-allocated, either on disk or in the cloud. The limit you set will be used against the license you have installed on your installation however. Each MMS (slots) license you have will provide 4TB of addressable storage across all the disk libraries in this installation of P5.

Finally, choose 'do not label' for the final option in this window. This allows us to have control over this in the next step. Click Apply to create the disk library.

Close the disk library set up window and double click on the disk library again to view it's settings. Here you can tweak the total number of volumes and the total size of each of them. These two figures multiplied together will equal the total space you requested in the previous step. You may leave them as they are, but now is your chance to change these settings - perhaps you would like smaller volumes but more of them. Again, apply and close this window.

4. Labelling volumes

Highlight your newly created disk library and click the 'label' button at the bottom of the browser window.

You will now see an overview of the 'slots', one per volume, available in your disk library. Select a range of these slots depending on how many volumes in a single pool you wish to create in this process. Click on the first slot and shift-click on the last to select a range.

Next, choose the Pool into which these volumes will be labelled. This will be the pool you created in step-2 above. If you're re-labelling a volume that already exists in the selected slot(s), select 'erase or recycle' - otherwise leave the 'exceptions' setting on the default, 'skip' setting. Be careful here because if you re-label a slot that already contains a volume, that volumes data will be deleted from disk and cloud and the files stored on that volume deleted from the index.

Click the 'Label' button to commence the labelling process. For each volume to be created, P5 will create the volume locally on disk and initialise some chunks. These chunks will be copied to your cloud storage. Initially, before you write and real data to the volumes via a Backup or Archive plan, these volumes will use very little space both on disk and on your cloud storage.

You can use the job monitor window to observe this labelling process and see when it is completed. Upon completion, you will be able to see the volumes that you just created in the 'Volumes' section of the admin interface. Since, as part of the labelling process, chunks are initialised in the cloud storage, this step may take some time.

Once the labelling process has completed, you'll be able to view folders within your cloud container that match the volume ID's that have been created within your P5 installation. The chunks that comprise the data stored in each volume are contained within a folder structure that P5 manages.

P5 will always keep the volumes consistent with the cloud storage. An alert symbol will be shown against any cloud volume that is not currently 'synced' with cloud storage. You can right-click on these volumes and chose to 're-sync cloud volume..' to force the copy operation to re-run.

Conclusions

You have now configured your cloud storage for either a Backup of Archive workflow - depending upon how you configured the pool.You may now create your Backup or Archive plan and point it to your cloud pool. Running the job will then write data to the cloud according to your set up. There are a few different moving parts here so you might find things don't work first time and be required to troubleshoot.

Run your job, observe which volume is written, see how much space that volume is using on disk compared with how much data the volume list reports is stored in it. Use a web interface of GUI tool browse your cloud storage and compare with above to satisfy yourself that it's all working as expected. Finally - restore data - always the best test!