



Google Workspace Plugin

Bacula Systems Documentation

Contents

1 Overview	3
2 Scope	6
3 Architecture	8
4 Services and Features	10
5 Special features	28
6 Installation	32
7 Configuration	33
8 Operations	56
9 Best practices	67
10 Troubleshooting	69

Contents

- *Overview*
- *Scope*
- *Architecture*
- *Services and Features*
- *Special features*
- *Installation*
- *Configuration*
- *Operations*
- *Best practices*
- *Troubleshooting*

1 Overview

This white-paper presents how to protect the most relevant elements of Google Workspace services using **Bacula Enterprise**.

1.1 Features

The **Bacula Enterprise Google Workspace Plugin** is a very easy to deploy and configure plugin supporting the following services:

- Google Drive
- Google Mail

It is shipped with advanced concurrency, resiliency, and flexibility features in addition to covering the most relevant Google Workspace backup use cases. A full feature list is presented below:

- Common features
 - Google Workspace APIs based backups
 - Support for free Gmail accounts
 - Support for accounts under a Google Workspace subscription
 - Multi-service concurrency capabilities
 - Multi-threaded processes
 - Advanced tuning configurations
 - Automatic concurrency of fetching processes
 - Generation of user-friendly report for restore operations
 - Network resiliency mechanisms
 - Latest Google Authentication mechanisms
 - Discovery/List/Query capabilities
 - Restore objects to Google Workspace
 - * To original entity
 - * To any other entity
 - Restore any object to file-system
 - Restore HTML report to user mailbox or user drive
- Backup and Restore of Google Drive
 - Backup and Restore of Users My Drive
 - Backup and Restore of Shared Drive Units
 - Hash check during backup and restore to ensure data integrity
 - Incremental & Differential backup
 - * Includes advanced delta function for improved performance
 - Advanced selection capabilities
 - * Include/exclude by name

- * Automatic discovery to backup everything
- * Include/exclude by RegEx
- * Folder selection capabilities for backup
 - Include/exclude by name
 - Automatic discovery to backup everything
 - Include/exclude by regular expressions
- Support for regular files and also native Google Workspace files (export)
- Folder and file granularity for restore
- Computed hash check at backup and restore time
- Backup and restore of permissions shares
- Backup and restore of shared elements with users
- Backup and restore of Google Drive file versions
- Backup and restore of file comments
- Backup and restore of trash
- Backup and Restore of Google Mail (GMail)
 - Backup and Restore of email messages
 - * Messages metadata
 - * Messages content
 - Backup and Restore of attachments
 - Backup and Restore of mailbox settings
 - * Auto-Forwarding, Imap, Language and Pop settings
 - * Delegates
 - * Filters
 - * SendAs addresses
 - * Forwarding addresses
 - Incremental & Differential backup with Delta function
 - * Includes advanced delta function for improved performance
 - Advanced selection capabilities
 - * Include/exclude users by name
 - * Automatic discovery to backup all Workspace users
 - * Include/exclude users by RegEx
 - * Label selection capabilities for backup
 - Include/exclude by name
 - Automatic discovery to backup all of them
 - Include/exclude by regular expressions
 - Export mail messages to mime RFC 822 local files

- Export attachments to local files
- Restore to original GMail mailbox
- Restore to a different user GMail mailbox
- Restore to the original labels
- Restore to a specific label
- Fully indexed information into Bacula Catalog
- Advanced search capabilities for restore operations
- Privacy excluding features:
 - * Ability to exclude message fields from the index
 - * Exclude private or spam messages through powerful filtering capabilities

Note: Future modules

Bacula Google Workspace Plugin will include more modules in the future, like Google Calendar among others.

1.2 Requirements

Bacula Google Workspace Plugin supports free Gmail accounts and Workspace accounts.

In order to protect Workspace accounts it is needed to have a Google Workspace active subscription: <https://workspace.google.com/intl/es-419/pricing.html>

On the other hand, it is necessary to have **full administrative access to the target associated Organization** to protect in order to generate a Google Application with all the needed permissions that will be used to communicate with this plugin.

In order to protect free accounts it is just needed to prepare some configurations in Google Cloud Platform, logging in with the user to protect, before using the plugin. Please refer to the authentication section of this document to have further details.

Currently, the plugin must be installed on a Linux based OS (RH, Debian, Ubuntu, SLES ..) where a Bacula Enterprise File Daemon is installed. *Bacula Systems may address support for running this plugin on a Windows platform in a future version.*

The OS where the File Daemon is installed must have installed Java version 11 or above.

Memory and computation requirements completely depend on the usage of this plugin (concurrency, environment size, etc). However, it is expected to have a minimum of 4GB RAM in the server where the File Daemon is running. By default, every job could end up using up to 512Mb of RAM in demanding scenarios (usually it will be less). However, there can be particular situations where this could be higher. This memory limit can be adjusted internally (see *Out of Memory*). Refer to the Scope section below for any service specific requirements.

1.3 Why protecting Google Workspace?

This is a common question that arises frequently among IT and Backup professionals when it comes to SaaS or Cloud services, so it is important to clearly understand it.

It is a fact that Google or any cloud provider offers some capabilities intended to prevent data loss such as:

- Usually, all data stored in cloud services is geo-replicated using the underlying cloud infrastructure to have the information stored into several destinations automatically and transparently. Therefore, complete data loss because of hardware failures are very unlikely to happen.
- Google Data Loss Prevention service: This is a policy based service capable of detecting filtered content and act upon it encrypting it or modifying it in order to protect it (remove headers, etc). This is not a backup tool, it is a service to prevent undesired actions to the content stored in Google Workspace (for example sharing confidential information with the wrong people).
- Retention policies of Google Workspace: Google retains a maximum of 30 days of deleted information from active subscriptions. Therefore it is possible to recover accidental deleted items inside that period.

There is no other data protection mechanism. Below we show a list of challenges that are not covered by cloud services:

- No Ransomware protection: If data suffers an attack and becomes encrypted, data is lost.
- No malicious attacker protection: If data is deleted permanently, data is lost.
- No real point-in-time recovery, and recoveries of partially deleted files are limited to 30 days.
- It is not possible to align data protection of Google Workspace services to general retention periods or policies longer than 30 days.
- No automated way to extract any data from the cloud to save it in external places (this could lead to eventual compliance problems)

2 Scope

Bacula Enterprise Google Workspace Plugin is applicable on environments using any Workspace subscription.

This paper presents solutions for **Bacula Enterprise** version 14.1 and later, and is not applicable to prior versions.

Note: Important considerations

Before using this plugin, please carefully read the elements discussed in this section.

2.1 Empty files

In general, empty files (files with 0 byte contents) are simply not backed up by Google Workspace plugin. In particular, Google Drive files will show a message in the joblog to inform about empty files detected and so not processed.

2.2 Files and objects spooling

In general, this plugin backups two types of information:

- Objects
- Files

Objects are elements representing some entity in Google Workspace such as a files metadata.

While objects are directly streamed from memory to the backup engine, files need to be downloaded to the FD host before being sent. This is done in order to make some checks and to improve overall performance, as this way operations can be paralleled. Every file is removed just after being completely downloaded and sent to the backup engine.

The path used for this purpose is established by the 'path' plugin variable, that usually is set up in the gw_backend script with the value: /opt/bacula/working

Inside the path variable, a 'spool' directory will be created and used for those temporary download processes.

Therefore, it is necessary to have at least enough disk space available for the size of the largest file in the backup session. If you are using concurrency between jobs or through the same job (by default this is the case through the concurrent_threads=5 parameter), you would need at least that size for the largest file multiplied by the number of operations in parallel you run.

For emails it is important to note that download operations are done in one step because of some API requirements. This means the jvm should have enough memory to load those downloaded files inside RAM. In case you suffer any memory issue, please refer to the troubleshooting section to find out how to increase it.

2.3 Accurate Mode and Virtual Full Backups

Accurate mode and Virtual Full backups are not supported. These features will be addressed in future versions of this plugin.

2.4 Google Workspace APIs General Disclaimer

Google Workspace APIs are owned by Google and they can change or evolve at any time. Almost all service APIs are actively developed, containing new features every week, even if the version number of the service is not changed as a result of any of those additions. Just as an example, Google Drive API now is tagged as v3 (and this plugin is using that version to work).

This situation is significantly different from traditional on-premise software, where each update is clearly numbered and controlled for a given server, so applications consuming that software, can clearly state what is offered and what are the target supported versions.

Google is committed to try not to break any existing functionality that could affect external applications. However, this situation can actually happen and therefore, cause some occasional problems with this plugin. Bacula Systems controls this with an advanced automatic monitoring system which is always checking the correct behavior of existing features, and will react quickly to that hypothetical event, but please be aware of the nature and implications of this kind of cloud technologies.

3 Architecture

Bacula Enterprise Google Workspace Plugin is using several Google Workspace APIs to perform almost all of its operations. Therefore, the plugin is working at the maximum granularity that the service provides.

Google Workspace developer products

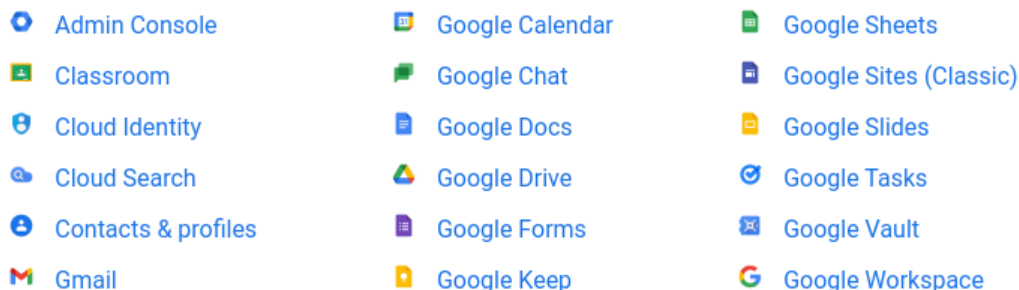


Fig. 1: Google Workspace APIs

All the information is gotten using HTTP requests to Google Cloud from the FD where the plugin is installed.

The plugin will contact a **Google Cloud Platform application** that needs to be **manually created and configured** before using the plugin. It will serve as a bridge to download the required data or objects during backup time and send them to the Storage Daemon. Conversely, the plugin will receive them from an SD and perform uploads as needed during a restore operation.

The implementation is done through a Java Daemon, therefore Java is a requirement in the FD host. For more information about how to create the application in GCP, please, consult auth section.

Below is a simplified vision of the architecture of this plugin inside a generic **Bacula Enterprise** deployment:

Listed below is the information that can be protected using this plugin:

- Google Drive
 - My Drive of users
 - * Folders
 - * Native Google services files (gdocs, gslides, gpresentation.. Export and download)
 - * All other files (regular download)
 - * File Versions
 - * Trash bin
 - Shared drives
 - * Folders
 - * Native Google services files (gdocs, gslides, gpresentation.. Export and download)
 - * All other files (regular download)
 - * File Versions
 - * Trash bin

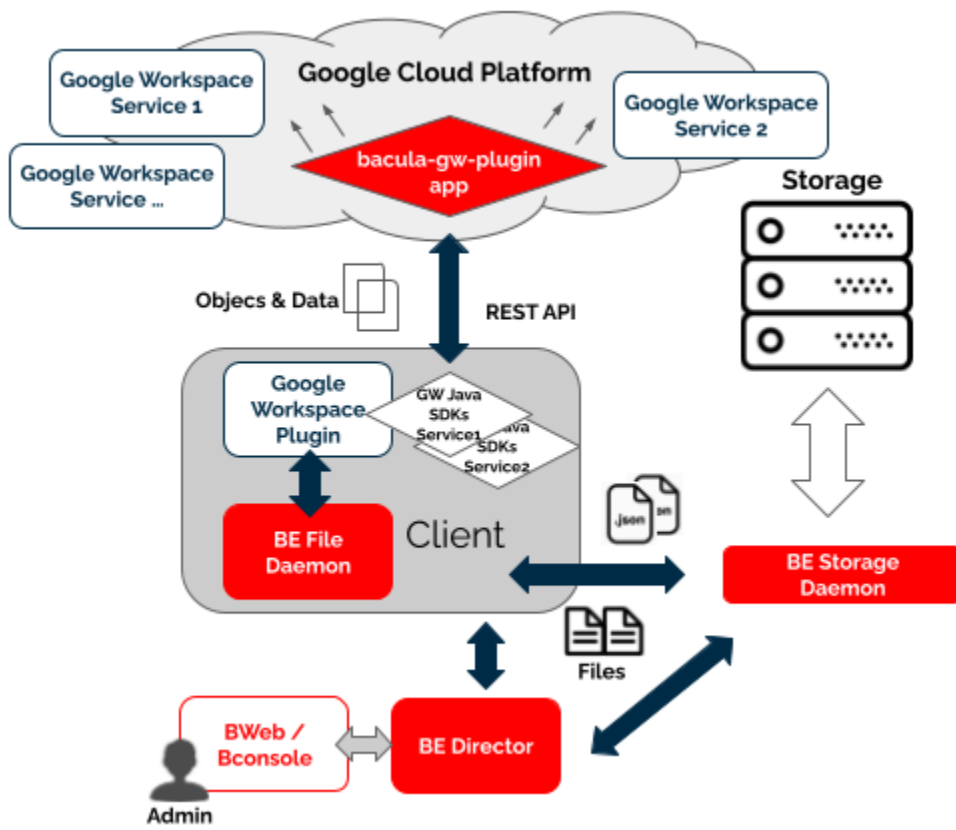


Fig. 2: Google Workspace Plugin Architecture

- Shared permissions (direct access, share links, expiration times..)
- SharedWithMe User files
- Files comments
- Google Mail
 - Mailbox user Labels
 - * System labels: Inbox, Sent, Draft, Spam ...
 - * User labels
 - Mailbox user Mails
 - * Metadata
 - * Contents
 - Mail Attachments
 - Mailbox user Settings
 - * Auto-Forwarding settings
 - * Imap settings
 - * Language settings
 - * Pop settings settings
 - Delegates addresses
 - Filters
 - SendAs addresses
 - Forwarding addresses

All the metadata information of each object is stored in JSON format preserving all their original values.

4 Services and Features

In this section we will dig into how this plugin behaves for each particular service, describing special features and behaviors that require an extended description.

4.1 Google Drive

Bacula Enterprise Google Workspace Plugin can protect My Drive units associated to users from a workspace, My Drive units of free accounts, as well as Shared Drive units.

It is possible to utilize advanced selection methods to decide exactly what is backed up, as well as control precisely which items to restore and their destinations.

The detailed list of the information protected with this service is:

- My Drive of users
 - Folders
 - Native Google services files (gdocs, gslides, gpresentation.. Export and download)
 - All other files (regular download)

- File Versions
- Trash bin
- Shared drives
 - Folders
 - Native Google services files (gdocs, gslides, gpresentation.. Export and download)
 - All other files (regular download)
 - File Versions
 - Trash bin
- Shared permissions (direct access, share links, expiration times..)
- SharedWithMe User files
- Files comments

Files will keep their names in the catalog and will be included in a path like this:

- `/@gw/customerId/entitykind/entityname/drives/unitname/path/to/file/name-file.extension`
(where `entitykind` can be `users` or `shared_drives`)

Version History

Google Drive can be configured to *retain the history for files/items*.

Google Drive hash check

Google Drive service stores a hash for every file hosted, using MD5 algorithm. Bacula Enterprise Google Workspace Plugin calculates this hash and compares it to ones stored in the cloud at backup time, and also at restore time in order to ensure data integrity. Debug mode shows information about these hashes. Please note that this is true only for non native Google files

Google Drive duplicated files

Google Drive stores its information in a different way compared to traditional filesystems. Instead of a tree structure, everything in google drive are pairs of keys and values (data maps). This makes some internal differences regarding the data structure and, for example, it is possible to have the same folder with the same name inside the exact same path. Similarly it is possible to have the same file with the same name several times in the same path.

Bacula Enterprise Google Drive plugin will combine the data inside folders with the same name. This is:

- **From Google Drive:**
 - mypath/**
 - DirA/ f1 f2**
 - DirA/ f3 f4**
- **To Bacula Catalog:**
 - mypath/DirA/ f1 f2 f3 f4**

For files with the same name, bconsole will only show one and will restore the last one when using common options to find the most recent backup or a specific job id where the same file is present more than once. However, for example when using BWeb it is possible to see the different versions of each file and get the desired one using the restore Wizard.

Google Apps files

Files that come from Google online services like Google Docs, Google Spreadsheets or Google Slides are exported transparently during the backup process to open formats.

These kind of files do not expose versioning through Google APIs, so `drive_version_history` will not take any effect on them. The specific list of files affected by this behavior is as follows:

- Google Docs to odt
- Google Draw to jpeg
- Google Photo to jpeg
- Google Sheets to ods
- Google Slides to odp
- Google Scripts to json

Google Photos and Google Sites

Historically, Google Photos and Google Drive have been very close modules. However today they work with separated APIs. Bacula Workspace Plugin is not supporting Google Photos, even if it is planned to support it in the future as a different module of this plugin.

Google Sites do not support export functions so the plugin cannot protect them. Associated files are simply ignored.

Google Drive shares

Bacula Enterprise Google Drive Plugin is able to backup and restore shared elements. These kind of elements require a special treatment, as they are composed of two parts:

- In the source account, shared elements include special information about the permissions of the share (who and how the share must work)
- In the destination account, shared elements appear within an special category called 'SharedWithMe'.

Shared permissions

Bacula Enterprise Google Workspace Plugin will query for the permissions of an item if this item has been shared directly. This means the plugin will not backup inherited permissions. In order to have inherited permissions in a backup, the top element where the original shared permissions were set needs to be included in the backup and in the restore. As an example, if a directory is shared, but we restore only specific files contained in it, those files will not be shared as they were originally. It is necessary to restore the whole directory in order to replicate the original inherited permissions as they were at the time of backup.

The method to store shared permissions is to include them as 'metadata' of every file. This implies that permissions can only be restored directly to the Google Workspace service. A File Daemon restore to a local filesystem will only restore files, and shared permissions will not be restored.

Shared permissions can include links. Links are pre-generated URLs that can include expiration dates and other configuration parameters such as scopes, types, or affected identities. Shared permissions restores have special characteristics that must be considered and they are described below:

- Permission will be generated exactly as it was, but it will be a new permission object. This is similar to the situation with files. A restored file has the same contents as the original, but can include slightly different metadata because creation process was different.
 - If the permission had a static link, a link will be generated, but the associated URL will be different from the original.

Shared permissions are not restored by default. You need to enable the option 'drive_restore_shared_permissions' during the restore session.

Shared with me

SharedWithMe elements of each target account, if included in the FileSet, are backed up in a predefined directory called SharedWithMe inside the top folder of every selected account. For example, for a given account `youraccount@yourdomain.com` in a workspace called `customer_id`:

```
:caption: **SharedWithMe**  
``/@gw/customer_id/users/youraccount@yourdomain.com/drive/my drive/sharedWithMe/``
```

At restore time, sharedWithMe elements are treated as any other regular file. However, it is important to note that sharedWithMe files, as we are in the receiver account, have no sharing permissions.

The plugin has a special parameter at restore time allowing it to skip sharedWithMe elements even if they are selected. This feature is intended to facilitate full restores where source and destination accounts are included. Please, note that a restore of a source account with share elements will present those elements to any receiver account if you enable the option to restore share permissions, as we have discussed in the upper section.

Please, go to the Configuration section of this document to see how to set up the sharedWithMe skip option.

Backup parameters

The list below shows the specific backup parameters that can be set up in order to control the behavior of the drive module.

In order to select the Google Drive module, the common *service* parameter must be equals or be containing the value *drive*.

Entities that can include one drive units are: users, groups or sites.

Option	Re- quired	De- fault	Values	Ex- ample	Description
drive_shared_units	No		Valid names of existing shared drives on the selected workspace separated by ‘,’	imagesShared-unit-My-Company	Will backup only selected drive units belonging to the specified entity (user, group or site)
drive_shared_units_exclude	No		Valid names of existing shared drives on the selected workspace separated by ‘,’	websites	Will backup all drives except the excluded ones in the list(s), belonging to the specified entity (user, group or site)
drive_shared_units_regex_include	No		Valid regex	*.pages	Backup matching drive units (based in the drive unit name)
drive_shared_units_regex_exclude	No		Valid regex	^site.*	Exclude matching drive units (based in the drive unit name)
drive_files	No		Strings representing existing folders for the given users or shared units separated by ‘,’	Customers, Partners	Backup only specified folders belonging to the selected users
drive_files_exclude	No		Strings representing existing folders for the given users or shared units separated by ‘,’	Personal	Exclude selected folders belonging to the selected users
drive_files_regex_include	No		Valid regex	.*Company	Backup matching drive folders. Please, only provide list parameters (files + files_exclude) or regex ones. But do not try to combine them.
drive_files_regex_exclude	No		Valid regex	.*Plan	Exclude matching drive folders from the selection. Please, only provide list parameters (files + files_exclude) or regex ones. But do not try to combine them. If this is the only parameter found for selection, all elements will be included and this list will be excluded.
drive_include_trash	No		0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	Yes	Include trashed elements from the user or shared drive selected to backup
drive_include_comments	No		0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	Yes	Include comments of every file. Please, notice that performance is lower when this option is enabled as extra requests are needed for every single file to backup
drive_shared_with_me	No		0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	No	Include SharedWithMe elements of every target entity in the backup process
drive_version_history	No		0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	Yes	Include Google Drive former versions of every file into the backup process

Restore

The list below shows the subset of restore parameters that can be used to control the behavior of Google Drive module restore operations:

- `destination_user`, `drive_destination_shared_unit`, `destination_path`, `send_report`, `allow_duplicates`
- `drive_skip_sharedwitme`, `drive_skip_versions`, `drive_skip_comments`, `drive_restore_share_permissions`,
- `debug`, `foreign_container_generation`

Use cases

The following restore scenarios are supported:

- Restore files, directories, or file versions to original drive or to a different drive
 - Restore parameters implied: **`destination_user`**, **`drive_destination_shared_unit`**
- Restore file(s)/dir(s) or file version(s) to original path or to a different path
 - Restore parameters implied: **`destination_path`**
- Restore file(s)/dir(s) or file version(s) to local file system (general restore **`where`** parameter must be set to a path)
- It is possible to make general restore selections, but avoid restoring versions
 - Restore parameters implied: **`drive_skip_versions`**
- It is possible to restore sharing permissions of implied files
 - Restore parameters implied: **`drive_restore_share_permissions`**
- It is possible to make general restore selections, but specify if backed up shared elements must be considered
 - Restore parameters implied: **`drive_skip_sharedwitme`**
- It is possible to make general restore selections, but specify if backed up file comments must be considered
 - Restore parameters implied: **`drive_skip_comments`**
- It is possible to control whether or not duplicate elements are allowed (based on file id):
 - Restore parameters implied: **`allow_duplicates`**

Particularities:

- If no **`destination user`** and no **`destination shared unit`** are provided, the destination user or unit will be looked for inside the backed up path, so the destination entity will be the same as the original one
- If no **`destination_path`** is provided, the destination path will be the same as the original one
 - If a destination entity was provided, but no `destination_path` was provided and the selected file did not belong to the destination entity:
 - * A new folder will automatically be created inside the target entity
 - * For each ‘foreign’ entity, a new folder will be created
 - * Inside each ‘foreign’ entity folder, the original path structure will be preserved when restoring the files
 - *Unless the parameter **`foreign_container_generation`** is disabled

For more details about the behavior of each restore parameter, please check the general section of restore parameters.

Fileset examples

Please note that Google Drive Plugin works with two kind of entities:

- User Drives
- Shared Drives

By default, if not specifying anything on any parameters, the plugin backups everything. Therefore, in order to only backup users, we need to exclude all shared drives ; in order to backup only shared drives, we need to exclude all users. Below examples should show this more clearly.

Full Google Drive of only one user:

```
:caption: **Fileset Example**

FileSet {
  Name = fs-gw-drive-adelev
  Include {
    Options { signature = MD5 }
    Plugin = "gw: service=drive credentials_file=/opt/bacula/etc/bacula-gw-plugin-
↪credentials.json customer_id=G39add3111 admin_user_email=super@baculasystemes.com
    user=adelev@baculasystems.com drive_shared_units_regex_exclude=\".*\""
  }
}
```

Folders of one user and include sharedWithMe elements:

```
:caption: **Fileset Example**

FileSet {
  Name = fs-gw-drive-adelev-shared
  Include {
    Options { signature = MD5 }
    Plugin = "gw: service=drive credentials_file=/opt/bacula/etc/bacula-gw-plugin-
↪credentials.json customer_id=G39add3111 admin_user_email=super@baculasystemes.com
    user=adelev@baculasystems.com drive_shared_units_regex_exclude=\".*\"" drive_files=
↪"dir1,dir2\" drive_shared_with_me=yes"
  }
}
```

Backup of some specific shared drives:

```
:caption: **Fileset Example**

FileSet {
  Name = fs-gw-drive-live-2-drives
  Include {
    Options { signature = MD5 }
    Plugin = "gw: service=drive credentials_file=/opt/bacula/etc/bacula-gw-plugin-
↪credentials.json customer_id=G39add3111 admin_user_email=super@baculasystemes.com
    drive_shared_units=myunit1,myunit2 user_regex_exclude=\".*\""
  }
}
```

Exclude directories of two users:


```
:caption: **Fileset Example**
```

```
FileSet {  
  Name = fs-gw-drive-adjon-users-notemp  
  Include {  
    Options { signature = MD5 }  
    Plugin = "gw: service=drive credentials_file=/opt/bacula/etc/bacula-gw-plugin-  
↳credentials.json customer_id=G39add3111 admin_user_email=super@baculasystemes.com  
    user=\"adelev@baculasystems.com,jonis@baculasystems.com\" drive_shared_units_regex_  
↳exclude=\".*\" drive_files_exclude=temp"  
  }  
}
```

4.2 Google Mail

Bacula Enterprise Google Workspace Plugin can protect Google Mailboxes associated to users. It is possible to utilize advanced selection methods to decide exactly what is backed up (labels included/excluded, users included/excluded), as well as control precisely which messages to restore and where (original user's account or another user's account). The information protected with this service is:

- Labels
 - System labels: Inbox, Sent, Draft, Spam ...
 - User labels
- Mails
 - Metadata
 - Contents
- Attachments
- Settings
 - AutoForwarding settings
 - Imap settings
 - Language settings
 - Pop settings settings
- Delegates
- Filters
- SendAs addresses
- Forwarding addresses

Mailbox backup includes the following features:

- Incremental/Differential backup with Delta function:
 - Delta function is applied always, independently of what is the labels selection in the fileset (email_files* parameters)
- MIME object (RFC 822) export:

- It is possible to restore to a local label a selection of emails in the RFC 822 format. This way emails could be read or imported in any other tool at will

- Attachments:

- This plugin backs up the message with its attachments in a single file. This means everything will be restored when selecting a given file.
- For export purposes there is a restore option to provoke the extraction of the attachments as separate files in addition to have the full MIME object.

Messages will be formatted in the catalog in order to not include sensitive information and will be included in a path like this:

- `/@gw/customer_id/users/user@customerdomain.com/email/labelname/AAAAXXXDDDDDIIIII.msg`
 - Where the message name corresponds to the message Id provided by Google GMail API

Other objects will look as follows:

- `/@gw/customer_id/users/user@customerdomain.com/email/settings/`
 - `addr1@customerdomain.com.mailbox.sasad` -> Send ass address
 - `addr2@customerdomain.com.mailbox.fwad` -> Forwarding address
 - `addr2@customerdomain.com.mailbox.del` -> Delegates address
 - `3838383aaadffdfdf.mailbox.fil` -> Filter
 - `settings.mailbox.autfw` -> Autoforwarding settings
 - `settings.mailbox.imap` -> Imap settings
 - `settings.mailbox.lan` -> Autoforwarding settings
 - `settings.mailbox.pop` -> Autoforwarding settings
 - `settings.mailbox.vac` -> Autoforwarding settings

Backup parameters

The list below shows the specific backup parameters that can be set up in order to control the behavior of the email module.

In order to select the email module, the common *service* parameter must be equals or be containing the value *email*.

Entities that can include mailboxes are: users.

Option	Re-quired	De-efault	Values	Example	Description
email_files	No		Strings representing existing labels for the given users separated by ‘,’	Inbox, Sent	Backup only specified labels belonging to the selected users
email_files_exclude			Strings representing existing labels for the given users separated by ‘,’	Archive, Personal	Exclude selected labels belonging to the selected users
email_files_regex_include			Valid regex	.*Com-pany	Backup matching labels. Please, only provide list parameters (files + files_exclude) or regex ones. But do not try to combine them.
email_files_regex_exclude			Valid regex	.*Plan	Exclude matching labels from the selection. Please, only provide list parameters (files + files_exclude) or regex ones. But do not try to combine them. If this is the only parameter found for selection, all elements will be included and this list will be excluded.
email_settings		No	0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	Yes	Backup mailbox settings of included users
email_spam_notrash		Yes	0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	No	Backup spam and trashed messages
email_messages_exclude_string			String representing a valid Boolean Javascript expression regarding email message fields	emailSubject.includes('private') && !emailIsRead	Exclude from backup all messages that match the provided expression
email_messages_exclude_index_expr			String representing a valid Boolean Javascript expression regarding email message fields	/.*private.com/test@mail.com	Exclude only from indexing (catalog email tables) messages matching the provided expression
email_fields_exclude_index			String representing a list of email message fields	email-From, emailSubject	Do not store into the index (catalog email tables) the provided list of message fields

Restore

The list below shows the subset of restore parameters that can be used to control the behavior of email module restore operations:

- `destination_user`, `destination_path`, `send_report`, `allow_duplicates`, `debug`, `foreign_container_generation`
- `email_export`, `email_export_attachments_extract`

Use cases

The following restore scenarios are supported:

- Restore labels, emails (with their attachments) to original user or to a different user mailbox
 - Restore parameters implied: **`destination_user`**
- Restore emails to a specific label of a user:
 - Restore parameters implied: **`destination_path`**
- Export emails to local file system in export mode :
 - Restore parameters implied: **`destination_path`**, **`email_export`**
 - Extract attachments during the export: **`email_export_attachments_extract`**
- It is possible to control whether or not duplicate elements are allowed (based on file id):
 - Restore parameters implied: **`allow_duplicates`**

Particularities:

- If no **`destination_user`** is set, every message will be restored into its original mailbox
- If no **`destination_path`** is set, every message will be restored into its original path
 - If the selection contains messages from several users:
 - * Original user messages will be restored in their original location
 - * For other users, a special label will be created with the email address of each of them, containing the full path and messages of the restored objects, unless the parameter **`foreign_container_generation`** is disabled

Restore of emails from 2 different users over a third mailbox without `destination_path` result in auto-generated `Restore_date` label containing those 2 foreign users with the restored label inside of them

- Restored elements will be duplicated by default, unless **`allow_duplicates`** variable is disabled
 - Even when disabling that variable, messages will be checked by id. So if there is an element with the same information but different ID, it will not be considered to be a duplicate

For more details about the behavior of each parameter, please check the general section of restore parameters.

Messages exclude expressions

Bacula Systems is aware about one of many privacy concerns that may arise when tools like this Google Workspace Plugin enables the possibility to backup and restore data coming from different users, so the backup administrator can restore potentially private data at his will. Moreover, emails are usually one of the most critical items in terms of privacy.

One of many strategies this plugin offers in order to deal with that problem is the possibility to exclude messages. This is a very powerful feature where it is possible to use quite flexible expressions that allow to select a subset of messages and simply exclude them from the backup:

- `email_messages_exclude_expr` new fileset parameter

Or only from the index (from the catalog)

- `email_messages_exclude_index_expr` new fileset parameter

Not only messages can be excluded but also select only a subset of email fields to be included in the protected information. It is possible to exclude fields from the backup index (the catalog):

- `email_fields_exclude_index` new fileset parameter

All four discussed expressions are based on an internal structure of fields to work with. Below you can see the entire list of fields that you can use:

- `emailTags`
- `emailSubject`
- `emailFolderName`
- `emailFrom`
- `emailTo`
- `emailCc`
- `emailBodyPreview`
- `emailImportance`
- `emailTime`
- `emailIsRead`
- `emailIsDraft`

Please note that it is very important to write the fields exactly as written above.

These fields can be used in a comma separated list in the `'email_fields_exclude'` parameter and also `'email_fields_exclude_index'` parameter.

Then, for `'email_messages_exclude_expr'` and `'email_messages_exclude_index_expr'` use them in a valid boolean expression in **Javascript** language syntax. Some examples are provided below:

Listing 1: Expression to exclude messages where subject includes the word 'private'

```
emailSubject.includes('private')
```

Listing 2: Complex expression to exclude messages that are not read and are Draft or their label name is named Private

```
!emailIsRead && (emailIsDraft || emailFolderName == 'Private')
```

Listing 3: **Expression to exclude messages based on the received or sent date**

```
!emailTime < Date.parse('2012-11-01')
```

Listing 4: **Expression to exclude messages using a regex based on emailFrom**

```
/. *private.com/.test(emailFrom)
```

Note: This feature is available since Bacula Enterprise version 14.0

Expression tester

This expression mechanism can sometimes be uncertain for end users, where they can have doubts about the correct behavior of their prepared expressions. In order to help with that, Google Workspace Plugin presents a query method that allows to test those expressions against a static pre-loaded set of data.

There are two commands available:

- Show command: It will show the static data in json format, so it is possible to see the contents to adapt the expressions to test
- Test command: It will apply the expression parameters to the pre-loaded static data

The test command has the following format:

Listing 5: **Expression tester Show command**

```
.query client=<your-fd-client> plugin="gw: credentials_file=/opt/bacula/etc/gw-  
↪credentials.json customer_id=xxxxxx admin_user_email=admin@company.com"  
↪parameter=email-expr-show
```

The show command has the following format

Listing 6: **Expression tester Test command**

```
.query client=<your-fd-client> plugin="gw: credentials_file=/opt/bacula/etc/gw-  
↪credentials.json customer_id=xxxxxx admin_user_email=admin@company.com email_messages_  
↪exclude_expr = \"<your-js-expression>\" parameter=json|email-expr-test  
// Or  
.query client=<your-fd-client> plugin="gw: credentials_file=/opt/bacula/etc/gw-  
↪credentials.json customer_id=xxxxxx admin_user_email=admin@company.com email_messages_  
↪exclude_index_expr = \"<your-js-expression>\" parameter=json|email-expr-test
```

The test command produces some JSON output with objects with the same format that the plugin uses to store data into the catalog. Please note the 'total' value at the end, where the value of 12 total pre-loaded messages is shown

Listing 7: Expression tester Show command output

```
.query client=<your-fd-client> plugin="gw: credentials_file=/opt/bacula/etc/gw-
↳credentials.json customer_id=xxxxx admin_user_email=admin@company.com"
↳parameter=json|email-expr-show
....
  "email-12": {
    "body": {
      "content": "These are the contents in text format of the 12 email of test data.
↳It has the following categories:orange, black, white, purpleYou can try to filter this.
↳body using any JS method like /. *12.* /.test(emailBody) or emailBody.includes(12)",
      "contentType": "TEXT"
    },
    "ccRecipients": [
      {
        "emailAddress": {
          "address": "danny@other.com"
        }
      },
      {
        "emailAddress": {
          "address": "lucas@other.com"
        }
      },
      {
        "emailAddress": {
          "address": "terese@other.com"
        }
      }
    ],
    "from": {
      "emailAddress": {
        "address": "elon@other.com"
      }
    },
    "hasAttachments": false,
    "isDraft": false,
    "isRead": false,
    "replyTo": [
      {
        "emailAddress": {
          "address": "elon@other.com"
        }
      }
    ],
    "sentDateTime": {
      "dateTime": {
        "date": {
          "year": 2021,
          "month": 12,
          "day": 5
        },
        "time": {
```

(continues on next page)

```

        "hour": 11,
        "minute": 30,
        "second": 0,
        "nano": 0
    }
},
"offset": {
    "totalSeconds": 0
}
},
"subject": "This is private subject 12",
"toRecipients": [
    {
        "emailAddress": {
            "address": "laura@other.com"
        }
    },
    {
        "emailAddress": {
            "address": "jack@other.com"
        }
    },
    {
        "emailAddress": {
            "address": "john@other.com"
        }
    }
],
"categories": [
    "orange",
    "black",
    "white",
    "purple"
]
}
{
    "total": "12"
}

```

The test command, on its side will produce two different outputs. The first part presents the same format as the show format, and those are the messages that would be included in the backup. The second part presents a different format, so an output like:

Listing 8: Expression tester Test command, index part output

```

.query client=<your-fd-client> plugin="gw: credentials_file=/opt/bacula/etc/gw-
↳credentials.json customer_id=xxxxxx admin_user_email=admin@company.com"
↳parameter=json|email-expr-show
....
{
    "meta-email-12": {

```

(continues on next page)


```

"EmailId": "",
"EmailOwner": "test@test.com",
"EmailTenant": "johndoe.onmicrosoft.com",
"EmailTags": "orange,black,white,purple",
"EmailSubject": "This is private subject 12",
"EmailFolderName": "/",
"EmailFrom": "elon@other.com",
"EmailTo": "laura@other.com,jack@other.com,john@other.com",
"EmailCc": "danny@other.com,lucas@other.com,terese@other.com",
"EmailInternetMessageId": "",
"EmailBodyPreview": "",
"EmailImportance": "",
"EmailConversationId": "",
"EmailSize": 235,
"EmailIsRead": 0,
"EmailIsDraft": 0,
"EmailHasAttachment": 0,
"Type": "EMAIL",
"Version": 1,
"Plugin": "gw"
}
},
{
  "total-backup": "12"
},
{
  "total-index": "12"
}

```

That part represents the information that would be indexed in the backup (included into the catalog). You can also see the total entries at the end, this is very useful to quickly compare with the original 12 value and knowing if our expression is filtering the expected data or not. Below we provide an example where some filtering is applied to the backup, but also to the index:

Listing 9: Expression tester Test command, index part output

```

.query client=127.0.0.1-fd plugin="gw: credentials_file=/opt/bacula/etc/gw-credentials.
↪ json customer_id=xxxxxyyy admin_user_email=jorge@company.com email_messages_exclude_
↪ expr="\emailFrom == 'elon@other.co == 'elon@otessages_exclude_index_expr=\
↪ "emailSubject.includes('private')\" parameter=email-expr-test
...
meta-email-4={
  "EmailId": "";
  "EmailOwner": "jorge@company.com";
  "EmailTenant": "xxxxxyyy";
  "EmailTime": "2021-08-05 12:30:00";
  "EmailTags": "SENT;UNREAD;SENT;orange;black;white;purple";
  "EmailSubject": "This is orange subject 8";
  "EmailFolderName": "sent";
  "EmailFrom": "bob@company.com";
  "EmailTo": "john@company.com";
  "EmailCc": "terese@company.com";

```

(continues on next page)

```

    "EmailInternetMessageId": "1533123860.7.1655130748637@jorge-Bravo-15-Bac";
    "EmailBodyPreview": "These are the contents in text format of the 8 email of test_
↪data. It has the following categories:orange; black; white; purpleYou can try to_
↪filter this body using any JS method like /. *8.* /.test(emailBody) or emailBody.
↪includes(8)";
    "EmailImportance": "";
    "EmailConversationId": "";
    "EmailIsRead": 1;
    "EmailIsDraft": 0;
    "EmailHasAttachment": 0;
    "Type": "EMAIL";
    "Version": 1;
    "Plugin": "gw"
}
total-backup=6
total-index=4

```

In case your expression is not valid, the plugin will also inform about that with the following message:

- error=Error listing elements. Cause: Predicate test error!! Review your query

Fileset examples

Backup Full MailBox of some users, but excluding some labels:

Listing 10: Fileset Example

```

FileSet {
  Name = fs-gw-drive-adjon-users-notemp
  Include {
    Options { signature = MD5 }
    Plugin = "gw: service=email credentials_file=/opt/bacula/etc/bacula-gw-plugin-
↪credentials.json customer_id=G39add3111 admin_user_email=super@baculasystems.com
  user=\"adelev@baculasystems.com,jonis@baculasystems.com\" email_files_exclude=\"*.
↪temporary\"""
  }
}

```

Backup all MailBoxes:

Listing 11: Fileset Example

```

FileSet {
  Name = fs-gw-email-all
  Include {
    Options { signature = MD5 }
    Plugin = "gw: service=email credentials_file=/opt/bacula/etc/bacula-gw-plugin-
↪credentials.json customer_id=G39add3111 admin_user_email=super@baculasystems.com"
  }
}

```

Backup only the Inbox label of some users:

Listing 12: Fileset Example

```
FileSet {
  Name = fs-gw-email-2user-inbox
  Include {
    Options { signature = MD5 }
    Plugin = "gw: service=email credentials_file=/opt/bacula/etc/bacula-gw-plugin-
↳credentials.json customer_id=G39add3111 admin_user_email=super@baculasystems.com
    user="peter@baculasystems.com, john@baculasystems.com" email_files=inbox"
  }
}
```

Backup some users and include settings:

Listing 13: Fileset Example

```
FileSet {
  Name = fs-gw-email-2user-mime
  Include {
    Options { signature = MD5 }
    Plugin = "gw: service=email credentials_file=/opt/bacula/etc/bacula-gw-plugin-
↳credentials.json customer_id=G39add3111 admin_user_email=super@baculasystems.com
    user="peter@baculasystems.com, miriam@baculasystems.com" email_settings=yes"
  }
}
```

System labels

Google Mail can present the labels information in local languages to the user.

In general, there is no ‘multilanguage’ support, in the sense that labels must be included with their original name. For example, if you create a label named ‘books’, you cannot expect it to be backed up if you use something like ‘livres’ or ‘libros’ from other languages. You need to use the real name that was used to create such label.

There is one very important special case though, which is ‘system labels’. System labels are labels like ‘inbox’, ‘sent’ ... A full list can be found here: <https://developers.google.com/gmail/api/guides/labels>

These kind of labels can be ‘found’ by the plugin using their standard name, instead of their internal id, as it’s the general case. Therefore, for them it is possible to get the label using their English well known word even if the user sees the label with a translated word.

For example, to backup inbox it is needed to use ‘inbox’ even if for some users it is ‘Posteingang’ or ‘boîte de réception’. Google Workspace Plugin will recognize these special words and will query the information through them.

To summarize:

- System labels -> Use English word
- Other user labels -> Use original name

5 Special features

In the following section, special features and behaviors are detailed.

5.1 File revisions

Google Drive service can be configured to retain the history for files/items (this is versions or revisions of the same file).

Depending on the service and configuration, a new version can be created for each edit, each time the file is saved, manually, or never. Previous versions of a document may be retained for a finite period of time depending on admin settings which may be unique per user or location. By default, this feature is enabled and Google will store up to 100 revisions for each item and/or revisions younger than 30 days (they are deleted after 30 days).

This feature is usually enabled and the information may be accessed through the ‘Manage versions’ option available once a file has been selected.

Bacula Enterprise Google Workspace Plugin is able to backup this information if the special **drive_version_history** backup parameter is activated.

File revisions have some particularities compared to normal files:

- They are backed up as a regular file. This means a revision has its own full metadata as the parent file itself. All the metadata is the same as the file contains, except for size, dates and name.
- The name of the file is modified, so at restore time you can see the version number and the version date in the filename. Example:
 - Parent file: myDoc.doc
 - Versions:
 - * myDoc###v25.0_2021-01-19_234537.doc
 - * myDoc###v24.0_2021-01-17_212537.doc
 - * myDoc###v23.0_2021-01-12_104537.doc
 - * ...
 - * *Notice that the extension of the file is kept in order to easily identify a possible name modification in GDrive, once the file is restored
- Versions are not restored by default. You need to disable the special restore parameter ‘drive_skip_versions’, setting it to 0.

File versions are backed up in all backup levels (Full, Incremental, Differential), this means you can track all the changes of the files in your backup. For example, every Incremental run is going to backup only the new modified versions since the last Full or Incremental execution.

Here is an example of some files backed up with revisions included, listed in a restore session:

Listing 14: **versions in a job**

```
cwd is: /@gw/XXXXX/users/jorge@baculasystems.com/drive/my drive/SOURCE_REGRESS_
↪20220512104335/
$ ls
Contentiones/
Dolores###v_2022-05-12_104436729.doc
Dolores###v_2022-05-12_104444796.doc
```

(continues on next page)

```
Dolores###v_2022-05-12_104448264.doc
Dolores.doc
Dolores.doc__comments/
Legimus###v_2022-05-12_104518541.mp4
Legimus###v_2022-05-12_104527444.mp4
Legimus###v_2022-05-12_104530638.mp4
Legimus.mp4
Legimus.mp4__comments/
Netus.ppt
Posse###v_2022-05-12_104456414.docx
Posse###v_2022-05-12_104506748.docx
Posse###v_2022-05-12_104510261.docx
Posse.docx
Posse.docx__comments/
Ridiculus.jpeg
```

Note: It is important to keep in mind that versions have no Delta or Changes in special function on the API side and cannot be filtered by date. Therefore, the process to decide if a version needs to be backed up or not requires the plugin to walk through all existent versions of a modified item. In some situations this could have some undesired impact on backup performance.

5.2 Delta Backup

Some modules of the Google Workspace Plugin implement a specific mechanism to handle incremental or differential backups in order to optimize them and offer good performance. In general, the plugin will store information about the current state of the running backup. It will use that checkpoint information in the next backup, so Google APIs return only the new elements or modified elements since then.

The Google Workspace Drive API provides a 'Changes' function to track changes of some objects in an efficient way. Bacula Enterprise Google Workspace Plugin uses this function for its Delta mechanism in order to speed up Incremental/Differential processes.

For Google Email API there is a 'History' function where each email stores a sequential number which is tied to the date when it was created. Bacula Enterprise Google Workspace Plugin uses this function for its Delta mechanism in order to speed up Incremental/Differential processes.

Please note that using this function is not a mandatory requirement, therefore Incremental or Differential backups will function also with the services that may not support it by navigating through the information and comparing the date with the last valid backup of the current chain.

Delta function has some important characteristics:

- In Google Drive, the function can only be used for full entities (full user or full shared drive). This means that selecting specific paths to backup will not trigger the Delta function.
- In Google Email it is possible to select specific labels and still get the benefits of the delta function.
- Google Drive delta tokens can expire at some point, or even become invalid due to internal Google issues. If this situation happens, the plugin will try to start a new Delta cycle
- Any situation where the Delta function cannot be used will trigger a regular Full/Inc/Diff where every element is listed and selected or discarded according to the item dates.

The Delta backup cycle is described below:

- Full backup: All entity elements are backed up. The current token (token_1) is received from the API (or the most recent historyId for Email module). This token is stored locally by the FD.
- Incremental 1 backup: token_1 is used to retrieve changes since token_1's generation so every change is backed up. A new token is generated and stored locally by the FD.
- Incremental 2 backup: token_2 is used to retrieve changes since token_2's generation so every change is backed up. A new token is generated and stored locally by the FD.
- And so on...

Tokens are stored in a file placed in a path defined by the **path** parameter of the plugin. The name is: jobname.deltaLink

The file stores tokens required for every execution and it is renewed (emptied) during every Full backup execution.

This file is also backed up in the backup itself, so it can be restored manually, before an Incremental/Differential execution in case it was lost and in case you don't want to run a Full backup again.


Here we can see an example of the contents of the file, with 3 executions and one user entity involved. The structure is tree-based, so it is easy to understand what would be generated in the case of backing up other entities:

Listing 15: **deltaLink**

```
{
  "jobName": "GW_DEMO_JOB",
  "deltaServices": {
    "DRIVE": {
      "entities": {
        "JorgeShared1": {
          "id": "JorgeShared1",
          "name": "jorgeshared1",
          "containers": {
            "jorgeshared1": {
              "id": "jorgeshared1",
              "description": "jorgeshared1",
              "deltaEntries": [
                {
                  "job": "pluginTest.2022-05-12_10.54.30_15",
                  "date": "May 12, 2022, 10:54:30 AM",
                  "delta": "5677"
                },
                {
                  "job": "pluginTest.2022-05-12_11.00.59_18",
                  "date": "May 12, 2022, 11:00:59 AM",
                  "delta": "5677"
                },
                {
                  "job": "pluginTest.2022-05-12_11.03.35_21",
                  "date": "May 12, 2022, 11:03:35 AM",
                  "delta": "5683"
                }
              ]
            }
          }
        }
      }
    },
    "kara@baculasystems.com": {
      "id": "kara@baculasystems.com",
```

(continues on next page)

Destination entity: JorgeShared1					Date: 2022/05/06 16:22:57
#	File	Unit	Destination	Description	Result
1	REGRESS_20220506161411/Sagittis/Causae.txt	jorgeshared1	AAA123	File was created	SUCCESS
2	REGRESS_20220506161411/Vocibus.mp4	jorgeshared1	AAA123	File was created	SUCCESS
3	REGRESS_20220506161411/Possit.docx	jorgeshared1	AAA123	File was created	SUCCESS
4	REGRESS_20220506161411/Vis.ppt	jorgeshared1	AAA123	File was created	SUCCESS
5	REGRESS_20220506161411/Efficitur.doc	jorgeshared1	AAA123	File was created	SUCCESS
6	REGRESS_20220506161411/Amet.jpeg	jorgeshared1	AAA123	File was created	SUCCESS
7	REGRESS_20220506161411/Amet.jpeg	jorgeshared1	AAA123	File was created	SUCCESS
8	REGRESS_20220506161411/Amet.jpeg	jorgeshared1	AAA123	File was created	SUCCESS
9	REGRESS_20220506161411/Amet.jpeg	jorgeshared1	AAA123	File was created	SUCCESS
10	REGRESS_20220506161411/Amet.jpeg	jorgeshared1	AAA123	File was created	SUCCESS



Summary

Restored with Error:	0
Restored with Warning:	0
Restored Successfully:	10

Fig. 3: Restore Example Drive Report

6 Installation

The Bacula File Daemon and the Google Workspace Plugin need to be installed on the host that is going to connect to for cloud based services. The plugin is implemented over a Java layer, therefore it can be deployed on the platform better suited for your needs among any of the officially supported platforms of **Bacula Enterprise** (RHEL, SLES, Debian, Ubuntu, etc). Please, note that you may want to deploy your File Daemon and the plugin on a virtual machine directly deployed in Google Cloud Platform in order to reduce the latency between it and the Google Workspace APIs.

The system must have Java ≥ 11 installed (openjdk-11-jre for example) and the Java executable should be available in the system PATH.

6.1 Bacula Packages

We are taking Debian Buster as the example base system to proceed with the installation of the **Bacula Enterprise** Google Workspace Plugin. In this system, the installation is most easily done by adding the repository file suitable for the existing subscription and the Debian version utilized. An example would be /etc/apt/sources.list.d/bacula.list with the following content:

Listing 16: APT

```
# Bacula Enterprise
deb https://www.baculasystems.com/dl/@customer-string@/debs/bin/@version@/buster-64/
↪buster main
deb https://www.baculasystems.com/dl/@customer-string@/debs/gw/@version@/buster-64/
↪buster gw
```

After that, a run of apt update is needed:

Listing 17: APT install

```
apt update
```

Then, the plugin may be installed using:

Listing 18: APT install

```
apt install bacula-enterprise-google-workspace-plugin
```

The plugin has two different packages implied that should be installed automatically with the command shown:

- bacula-enterprise-google-workspace-plugin
- bacula-enterprise-google-workspace-plugin-libs

Alternately, manual installation of the packages may be done after downloading the packages from your Bacula Systems provided download area, and then using the package manager to install. An example:

Listing 19: APT install

```
dpkg -i bacula-enterprise-*
```

The package will install the following elements:

- Jar libraries in /opt/bacula/lib (such as bacula-google-workspace-plugin-x.x.x.jar and bacula-google-workspace-plugin-libs-x.x.x.jar). Please note that the version of those jar archives is not aligned with the version of the package. However, that version will be shown in the joblog in a message like 'Jar version:X.X.X'.
- Plugin connection file (gw-fd.so) in the plugins directory (usually /opt/bacula/plugins)
- Backend file (gw_backend) that invokes the jar files in /opt/bacula/bin. This backend file searches for the most recent bacula-google-workspace-plugin-x.x.x.jar file in order to launch it, even though usually we should have only one file.

7 Configuration

7.1 Authorization

The first step in order to use the **Bacula Enterprise** Google Workspace Plugin is to authorize it to handle data of the target workspace to backup.

The way of doing this is to:

- Define a Project in Google Cloud Platform
- Activate the proper APIs
- Generate a service account on that project with permissions
- Generate and get the credentials for that service account
- Connect the project and service credentials to the target Google Workspace to protect using domain wide permissions

Once those steps are completed, we also need to find our customer_id as well as an admin user email.

For protecting free users instead of Workspace users, the steps are very similar:

- Define a Project in Google Cloud Platform

- Activate the proper APIs
- Generate and get key credentials
- Add the target addresses to the allowed list of addresses

Google Cloud Platform Project selection

We need to login with an administrator user to the Google Cloud Platform Console (<https://console.cloud.google.com/>). Once there, we need to create a project inside our organization or select an existing one, using the combo-box located close to the Google Cloud Platform logo in the header.

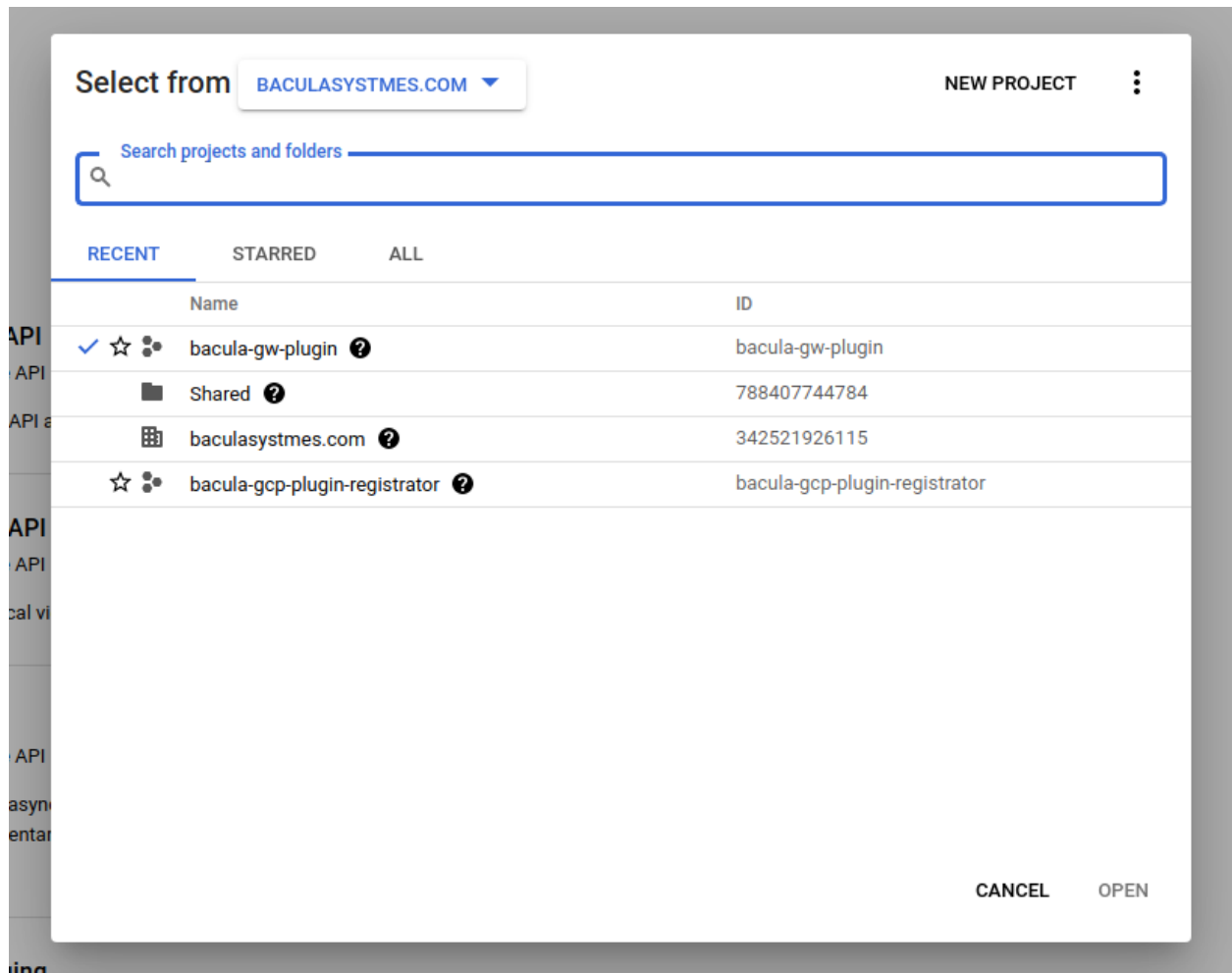


Fig. 4: Google Workspace Projects

Note: Workspace And Free users

This step is exactly the same for Google Workspace environment or free users.

Activate APIs

Once the project is selected, we need to go to API & Services > Enabled API & services.

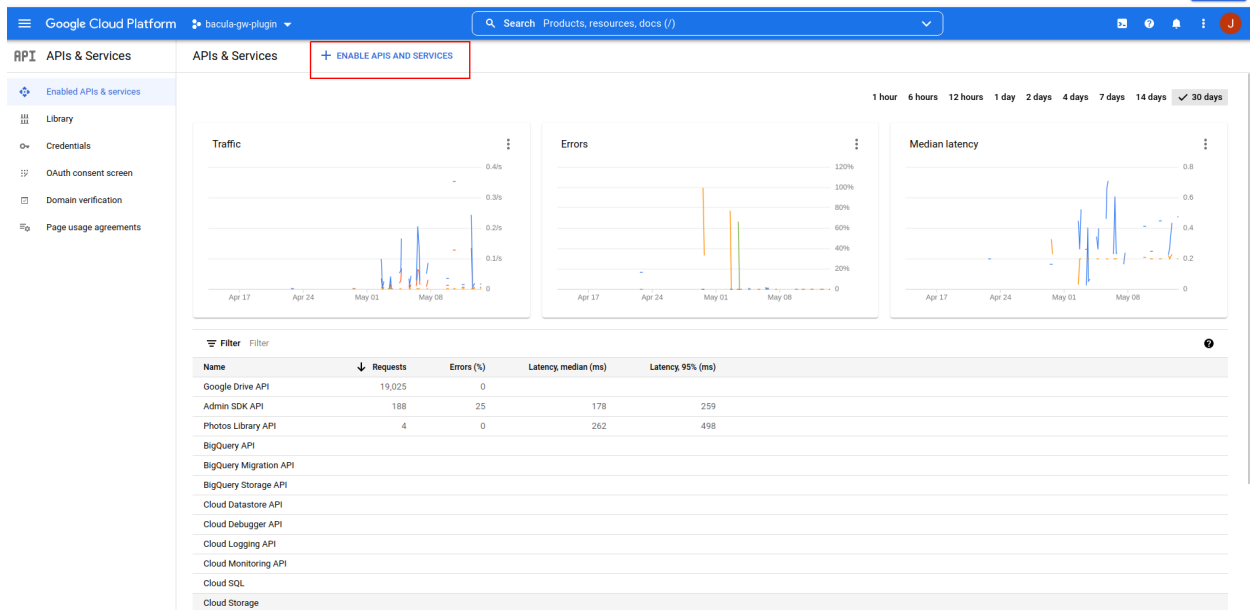


Fig. 5: Google Workspace Project APIs

From there, it is needed to click on 'Enable APIs and Services' button. Once there, we can search and activate the required APIs.

We can search for the name of each API in order to activate it. In the sample, we look for the 'Google Drive API'.

We select it and we need to enable it. Once activated, the activation button will change to show 'Manage' as the image below.

The APIs that we need to be enabled are: **Google Drive API, Gmail API, Admin SDK API and Photos Library API.**

Note: Photos

Bacula Google Workspace Plugin is not supporting right now Photos module, but it is planned to be supported on next versions and that's the reason of enabling the api.

Note: Workspace And Free users

This step is exactly the same for Google Workspace environment or free users.

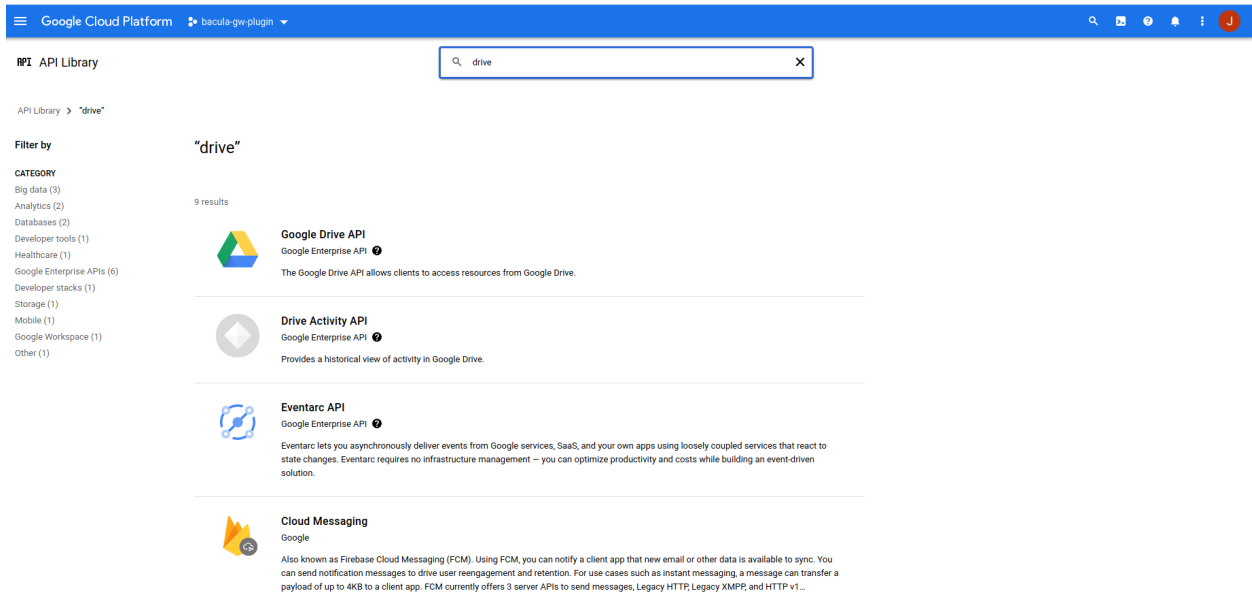


Fig. 6: Google Workspace Project Search Drive

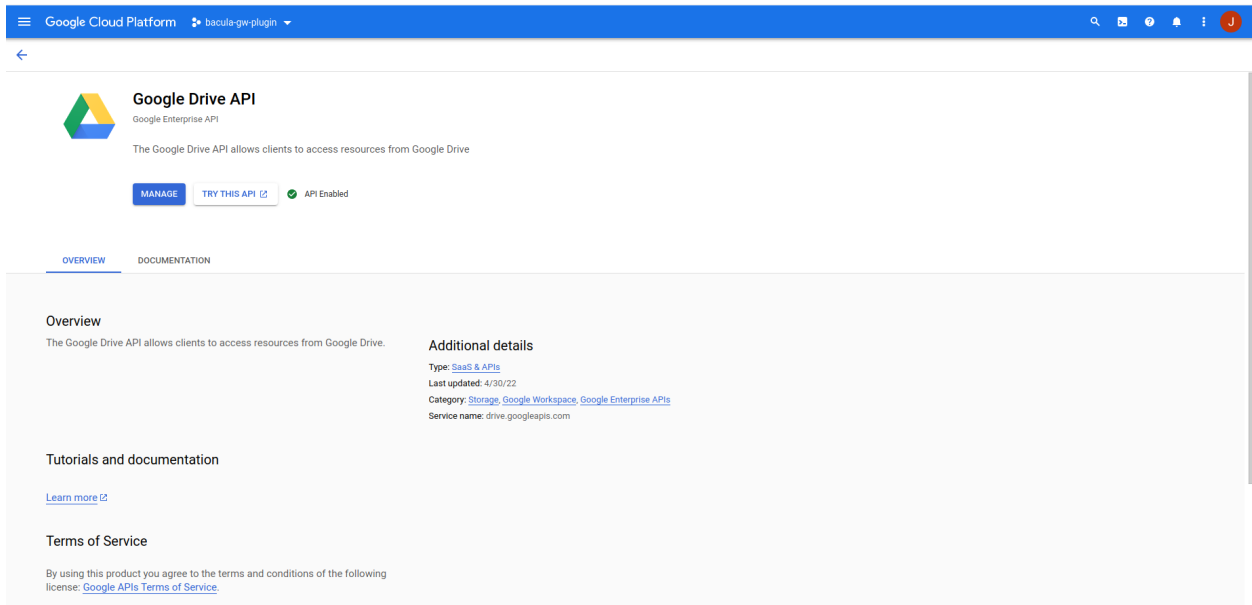


Fig. 7: Google Workspace Project Enabled APIs

Service account

Note: Workspace only

This step is only needed for Workspace environments.

From the same project, now it is needed to go to ‘IAM & Admin’ > Service Accounts.

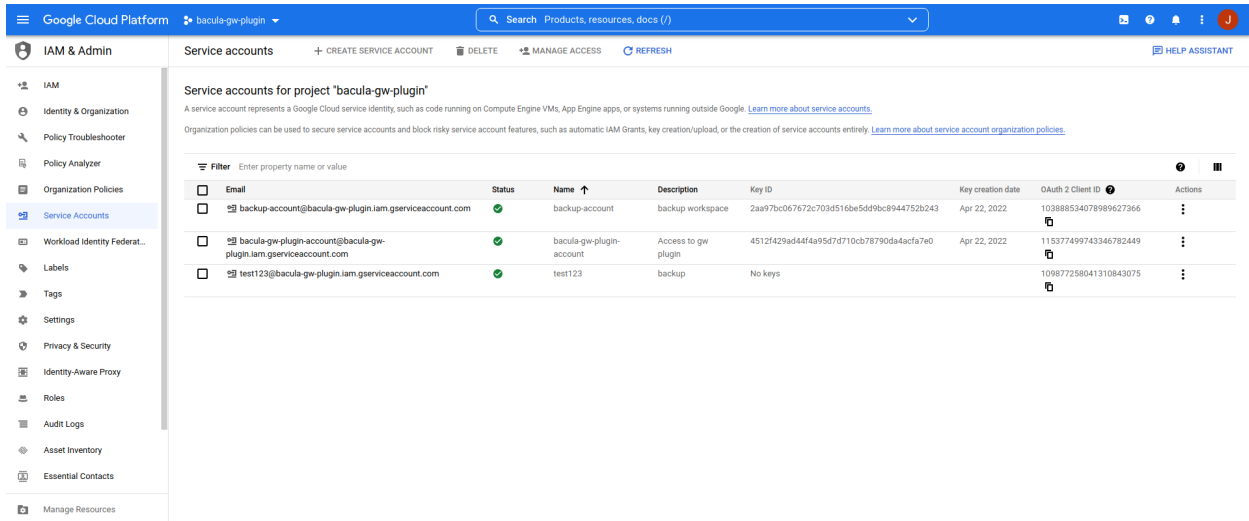


Fig. 8: Google Workspace Project Service accounts

We click on ‘Create Service Account’ and fill the form with values as the ones shown here:

In the second step it is very important to select the ‘owner’ role:

Service account key

This step is only needed for Workspace environments.

Now the service account is created and we will see it in the list. We need to select it now in order to generate a key, which will be the ‘credentials_file’ to use in any fileset of this plugin:

We need to generate a new one using the ‘ADD KEY > Create new key’ button. We select JSON format:

Once we accept, a json file containing all the information we need to use to connect to the project will be downloaded. We need to securely store that file and be referencing it in any fileset through the ‘credentials_file’ parameter.

Connect Project to Google Workspace

Note: Workspace only

This step is only needed for Workspace environments.

Without closing the Google Cloud Platform window, we need now to open a new tab and login to the Google Workspace Admin console: <https://admin.google.com/>

Once there, it is needed to open the ‘Security > Access and data control > API controls’ option:

1 Service account details

Service account name

bacula-gw-plugin-backup-sa

Display name for this service account

Service account ID *

bacula-gw-plugin-backup-sa



Email address: bacula-gw-plugin-backup-sa@bacula-gw-plugin.iam.gserviceaccount.com



Service account description

Service account to backup Workspace data

Describe what this service account will do

CREATE AND CONTINUE

2 Grant this service account access to project (optional)

3 Grant users access to this service account (optional)

DONE

CANCEL

Fig. 9: Google Workspace Project Service accounts Step 1

✓ Service account details

2 Grant this service account access to project (optional)

Grant this service account access to bacula-gw-plugin so that it has permission to complete specific actions on the resources in your project. [Learn more](#)

Role	Condition
Owner	Add condition
Full access to most Google Cloud resources. See the list of included permissions.	
+ ADD ANOTHER ROLE	
CONTINUE	

3 Grant users access to this service account (optional)

[DONE](#) [CANCEL](#)

Fig. 10: Google Workspace Project Service accounts Step 2

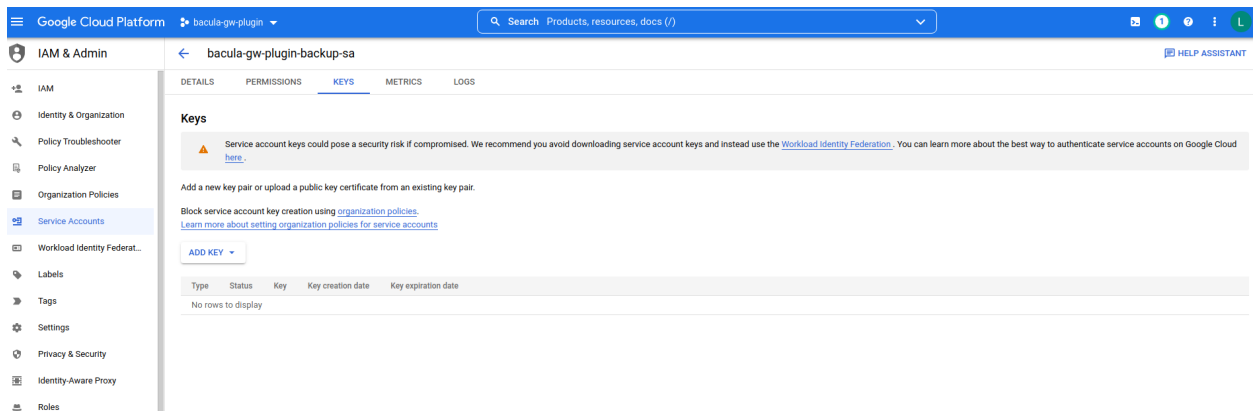


Fig. 11: Google Workspace Service account keys

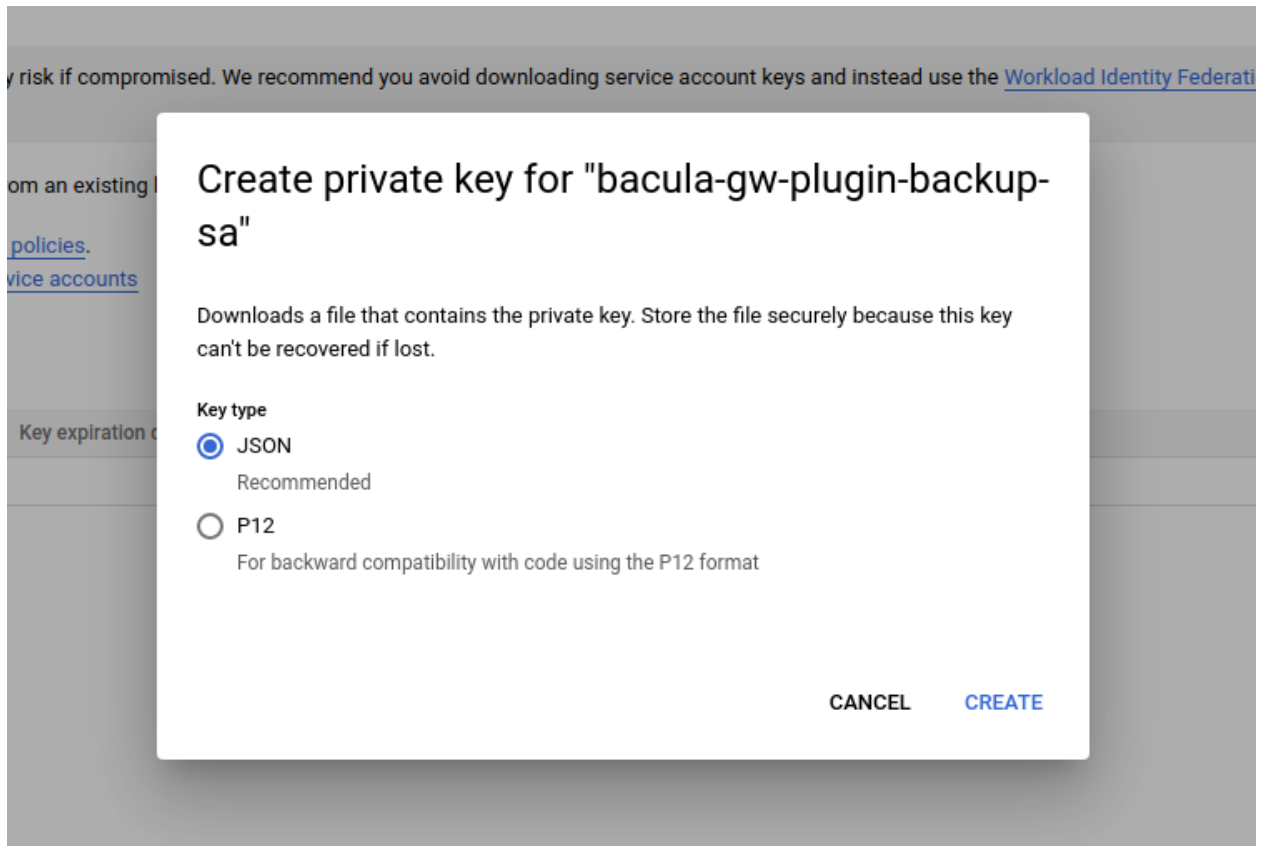


Fig. 12: Google Workspace Project Service account key creation

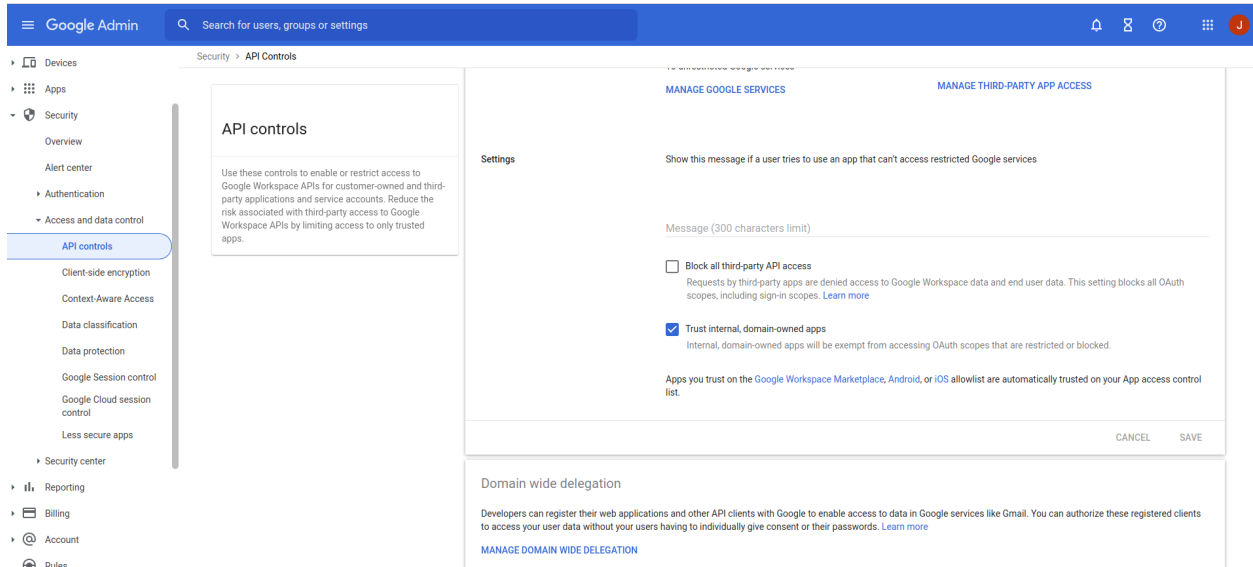


Fig. 13: Google Workspace Admin API controls

We need to make sure we have enabled the check 'Trust internal, domain-owned apps'. Then, click on 'Manage Domain Wide delegation' option located at the bottom of the screen. Then we click the 'Add new' button to see:

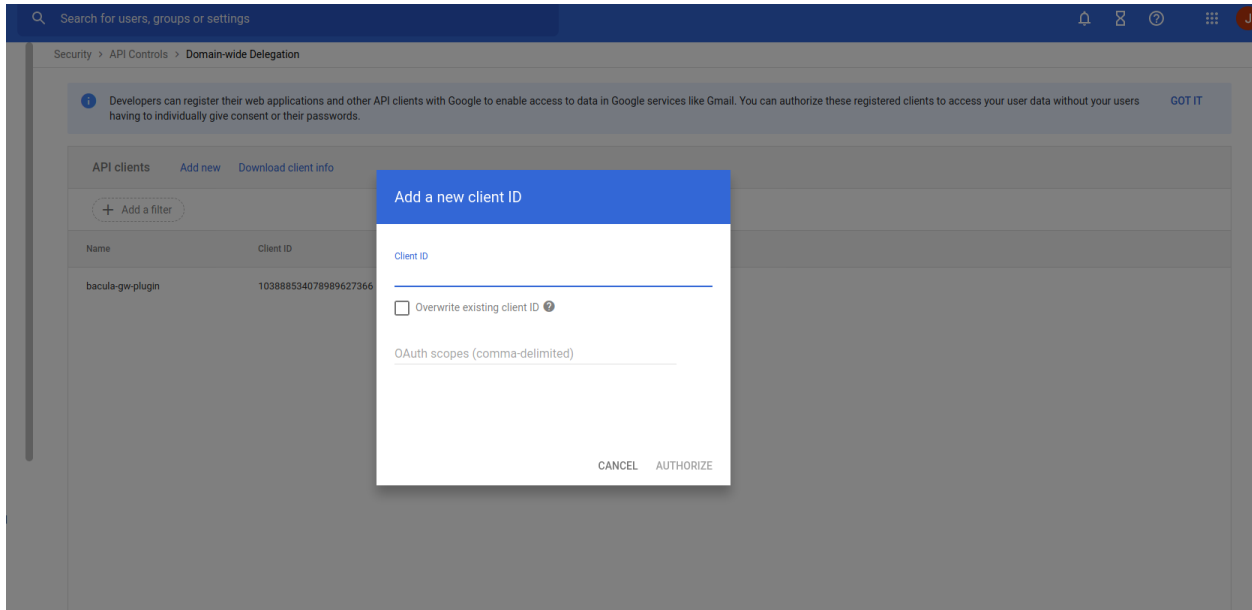


Fig. 14: Google Workspace Admin new domain wide client

The value we need there is the ID of the Service account we created in the previous step of this authentication guide. So we go back to the tab where we had services accounts and click on the 'Details' tab of our created service account. There we will find the required ID:

We copy that value into the Client ID field. For OAuth scopes, we need to put all the following ones:

- <https://www.googleapis.com/auth/drive>
- <https://www.googleapis.com/auth/admin.directory.user.security>
- <https://www.googleapis.com/auth/admin.directory.user>
- <https://www.googleapis.com/auth/drive.appdata>
- <https://www.googleapis.com/auth/photoslibrary>
- <https://mail.google.com/>
- <https://www.googleapis.com/auth/gmail.settings.basic>
- <https://www.googleapis.com/auth/gmail.settings.sharing>

Once everything is put in the form:

We click on authorize and congratulations! You should have successfully prepared your environment to use Bacula Enterprise Google Workspace Plugin.

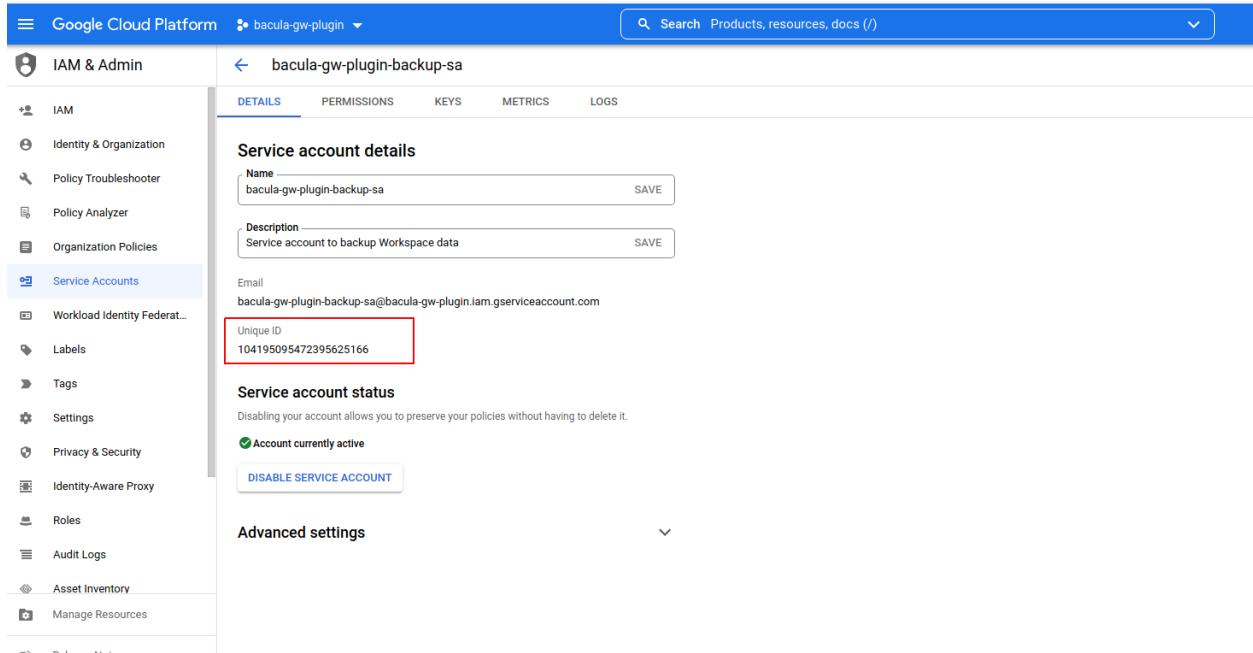


Fig. 15: Google Workspace Service account id

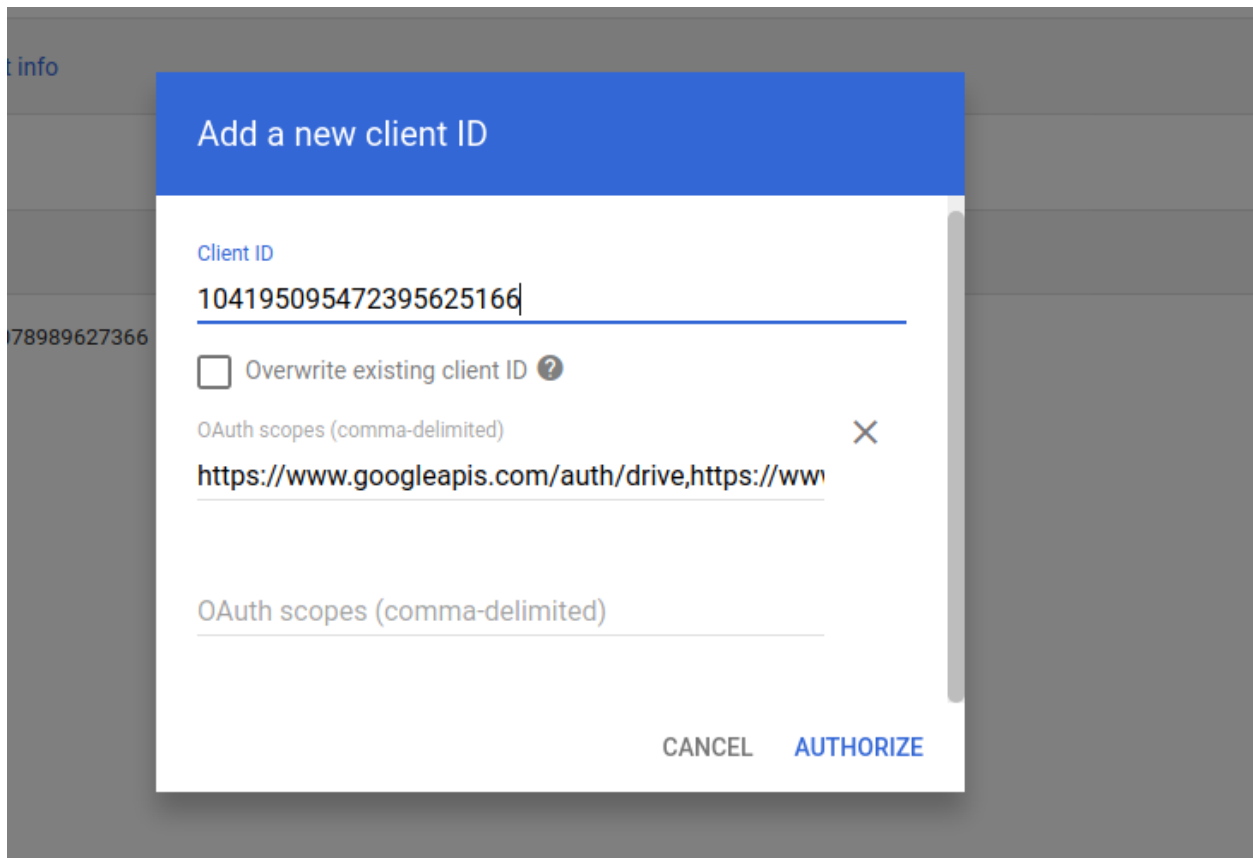


Fig. 16: Google Workspace Service account id completed

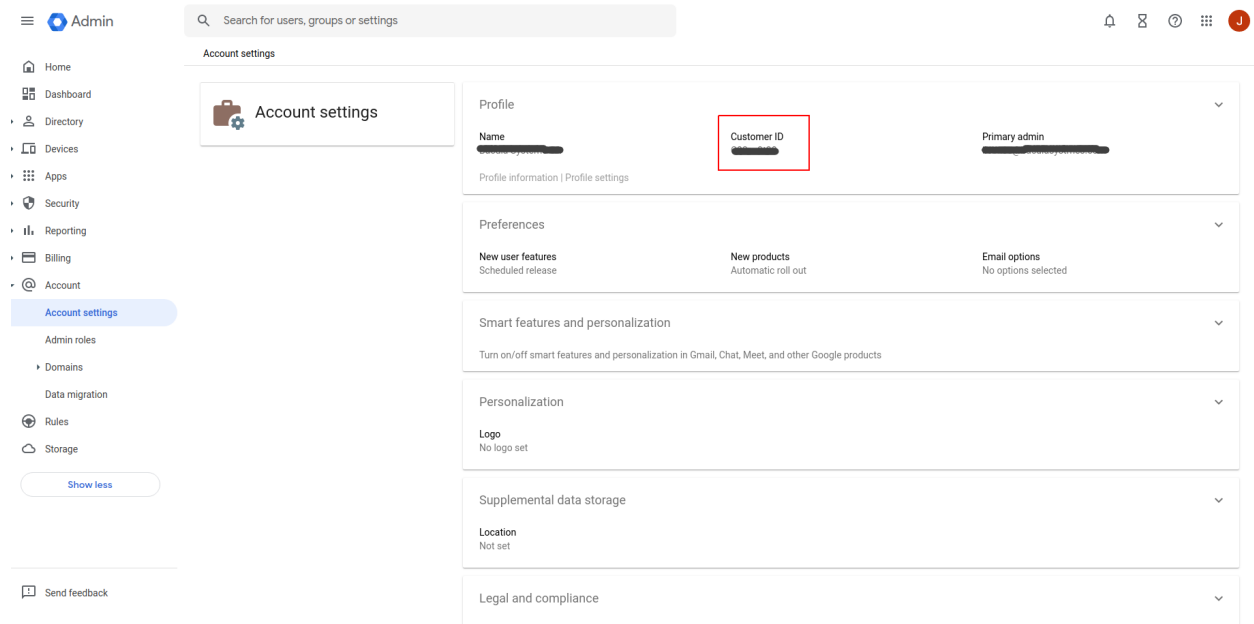
Customer id and admin user email

Note: Workspace only

This step is only needed for Workspace environments.

In order to use this plugin, in addition to a credentials file pointing to a properly configured project and workspace, it is needed to specify the customer id, as well as an admin user email.

We can find customer id in the ‘Google Admin Workspace’ console. Just go to ‘Account > Account settings’.



For the email address of an admin user, we can use the same console, but going to ‘Directory > Users’. Clicking on users, we can see the roles and privileges they have. We need to get the email of a user having the role of ‘Super Admin’ as the image below is showing:

Then we need to be use the email associated to that user that is shown in the same screen just below the user name.

Key credentials for free accounts

Note: Free users only

This step is only needed for Free Gmail users

We need to go to API & Services > credentials. From this section we will create a OAuth type credentials. We need to click in ‘Create credentials’ as the image shows:

As this is the first time, we will be redirected to configure the OAuth consent screen.

We just need to select ‘External’ and then put a name and our email as the following images show:

In the next step we need to add any user that we want to be allowed to be backed up:

Now we are ready to continue with the OAuth type credentials, so we go again with the ‘Create credentials’ button we saw in the first step, where we need to select ‘Desktop application’ and put a name:

ADMIN

Jorge Gea
 jorge@baculasystems.com
 Active
 Last sign in: 41 minutes ago
 Created: Apr 21, 2022

Organizational unit
Bacula Systems SA

RESET PASSWORD

UPDATE USER

UPLOAD PROFILE PHOTO

ADD ALTERNATE EMAILS

ADD TO GROUPS

EMAIL

SUSPEND USER

RESTORE DATA

DELETE USER

CHANGE ORGANIZATIONAL UNIT

Admin roles and privileges

Roles
 Super Admin

Privileges
 159 Admin console privileges, 34 Admin API privileges

Apps

Google apps
 76 of 78 available Google services are on for Jorge. [Turn apps on or off](#)

Other cloud apps
 No other cloud apps have been added to this organization. [Browse for apps](#) [Turn apps on or off](#)

Managed devices

Your organization doesn't have mobile device management.
[SETUP MOBILE DEVICE MANAGEMENT](#)

Licenses

1 license assigned
 Google Workspace Enterprise Plus

Total estimated monthly bill
 -

Shared drives

Member of 1 shared drive

Fig. 17: Google Workspace Super Admin User

Google Cloud Platform my-test-project

Search Products, resources, docs (1)

APIs & Services

Credentials [+ CREATE CREDENTIALS](#) [DELETE](#)

Create credentials to access your enabled APIs. [Learn more](#)

Remember to configure the OAuth consent screen with information about your application. [CONFIGURE CONSENT SCREEN](#)

API Keys

Name	Creation date	Restrictions	Actions
No API keys to display			

OAuth 2.0 Client IDs

Name	Creation date	Type	Client ID	Actions
No OAuth clients to display				

Service Accounts [Manage service accounts](#)

Email	Name	Actions
No service accounts to display		

API	APIs & Services	OAuth consent screen
	Enabled APIs & services	<p>Choose how you want to configure and register your app, including your target users. You can only associate one app with your project.</p> <p>User Type</p> <p><input type="radio"/> Internal </p> <p>Only available to users within your organization. You will not need to submit your app for verification. Learn more about user type</p> <p><input checked="" type="radio"/> External </p> <p>Available to any test user with a Google Account. Your app will start in testing mode and will only be available to users you add to the list of test users. Once your app is ready to push to production, you may need to verify your app. Learn more about user type</p> <p>CREATE</p> <p>Let us know what you think about our OAuth experience</p>
	Library	
	Credentials	
	OAuth consent screen	
	Domain verification	
	Page usage agreements	

Edit app registration

1 **OAuth consent screen** — 2 Scopes — 3 Test users — 4 Summary

App information

This shows in the consent screen, and helps end users know who you are and contact you

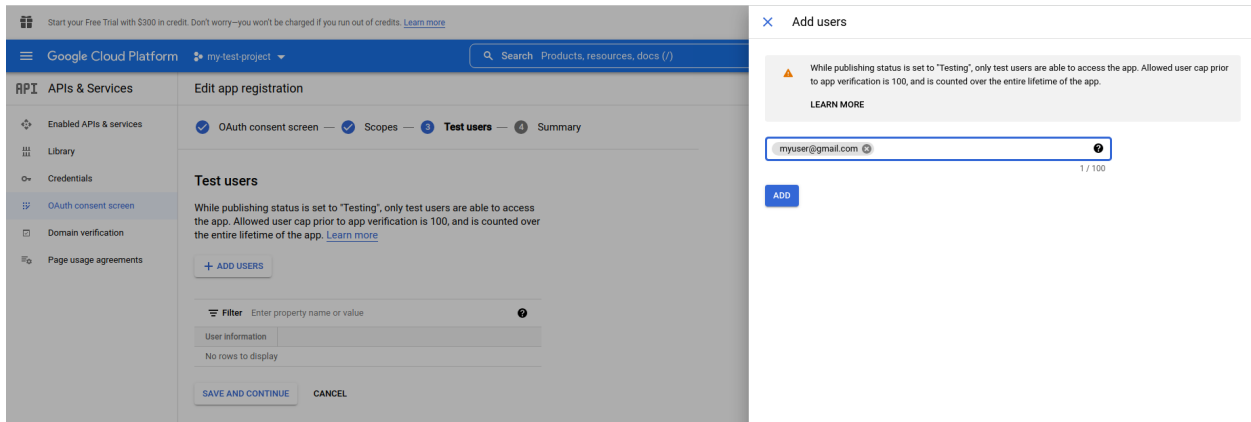
App name *

 The name of the app asking for consent

User support email *

 For users to contact you with questions about their consent

App logo [BROWSE](#)
 Upload an image, not larger than 1MB on the consent screen that will help users recognize your app. Allowed image formats are JPG, PNG, and BMP. Logos should be square and 120px by 120px for the best results.



← Create OAuth client ID

A client ID is used to identify a single app to Google's OAuth servers. If your app runs on multiple platforms, each will need its own client ID. See [Setting up OAuth 2.0](#) for more information. [Learn more](#) about OAuth client types.

Application type *
Desktop app

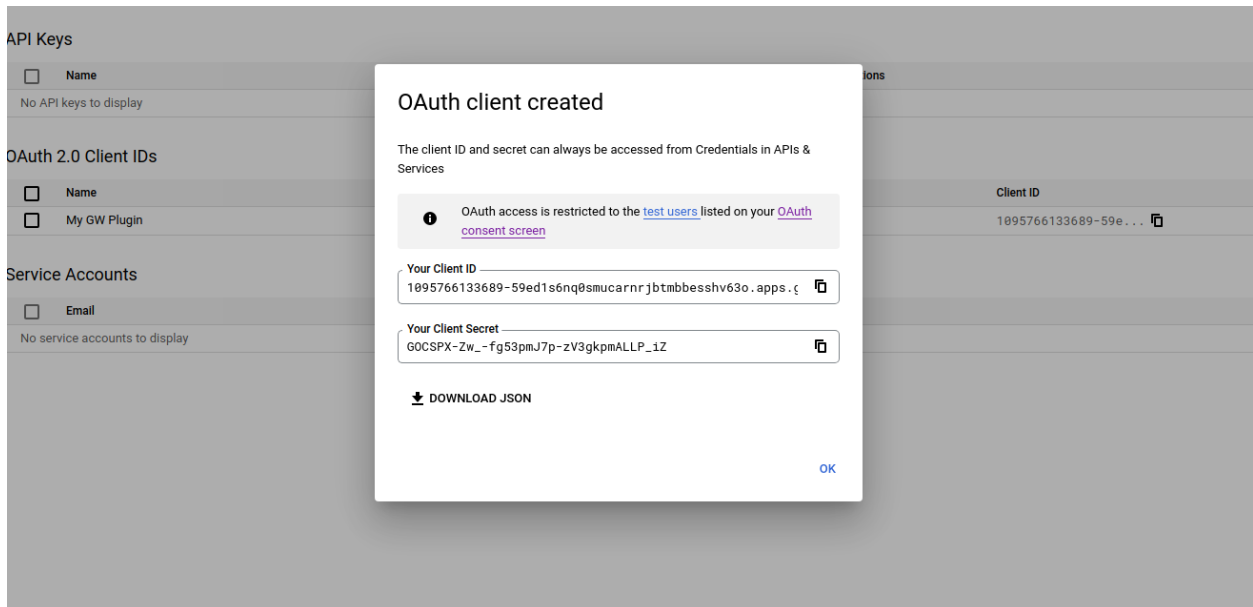
Name *
My GW Plugin

The name of your OAuth 2.0 client. This name is only used to identify the client in the console and will not be shown to end users.

Note: It may take 5 minutes to a few hours for settings to take effect

CREATE CANCEL

The OAuth Id will be generated and we need to click on the download json button the image shows:



That JSON downloaded file is the file we need to refer to in the 'credentials_file' fileset parameter in order for the plugin to authenticate our user. The first time we use the plugin with it it will ask to open a URL so we can confirm the permissions the plugin needs to perform the backup in our account, so it is needed to open that URL, to login with the proper user, select all the permissions shown and accept them.

Log example:

```
time: 2022-06-13 17:44:14 logtext: 127.0.0.1-fd JobId 4: gw: Please, open the following address in your browser and login with 'jorgebacula@gmail.com'
```

```
time: 2022-06-13 17:44:14 logtext: 127.0.0.1-fd JobId 4: gw: https://accounts.google.com/o/oauth2/auth?access_type=offline&client_id=105945873553-e4bu9nipqljhbh1edgjtgueleqe4do7r.apps.googleusercontent.com&redirect_uri=http://localhost:8888/Callback&response_type=code&scope=https://www.googleapis.com/auth/drive%20https://www.googleapis.com/auth/drive.appdata%20https://www.googleapis.com/auth/photoslibrary%20https://mail.google.com/%20https://www.googleapis.com/auth/gmail.settings.basic
```


Once we open the URL and login with the proper user, we need to select all the items as the image shows and click on 'Continue':

The job will automatically continue its execution after that.


It is important to note that the service expecting to receive the result of this interaction will automatically listen on port 8888, this port can be adjusted using the plugin parameter 'auth_port' if needed.


The credentials will be stored by default into the 'tokens' path inside the 'path' directory. This can be changed using the 'tokens_path' variable. Those persistent credentials will avoid performing the authentication for every job execution.

bacula-google-workspace-plugin wants access to your Google Account


 jorgebacula@gmail.com


Select what **bacula-google-workspace-plugin** can access

 Read, compose, send, and permanently delete all your email from Gmail. [Learn more](#)

 See, edit, create, or change your email settings and filters in Gmail. [Learn more](#)

 See, edit, create, and delete all of your Google Drive files. [Learn more](#)

 See, create, and delete its own configuration data in your Google Drive. [Learn more](#)

 See, upload, and organize items in your Google Photos library. [Learn more](#)

Make sure you trust bacula-google-workspace-plugin

You may be sharing sensitive info with this site or app. You can always see or remove access in your [Google Account](#).

Learn how Google helps you [share data safely](#).

See bacula-google-workspace-plugin's [Privacy Policy](#) and [Terms of Service](#).

Cancel

Continue

7.2 Fileset Configuration

Once the plugin is successfully authorized, it is possible to define regular filesets for backup jobs in Bacula, where we need to include a line similar to the one below, in order to call the Google Workspace Plugin:

Listing 20: Fileset GW

```
FileSet {
  Name = FS_GW
  Include {
    Options {
      signature = MD5
      ...
    }
    Plugin = "gw: <gw-parameter-1>=<gw-value-1> <gw-parameter-2>=<gw-value-2> ..."
  }
}
```

It is **strongly recommended** to use only one ‘Plugin’ line in every fileset. The plugin offers the needed flexibility to combine different modules or entities to backup inside the same plugin line. Different workspaces, in case of existing, should be using different filesets and different jobs.

Below sub-sections list all the parameters you can use to control GW Plugin behavior.

In this plugin, any parameter allowing a list of values can be assigned with a list of values separated by ‘,’.

Common parameters

These parameters are common and applicable to all the modules of the Google Workspace Plugin.

Option	Required	Default	Values	Example	Description
abort_on_error	No	No	No, Yes	Yes	If set to Yes : Abort job as soon as any error is found with any element. If set to No : Jobs can continue even if they found a problem with some elements. They will try to backup or restore the other and only show a warning
config_file	No		The path pointing to a file containing any combination of plugin parameters	/opt/bacula/etc/pluginsettings	Allows to define a config file where configure any parameter of the plugin. Therefore you don't need to put them directly in the Plugin line of the fileset. This is specially useful for shared data between filesets and/or sensitive data as customer_id.
log	No	/opt/bacula/working/debug.log	A working path with enough permissions for File Daemon to create a file with the provided name	/tmp/gw.log	Generates additional log in addition to what is shown in job log. This parameter is included in the backend file, so, in general, by default the log is going to be stored in the working directory.
debug	No	0	0, 1, 2, 3, 4, 5, 6, 7, 8, 9	Debug level. Greater values generate more debug information	Generates the working/gw/gw-debug.log* files containing debug information which is more verbose with a greater debug number
path	No	/opt/bacula/working	A working path with enough permissions for File Daemon to create any internal plugin file	/mnt/my-vol/	Uses this path to store metadata, plugin internal information and temporary files
customer_id	No		String representing the customer id associated to the Google Workspace subscription	Cbdi2930doi	The customer id associated to the Google Workspace subscription to be backed up. Please, check the authentication section of this document for more detailed information. Note that this is mandatory if you want to protect a workspace environment, but not needed to protect gmail free accounts.
admin_user_email	No		A valid email address of one admin user of the Google Workspace subscription	rafael@custo	The email address of an admin user of the Google Workspace subscription to be protected. Please, check the authentication section of this document for more detailed information. Note that this is mandatory if you want to protect a workspace environment, but not needed to protect gmail free accounts.
credentials_file	Yes		The path of the file where credentials are stored	/opt/bacula/etc/credentials	The credentials file is downloaded from the configured Google Cloud application that will act as a bridge in order to allow the communication between this plugin and Google Workspace. Please, check the authentication section of this document for more detailed information.
tokens_path	No*	tokens	A path with enough permissions so File Daemon can write in it	/home/user/m	The path to tokens will be used to store the login cache for the device code flow authenticated users, which is relative to the path folder folder (usually working/gw/customer_id/tokens_path/). This is not used and not needed for protecting a workspace with a subscription
auth_port	No*	8888	An integer with an OpenPortNumber suitable of receiving the answer from Google Cloud services upon	9999	The port to be used to open the internal service to receive the authentication answer from Google Cloud services

The plugin supports two different kind of users: Workspace users and free Gmail users.

For Workspace users, in addition to ‘credentials_file’, the following parameters are mandatory: ‘customer_id’ and ‘admin_user_email’.

For free Gmail users those parameters are not used, but it is possible to customize ‘tokens_path’ and ‘auth_port’.

Advanced common parameters

Following parameters are common to all Google Workspace modules (and even with some other plugins), but are advanced ones. They should not be modified in most common use cases.

Option	Re-quired	Default	Values	Ex-ample	Description
stream_sleep	No	1	Positive integer (1/10 seconds)	5	Time to sleep when reading header packets from FD and not having a full header available
stream_max_wait	No	120	Positive integer (seconds)	360	Max wait time for FD to answer packet requests
time_max_last_modify	Modify	18400	Positive integer (seconds)	43200	Maximum time to wait to overwrite a debug log that was marked as being used by other process
log-ging_max_file_size	No	50MB	String size	300MB	Maximum size of a single debug log file
log-ging_max_backup_index	No	25	Positive integer (number of files)	50	Maximum number of log files to keep
log_rolling_file_pattern	Pattern	gw.log.%d{dd-MMM}.log.gz”	No, Yes	Yes	Log patter for rotated log files
split_config_file	No	=	Character	:	Character to be used in config_file parameter as separator for keys and values
opener_queue_timeout_secs	No	3600	Positive integer (seconds)	3600	Timeout when internal object opener queue is full
pub-lisher_queue_timeout_secs	No	1200	Positive integer (seconds)	3600	Timeout when internal object publisher queue is full

The internal plugin logging framework presents some relevant features that we are going to describe:

- The “.log” files are rotated automatically. Currently each file can be 50Mb at maximum and the plugin will keep 25 files.
 - This behavior can be changed using the internal advanced parameters: logging_max_file_size and logging_max_backup_index
- The “.err” file can show contents even if no real error happened in the jobs. It can show contents too even if debug is disabled. This file is not rotated, but it is expected to be a small file in general. If you still need to rotate it, you can include it in a general rotating tool like ‘logrotate’.
- Backups in parallel and also failed backups will generate several log files. For example: gw-debug-0.log, gw-debug-1.log...

Tuning parameters

These set of parameters are common to all modules and they are advanced ones. They should not be modified in general. They can be used to tune the behavior of the plugin to be more flexible in particular bad network environments or when significant job concurrency is happening, etc.

Option	Required	Default	Values	Example	Description
backup_queue_size	No	10	0-50	1	Number of maximum en-queued internal operations between service static internal threads (there are 3 communicating through queues with the set size: service fetcher, service opener and general publisher to bacula core). This could potentially affect google api concurrent requests and consequently, Google throttling. It is only needed to modify this parameter, in general, if you are going to run different jobs in parallel
concurrent_threads	No	5	0-10	1	Number of maximum concurrent backup threads running in parallel in order to fetch or open data for running download actions. This means every service fetcher and service opener will open this number of child concurrent threads. This will affect google api concurrent requests. Google API can throttle requests depending on a variety of circumstances, but it is directly attached . It is only needed to modify this parameter, in general, if you are going to run different jobs in parallel. If you want to have a precise control of your concurrency through different jobs, please set up this value to 1. Please be careful also with the memory requirements, multi-threaded increases very significantly memory consumption per job
api_list_page_size	No	50	1-500	350	Number of maximum elements got from Google API for each page of objects. Higher number implies less requests, but more memory and more time for each request
api_timeout	No	9000	Positive integer (milliseconds)	60000	Google call timeout inside HttpClient
api_read_timeout	No	3000	Positive integer (milliseconds)	30000	Google read timeout inside HttpClient
api_retries	No	5	Positive integer (number of retries)	10	Google number of retries for retry-candidate requests
api_retry_delay	No	10	Positive integer (seconds)	10	Google API delay between retries
general_network_retries	No	5	Positive integer (number of retries)	10	Number of retries for the general external retry mechanism
gen-	No	50	Pos-	100	General Plugin delay between retries

Entity parameters

The following list of parameters are commonly shared through any module used in the same fileset line and are intended to select the target entities to backup. Every module subsection mentions what entities are supported too.

Option	Required	Default	Values	Example	Services	Description
user	No		Valid email addresses of existing users on the selected workspace separated by ‘;’	AlexW@yourdomain.com LeeY@yourdomain.com	drive	Backup selected services of this list of users. If no user is provided, and no other user parameter is set, all users will be discovered and included in the backup
user_exclude	No		Valid email addresses of existing users on the selected workspace separated by ‘;’	LauraG@yourdomain.com AmandaT@yourdomain.com	drive email	Exclude selected services of selected users. If this is the only parameter found for selection, all elements will be included and this list will be excluded
user_regex_include	No		Valid regex	.*@management\.mydomain.com	drive email	Backup selected services of matching users.
user_regex_exclude	No		Valid regex	.*@guests\.mydomain.com	drive email	Exclude selected services of matching users. If this is the only parameter found for selection, all elements will be included and this list will be excluded

Backup parameters

Please, check the specific module pages in order to see backup parameters that are applicable only to each of them:

- [Google Drive](#)
- [Google Email](#)

Restore parameters

The plugin is able to restore to the local file system on the server where the File Daemon is running or to the Google Workspace environment. The method is selected based on the value of the *where* parameter at restore time:

- Empty or ‘/’ (example: where=/) → Google Workspace restore will be triggered
- Any other path for where (example: where=/tmp) → Local file system restore will be triggered

When using Google Workspace restore option, the following parameters may be modified by selecting ‘Plugin Options’ during the bconsole restore session:

Option	Required	Default	Values	Example	Services	Description
destination_user	No		Existing email address on the target Google Workspace	AlexW	Drive, email	Destination User where restore data will be uploaded. If no user is set, every selected file will be restored in the original account
destination_path	No		Destination path to be created (or existing) into the selected user (drive folder path)	Restore-Folder	drive, email	Destination folder where all selected files to restore will be restored. If no path is set: - If no user is set either, every element will go to its original location - If a user is set using the variable destination_user : - Elements belonging to destination_user will be restored in their original location - Elements belonging to different users than destination_user will be restored in a new folder using the email address of the original user of the element
send_report	0	0	0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	1	drive, email	Send a report to the user where every restore action is listed. - In drive service this will generate a new text file in the top restore folder
allow_duplicates	No	1	0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	0	drive, email	Set if we allow to have several files with the same name in the same path or not (if not, we can overwrite the file using the 'Replace' general restore variable)
drive_destination_shared_unit			Existing shared drive name	MyShared Drive	drive	Destination drive shared unit where restored data will be uploaded. If no drive is set, every selected file will be restored in the original shared drive
drive_skip_versions			0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	0	drive	Skip restoring former file versions (tagged with '###date') even if they are selected. Important : Notice that this parameter is enabled by default , as we consider not restoring file versions the most common case. You need to disable it in order to have this kind of files restored
drive_skip_comments			0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	0	drive	Skip restoring file comments (located inside the 'filename_comments' folder) even if they are selected. Important : Notice that this parameter is enabled by default , as we consider not restoring file comments the most common case. You need to disable it in order to have this kind of information restored
drive_skip_shared_with_me			0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	1	drive	Skip restoring shared with me elements even if they are selected.
drive_restore_share_permissions			0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	1	drive	Restore share permissions of every element in order to regenerate sharing information as allowed identities, shared links, etc. Important : Notice that this parameter is disabled by default , as we consider not restoring sharing permissions the most common case. You need to enable it in order to have shared permissions restored
email_export		0	0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	1	email	Export selected emails to MIME format in local filesystem (RFC 822)
email_export_attachments_extract			0, no, No, false, FALSE, false, off ; 1, yes, Yes, TRUE, true, on	0	email	Extract attachments of exported emails as independent files
customer_id	No		String representing the customer id associated to the Google	Customer ID	email	The customer id associated to the Google Workspace subscription to be backed up. Please, check the authentication section of this document for more detailed information. Note that this is mandatory if you want to protect

8 Operations

8.1 Backup

Google Workspace plugin backup configurations currently have just one specific requirement in the Job resource. Below we show some examples.

Job Example

The only special requirement with Google Workspace jobs is that Accurate mode backups must be disabled, as this feature is not supported at this time.

Listing 21: **Job Example**

```
Job {
  Name = gw-myworkspace-backup
  FileSet = fs-gw-drive-all
  Accurate = no
  ...
}
```

FileSet Examples

The plugin supports enough flexibility to configure almost any type of desired backup. Multiple *Plugin=* lines should not be specified in the Include section of a FileSet for the Google Workspace Plugin.

Fileset examples for every supported service are linked below. For common purposes, the following two examples show how to configure an external config file or configure the number of threads:

Setup external config file:

Listing 22: **Fileset Example**

```
FileSet {
  Name = FS_GW_DRIVE
  Include {
    Options {
      signature = MD5
    }
    Plugin = "gw: config_file=/opt/bacula/etc/gw.settings service=drive"
  }
}
```

Listing 23: **Settings file**

```
$ cat /opt/bacula/etc/gw.settings
```

Increase number of threads:

Listing 24: Fileset Example

```
FileSet {
  Name = fs-gw-drive-kara
  Include {
    Options {
      signature = MD5
    }
    Plugin = "gw: credentials_file=/opt/bacula/etc/bacula-gw-plugin-sa-credentials.json
customer_id=\"B01ua5i29\" admin_user_email=\"peter@baculasystems.com\" service=drive
user=kara@baculasystems.com backup_threads=10"
  }
}
```

More fileset examples for:

- [Google Drive](#)
- [Google Mail](#)

8.2 Restore

Restore operations are done using standard **Bacula Enterprise** bconsole commands.

The *where* parameter controls if the restore will be done locally to the File Daemon's file system or to the Google Workspace service:

- *where=/* or empty value → Restore will be done over Google Workspace
- *where=/any/other/path* → Restore will be done locally to the File Daemon file system

Restore options are described in the *restore-params* section of this document, so here we are going to simply show an example restore session:

Listing 25: Restore Drive Bconsole Session

```
*restore where=/

First you select one or more JobIds that contain files
to be restored. You will be presented several methods
of specifying the JobIds. Then you will be allowed to
select which files from those JobIds are to be restored.

To select the JobIds, you have the following choices:
  1: List last 20 Jobs run
  2: List Jobs where a given File is saved
  3: Enter list of comma separated JobIds to select
  4: Enter SQL list command
  5: Select the most recent backup for a client
  6: Select backup for a client before a specified time
  7: Enter a list of files to restore
  8: Enter a list of files to restore before a specified time
  9: Find the JobIds of the most recent backup for a client
 10: Find the JobIds for a backup for a client before a specified time
 11: Enter a list of directories to restore for found JobIds
 12: Select full restore to a specified Job date
```

(continues on next page)

```

13: Select object to restore
14: Cancel
Select item: (1-14): 5
Automatically selected Client: 127.0.0.1-fd
Automatically selected FileSet: FS_GW
+-----+-----+-----+-----+-----+-----+-----+-----+
| jobid | level | jobfiles | jobbytes | starttime          | volumename          |
+-----+-----+-----+-----+-----+-----+-----+-----+
|    1  | F    |    29   | 125,994 | 2022-05-12 17:49:27 | TEST-2022-05-12:0  |
+-----+-----+-----+-----+-----+-----+-----+-----+
You have selected the following JobId: 1

Building directory tree for JobId(s) 1 ...
27 files inserted into the tree.

You are now entering file selection mode where you add (mark) and
remove (unmark) files to be restored. No files are initially added, unless
you used the "all" keyword on the command line.
Enter "done" to leave this mode.

cwd is: /
$ cd "/@gw/C02uv9t30/users/jorge@baculasystemes.com/drive/my drive/"
cwd is: /@gw/C02uv9t30/users/jorge@baculasystemes.com/drive/my drive/
$ ls
REGRESS_20220512174729/
sharedWithMe/
$ cd REGRESS_20220512174729/
cwd is: /@gw/C02uv9t30/users/jorge@baculasystemes.com/drive/my drive/REGRESS_
↪20220512174729/
$ ls
Elitr.mp4
Elitr.mp4__comments/
Graeco.docx
Graeco.docx__comments/
Interpretaris/
Mnesarchum.ppt
Scelerisque.jpeg
Vivamus.doc
Vivamus.doc__comments/
$ mark *
20 files marked.
$ done
Bootstrap records written to /tmp/regress/working/127.0.0.1-dir.restore.2.bsr

The Job will require the following (*=>InChanger):
  Volume(s)                Storage(s)                SD Device(s)
=====
  TEST-2022-05-12:0        File                      FileStorage

Volumes marked with "*" are in the Autochanger.

```

20 files selected to be restored.

Using Catalog "MyCatalog"

Run Restore job

```

JobName:      RestoreFiles
Bootstrap:    /tmp/regress/working/127.0.0.1-dir.restore.2.bsr
Where:        /
Replace:      Always
FileSet:      Full Set
Backup Client: 127.0.0.1-fd
Restore Client: 127.0.0.1-fd
Storage:      File
When:         2022-05-12 18:03:23
Catalog:      MyCatalog
Priority:      10
Plugin Options: *None*

```

OK to run? (Yes/mod/no): mod

Parameters to modify:

- 1: Level
- 2: Storage
- 3: Job
- 4: FileSet
- 5: Restore Client
- 6: When
- 7: Priority
- 8: Bootstrap
- 9: Where
- 10: File Relocation
- 11: Replace
- 12: JobId
- 13: Plugin Options

Select parameter to modify (1-13): 13

```

Automatically selected : gw: credentials_file="/home/jorge/projects/bacula-gw-plugin-sa-
↳2.json" customer_id="C02uv9t30" admin_user_email="jorge@baculasystemes.com" service=
↳"drive" user="jorge@baculasystemes.com" drive_files="REGRESS_20220512174729" drive_
↳shared_units_regex_exclude=".*" debug=6

```

Plugin Restore Options

Option	Current Value	Default Value
destination_user:	*None*	(*None*)
destination_path:	*None*	(*None*)
send_report:	*None*	(0)
allow_duplicates:	*None*	(1)
drive_destination_shared_unit:	*None*	(*None*)
drive_skip_versions:	*None*	(1)
drive_skip_comments:	*None*	(1)
drive_skip_sharedwithme:	*None*	(0)
drive_restore_share_permissions:	*None*	(0)
customer_id:	*None*	(*None*)
admin_user_email:	*None*	(*None*)
credentials_file:	*None*	(*None*)
tokens_path:	*None*	(*None*)

```

auth_port:          *None*          (*None*)
foreign_container_generation: *None*          (1)
debug:             *None*          (*None*)
Use above plugin configuration? (Yes/mod/no): mod
You have the following choices:
  1: destination_user (Destination User)
  2: destination_path (Destination Path in google-workspace)
  3: send_report (Send report of the restore operation to the affected user)
  4: allow_duplicates (Allow Duplicate Objects (Files with the same name in the same_
↳ folder, emails with same id..))
  5: drive_destination_shared_unit (Destination Shared Unit name)
  6: drive_skip_versions (Skip restoring file former versions (tagged with '###date')_
↳ even if they are selected)
  7: drive_skip_comments (Skip restoring file comments even if they are selected)
  8: drive_skip_sharedwithme (Skip restoring shared with me elements even if they are_
↳ selected)
  9: drive_restore_share_permissions (Restore sharing permissions of the files, so_
↳ they get shared with the same people than original files)
 10: customer_id (Destination Workspace customer id)
 11: admin_user_email (Destination Workspace admin user email)
 12: credentials_file (Credentials file path to be used for authentication)
 13: tokens_path (Directory to store authorization tokens for delegated permissions)
 14: auth_port (Port to receive response from delegated authentication process)
 15: foreign_container_generation (Generate a general container (usually a folder) to_
↳ put inside restored objects coming from different entities)
 16: debug (Change debug level)
Select parameter to modify (1-16): 2
Please enter a value for destination_path: restored1
Plugin Restore Options
Option          Current Value          Default Value
destination_user:  *None*          (*None*)
destination_path:  restored1        (*None*)
send_report:      *None*          (0)
allow_duplicates:  *None*          (1)
drive_destination_shared_unit: *None*          (*None*)
drive_skip_versions: *None*          (1)
drive_skip_comments: *None*          (1)
drive_skip_sharedwithme: *None*          (0)
drive_restore_share_permissions: *None*          (0)
customer_id:      *None*          (*None*)
admin_user_email: *None*          (*None*)
credentials_file: *None*          (*None*)
tokens_path:      *None*          (*None*)
auth_port:        *None*          (*None*)
foreign_container_generation: *None*          (1)
debug:           *None*          (*None*)
Use above plugin configuration? (Yes/mod/no): yes
Run Restore job
JobName:         RestoreFiles
Bootstrap:       /tmp/regress/working/127.0.0.1-dir.restore.2.bsr
Where:           /
Replace:         Always

```

(continues on next page)

```

FileSet:      Full Set
Backup Client: 127.0.0.1-fd
Restore Client: 127.0.0.1-fd
Storage:      File
When:         2022-05-12 18:03:23
Catalog:      MyCatalog
Priority:      10
Plugin Options: User specified
OK to run? (Yes/mod/no): yes
Job queued. JobId=3

```

Restore by service

In this section some example restore configurations will be shown:

- [Google Drive](#)
- [Google Drive](#)

Cross workspace restore

You can perform cross-workspace restores using the restore variables:

- `customer_id`
- `admin_user_email`
- `credentials_file`

Obviously, it is needed to set up the destination workspace values, where a connection application should have been also set up previously to allow the connection.

8.3 List

It is possible to list information using the `bconsole .ls` command and providing a path. In general, we need to provide the service parameter, the implied entity and a path representing a folder.

There are some general commands (like listing users), while the rest of the commands need to have the service set

Below some examples:

List general info: Users of a workspace

Here we are showing these 3 commands using the `bconsole .ls` command, but notice you may also use them with the query interface (keep your variable values, but apply something like: `.query plugin="..." client=xxx parameter=xxx`)

Listing 26: **List example: General information**

```

*.ls plugin="gw: credentials_file=/opt/bacula/etc/bacula-gw-plugin-credentials.json
↪customer_id=Alo9783c12 admin_user_email=jorge@baculasystems.com" client=127.0.0.1-fd
↪path=users
Connecting to Client 127.0.0.1-fd at 127.0.0.1:8102
-rw-r----- 1 nobody nogroup -1 1970-01-01 00:59:59 /
↪jorge@baculasystems.com

```

(continues on next page)

(continued from previous page)

```
-rw-r----- 1 nobody nogroup          -1 1970-01-01 00:59:59 /  
↳kara@baculasystems.com  
-rw-r----- 1 nobody nogroup          -1 1970-01-01 00:59:59 /  
↳john@baculasystems.com  
2000 OK estimate files=3 bytes=0
```

List Google Drive contents

Listing 27: List example: General information

```
*.ls plugin="gw: credentials_file=/opt/bacula/etc/bacula-gw-plugin-credentials.json  
↳customer_id=Alo9783c12 admin_user_email=jorge@baculasystems.com  
↳user=jorge@baculasystems.com service=drive" client=127.0.0.1-fd path=  
Connecting to Client 127.0.0.1-fd at 127.0.0.1:8102  
-rw-r----- 1 nobody nogroup          373568 2022-05-13 11:24:51 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-13_11.24.  
↳33_06.html  
-rw-r----- 1 nobody nogroup          372115 2022-05-12 18:03:51 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-12_18.03.  
↳36_10.html  
drwxr-xr-x 1 nobody nogroup          -1 2022-05-12 18:03:41 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/restored1/  
-rw-r----- 1 nobody nogroup          373602 2022-05-12 17:49:52 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-12_17.49.  
↳34_06.html  
drwxr-xr-x 1 nobody nogroup          -1 2022-05-12 17:49:39 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220512174934/  
drwxr-xr-x 1 nobody nogroup          -1 2022-05-12 17:48:25 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220512174729/  
-rw-r----- 1 nobody nogroup          372471 2022-05-12 10:47:16 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-12_10.47.  
↳03_10.html  
drwxr-xr-x 1 nobody nogroup          -1 2022-05-12 10:47:08 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/RESTORED_SKIPVER_REGRESS_20220512104703/  
-rw-r----- 1 nobody nogroup          376389 2022-05-12 10:46:50 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-12_10.46.  
↳32_06.html  
drwxr-xr-x 1 nobody nogroup          -1 2022-05-12 10:44:35 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/SOURCE_REGRESS_20220512104335/  
-rw-r----- 1 nobody nogroup          376609 2022-05-10 12:56:36 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_12.56.  
↳12_06.html  
drwxr-xr-x 1 nobody nogroup          -1 2022-05-10 12:56:17 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220510125612/  
drwxr-xr-x 1 nobody nogroup          -1 2022-05-10 12:55:31 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/SRC_INCLUDE_REGRESS_20220510125529/  
drwxr-xr-x 1 nobody nogroup          -1 2022-05-10 12:54:57 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/trash/SRC_REMOVE_REGRESS_20220510125403/  
-rw-r----- 1 nobody nogroup          373736 2022-05-10 12:49:34 /@gw/C02uv9t30/  
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_12.49.  
↳18_10.html
```

(continues on next page)

(continued from previous page)

```
-rw-r----- 1 nobody nogroup          376810 2022-05-10 12:49:05 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_12.48.
↳42_06.html
-rw-r----- 1 nobody nogroup          374586 2022-05-10 12:31:51 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_12.31.
↳28_06.html
drwxr-xr-x  1 nobody nogroup           -1 2022-05-10 12:31:33 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220510123128/
drwxr-xr-x  1 nobody nogroup           -1 2022-05-10 12:30:10 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/trash/SRC_REMOVE_REGRESS_20220510123006/
-rw-r----- 1 nobody nogroup          372465 2022-05-10 12:29:46 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_12.29.
↳31_10.html
-rw-r----- 1 nobody nogroup          376395 2022-05-10 12:29:19 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_12.28.
↳58_06.html
-rw-r----- 1 nobody nogroup          372210 2022-05-10 12:25:23 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_12.25.
↳04_06.html
drwxr-xr-x  1 nobody nogroup           -1 2022-05-10 12:25:09 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220510122504/
drwxr-xr-x  1 nobody nogroup           -1 2022-05-10 12:23:48 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220510122254/
-rw-r----- 1 nobody nogroup          372218 2022-05-10 11:38:20 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_11.38.
↳04_06.html
drwxr-xr-x  1 nobody nogroup           -1 2022-05-10 11:38:09 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220510113804/
drwxr-xr-x  1 nobody nogroup           -1 2022-05-10 11:36:53 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220510113557/
-rw-r----- 1 nobody nogroup          372210 2022-05-10 11:34:28 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_11.34.
↳11_06.html
-rw-r----- 1 nobody nogroup          372196 2022-05-10 11:27:03 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_11.26.
↳47_06.html
-rw-r----- 1 nobody nogroup          373412 2022-05-10 11:23:53 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_11.23.
↳39_06.html
-rw-r----- 1 nobody nogroup          373574 2022-05-10 11:21:10 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-10_11.20.
↳53_06.html
drwxr-xr-x  1 nobody nogroup           -1 2022-05-09 18:26:39 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/DoComplicateMyLife/
-rw-r----- 1 nobody nogroup          370347 2022-05-09 18:24:18 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-09_18.24.
↳08_18.html
drwxr-xr-x  1 nobody nogroup           -1 2022-05-09 18:24:13 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/testingMyRestore/
-rw-r----- 1 nobody nogroup          373586 2022-05-09 17:56:49 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-09_17.56.
↳33_06.html
```

(continues on next page)

(continued from previous page)

```
drwxr-xr-x 1 nobody nogroup -1 2022-05-09 17:56:38 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220509175633/
-rw-r----- 1 nobody nogroup 373701 2022-05-07 14:16:03 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-07_14.15.
↳48_10.html
-rw-r----- 1 nobody nogroup 376757 2022-05-07 14:15:36 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-07_14.15.
↳15_06.html
-rw-r----- 1 nobody nogroup 371504 2022-05-07 13:57:22 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-07_13.56.
↳39_13.html
drwxr-xr-x 1 nobody nogroup -1 2022-05-07 13:56:44 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/ANADALRG14/
-rw-r----- 1 nobody nogroup 371491 2022-05-07 13:21:14 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-07_13.20.
↳31_12.html
drwxr-xr-x 1 nobody nogroup -1 2022-05-07 13:20:36 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/ADjoker/
-rw-r----- 1 nobody nogroup 382293 2022-05-07 12:35:15 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-07_12.34.
↳32_11.html
-rw-r----- 1 nobody nogroup 23352 2022-05-07 12:34:41 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/Brute.jpeg
-rw-r----- 1 nobody nogroup 18394 2022-05-07 12:34:40 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/Ultricies.txt
-rw-r----- 1 nobody nogroup 8693 2022-05-07 12:34:40 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/Suavitae.ppt
-rw-r----- 1 nobody nogroup 14981 2022-05-07 12:34:40 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/Cetero.txt
-rw-r----- 1 nobody nogroup 15752 2022-05-07 12:34:39 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/Fastidii.ppt
drwxr-xr-x 1 nobody nogroup -1 2022-05-07 12:34:38 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/AFullRestore1/
-rw-r----- 1 nobody nogroup 370327 2022-05-07 12:15:58 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-07_12.15.
↳48_10.html
drwxr-xr-x 1 nobody nogroup -1 2022-05-07 12:15:54 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/AverAlc/
drwxr-xr-x 1 nobody nogroup -1 2022-05-06 17:12:34 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220506171229/
drwxr-xr-x 1 nobody nogroup -1 2022-05-06 16:58:13 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220506165808/
-rw-r----- 1 nobody nogroup 372243 2022-05-06 13:49:21 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-06_13.49.
↳06_06.html
-rw-r----- 1 nobody nogroup 372215 2022-05-06 13:39:36 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-06_13.39.
↳20_06.html
-rw-r----- 1 nobody nogroup 372167 2022-05-06 13:13:51 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-06_13.13.
↳33_06.html
-rw-r----- 1 nobody nogroup 372444 2022-05-06 13:09:03 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-06_13.09.
↳47_10.html
```


(continued from previous page)

```
-rw-r----- 1 nobody nogroup          376549 2022-05-06 13:08:34 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-06_13.08.
↳11_06.html
-rw-r----- 1 nobody nogroup          372396 2022-05-06 11:20:51 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-06_11.20.
↳37_10.html
-rw-r----- 1 nobody nogroup          376416 2022-05-06 11:20:24 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-06_11.20.
↳06_06.html
-rw-r----- 1 nobody nogroup          373532 2022-05-06 09:51:58 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-06_09.51.
↳43_06.html
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:19:36 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/AutoSimple 2022-05-05 05.19.34/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:14:12 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/AutoSimple 2022-05-05 05.14.10/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:03:02 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:02:51 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:02:46 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:02:35 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:02:31 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:02:23 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:02:20 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:02:12 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:02:09 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 17:01:58 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505170105/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 16:59:34 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 16:59:22 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 16:59:18 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 16:59:06 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 16:59:02 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 16:58:51 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 16:58:49 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x  1 nobody nogroup            -1 2022-05-05 16:58:39 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
```

(continues on next page)

(continued from previous page)

```
drwxr-xr-x 1 nobody nogroup -1 2022-05-05 16:58:36 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x 1 nobody nogroup -1 2022-05-05 16:58:27 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220505165734/
drwxr-xr-x 1 nobody nogroup -1 2022-05-05 16:56:25 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/AutoSimple 2022-05-05 04.56.13/
drwxr-xr-x 1 nobody nogroup -1 2022-05-05 16:56:15 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/AutoSimple 2022-05-05 04.56.13/
-rw-r----- 1 nobody nogroup 373488 2022-05-05 13:55:46 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-05_13.55.
↳30_06.html
-rw-r----- 1 nobody nogroup 373499 2022-05-04 12:03:24 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-04_12.03.
↳07_06.html
drwxr-xr-x 1 nobody nogroup -1 2022-05-04 12:03:12 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220504120307/
drwxr-xr-x 1 nobody nogroup -1 2022-05-04 12:02:01 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220504120058/
-rw-r----- 1 nobody nogroup 373510 2022-05-03 13:32:34 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-03_13.32.
↳18_06.html
drwxr-xr-x 1 nobody nogroup -1 2022-05-03 13:32:23 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220503133217/
drwxr-xr-x 1 nobody nogroup -1 2022-05-03 13:31:04 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220503133010/
-rw-r----- 1 nobody nogroup 371497 2022-05-03 13:26:48 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-03_13.26.
↳36_06.html
drwxr-xr-x 1 nobody nogroup -1 2022-05-03 13:26:41 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/REGRESS_20220503132636/
-rw-r----- 1 nobody nogroup 372431 2022-05-02 17:53:43 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-02_17.53.
↳29_10.html
-rw-r----- 1 nobody nogroup 376394 2022-05-02 17:53:17 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-02_17.52.
↳56_06.html
-rw-r----- 1 nobody nogroup 373668 2022-05-02 13:59:36 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-02_13.59.
↳19_10.html
-rw-r----- 1 nobody nogroup 375148 2022-05-02 13:59:09 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-02_13.58.
↳48_06.html
drwxr-xr-x 1 nobody nogroup -1 2022-05-02 13:56:52 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/SOURCE_REGRESS_20220502135650/
-rw-r----- 1 nobody nogroup 372168 2022-05-02 13:39:07 /@gw/C02uv9t30/
↳users/jorge@baculasystems.com/drive/my drive/BEE_RestoreReport_DRIVE_2022-05-02_13.38.
↳51_06.html
2000 OK estimate files=100 bytes=15,775,509
```

Other query/list examples

Listing 28: Query/List examples

```
*.ls plugin="gw: credentials_file=/opt/bacula/etc/bacula-gw-plugin-credentials.json
↳customer_id=Alo9783c12 admin_user_email=jorge@baculasystems.com
↳user=jorge@baculasystems.com service=drive" client=127.0.0.1-fd path=/folder1

List emails inside inbox
*.ls plugin="gw: credentials_file=/opt/bacula/etc/bacula-gw-plugin-credentials.json
↳customer_id=Alo9783c12 admin_user_email=jorge@baculasystems.com
↳user=jorge@baculasystems.com service=email" client=127.0.0.1-fd path=/inbox

Show free users loggedin
*.query plugin="gw: credentials_file=/opt/bacula/etc/bacula-gw-plugin-free.json"
↳client=127.0.0.1-fd parameter=logged-users

Force login of a particular free user
*.query plugin="gw: credentials_file=/opt/bacula/etc/bacula-gw-plugin-free.json"
↳client=127.0.0.1-fd parameter=login:myuser@gmail.com
```

9 Best practices

9.1 Jobs Distribution

It is recommended to split the target backup between different groups of entities or even having one job per entity (user, drive unit, etc). This way errors in one job will not invalidate a whole backup cycle where some entities have been successful and some others had errors. This also makes it easier to identify the cause of the error.

9.2 Concurrency

Google Workspace APIs impose a variety of boundaries that need to be considered. If a boundary is crossed, the corresponding API call will fail and the application will need to wait some amount of time to retry, which is different depending on the boundary crossed.

It is crucial to plan an adequate strategy to backup all the elements without reaching API boundaries. A single job implements some parallelism which can be reduced until a point, if necessary, using the variable **backup_queue_size** (default value is 30). This variable controls the size of the internal queues communicating the internal threads, that are designed to fetch, open and send every item to Bacula core. Reducing its size will produce, ultimately (with a value of 1 for example), an execution very similar to a single threaded process. On the other hand the plugin has **concurrent_threads** which controls the number of simultaneous processes fetching and downloading data (default value is 5).

Caution is recommended with the concurrency over the same service (in general, it is recommended a maximum of 4-5 jobs or threads working with the same service) and plan a step-by-step testing scenario before putting it into production. Other important point is the timing schedule, as some boundaries are related to time-frames (number of request per 10 minutes or 1 hour, for example). If you detect you reach boundaries when running all your backups during a single day of the week, please try to use 2 or 3 days and spread the load through them in order to achieve better performance results.

Specifically for the GMail module, in addition to concurrency the plugin uses batch requests that are processed in parallel as soon as it gets the answer. Therefore, throttling can be reached very easily and it's recommended to not use

almost any concurrency with this module. By default, GMail uses 2 threads in parallel. Even with 2 it is expected to have some request throttled. Limits can be raised under request with Google, but if this is not a possibility and you experience throttling problems with parallelism we recommend to disable it completely (setting `concurrent_threads = 1`).

More information about Google Workspace API boundaries may be found here:

<https://developers.google.com/drive/api/guides/limits> <https://developers.google.com/gmail/api/reference/quota>

9.3 Performance

The performance of this plugin is highly dependent on many external factors:

- ISP latency and bandwidth
- Network infrastructure
- FD Host hardware
- FD Load
- ...

In summary, it is not possible to establish an exact reference about how much time a backup will need to complete.

Some general guidelines to understand the performance we can get:

- Many little objects to protect -> More objects per second, but less speed (MB/s)
- Big files to protect -> Less objects per second, but greater speed (MB/s)

It is recommended to benchmark your own environment in base to your requirements and needs.

The automatic concurrency mechanism (using `concurrent_threads=x`, default is 5) should work well for most scenarios, however, fine tune is possible if we define one job per entity and we control how many of them run in parallel, together to decrease the `concurrent_threads` value in order to avoid throttling from Google Cloud APIs.

There are many different possible strategies to use this plugin, so please, study what is best suiting for your needs before deploying the jobs for your entire environment, so you can get best possible results:

- You can have a job per entity (users, shared drives...) and all services
- You can split your workload through a schedule, or try to run all your jobs together.
- You can run jobs in parallel or take advantage of `concurrent_threads` and so run less jobs in parallel
- You can backup whole services to backup or select precisely what elements you really need inside each service (folders, paths, exclusions...)
- etc.

Specifically for Drive service, in order to maximize the performance we recommend additionally to: - Disable comments backup - Disable version history backup - Run one job per user and use the full Drive (no path selection) so the Delta function is applied. Exclude all shared units in user jobs (`drive_shared_units_regex_exclude=.*`) - Run one job per shared unit and use the full Drive (no path selection) so the Delta function is applied. Exclude all users in shared unit jobs (`users_regex_exclude=.*`)

10 Troubleshooting

Listed in this section are some scenarios that are known to cause issues.

10.1 Out of Memory

If you ever face *OutOfMemory* errors of the Java daemon (you will find them in the gw-debug.err file),

you are likely using a high level of concurrency through internal `concurrent_threads` parameter and/or parallel jobs.

To overcome this situation you can:

- a) Reduce `concurrent_threads` parameter
- b) Reduce the number of jobs running in parallel
- c) If you cannot do that you should increase JVM memory.

To increase JVM memory, you will need to:

Create a this file: `/opt/bacula/etc/gw_backend.conf`.

Below, an example of the contents: `GW_JVM_MIN=2G GW_JVM_MAX=8G`

Those values will define the MIN (`GW_JVM_MIN`) and MAX (`GW_JVM_MAX`) memory values assigned to the JVM Heap size. In this example we are setting 2Gb for the minimum, and 8Gb for the maximum. In general, those values should be more than enough. Please, be careful if you are running jobs in parallel, as very big values and several jobs at a time could quickly eat all the memory of your host.

The `/opt/bacula/etc/gw_backend.conf` won't be modified through package upgrades, so your memory settings will be persistent.