Google Workspace Migrate: Data migration best practices

Copyright, Trademarks, and Legal
Google, LLC.
1600 Amphitheatre Parkway Mountain View, CA 94043
www.google.com
OCT112024

About this guide

The goal of this document is to guide approved Google Workspace delivery partners in planning migrations of data (mails, files and so on) to Google Workspace.

The document also does the following:

- Provides best practices about the installation and configuration of Google Workspace Migrate
- Provides best practices about planning a data migration
- Specifies considerations to take into account when planning a migration
- Identifies monitoring points to avoid issues that can occur during a migration

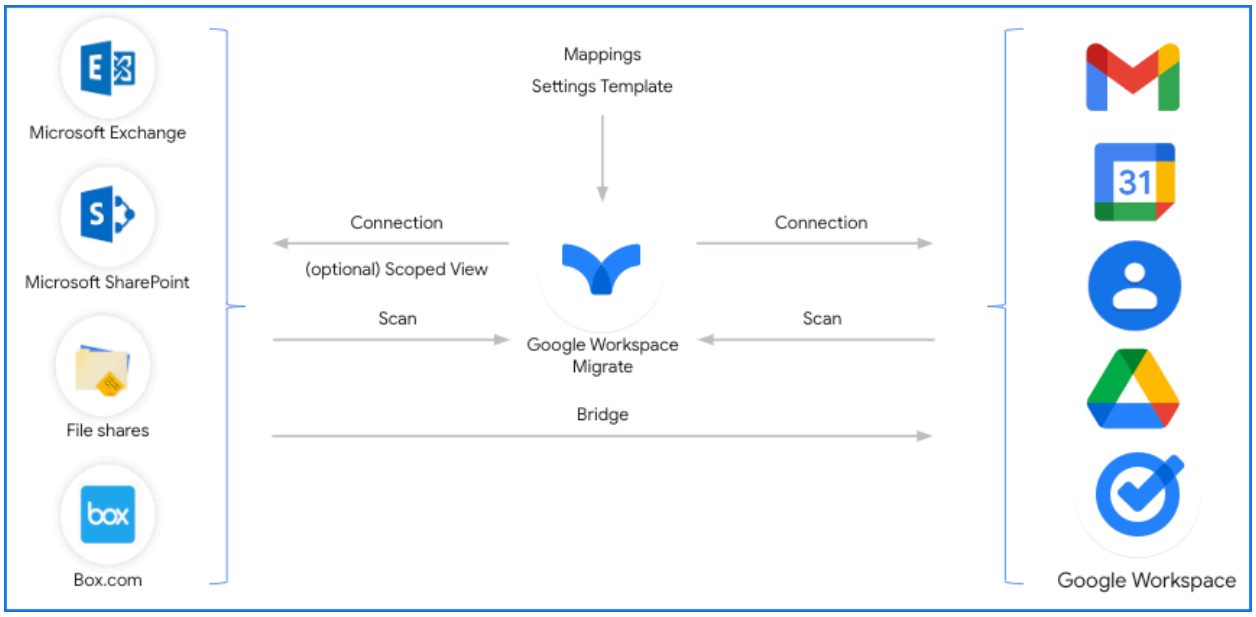
Important: Each customer setup is unique, and no migration is exactly the same. Although this document makes best practice recommendations, choose the precise settings for a specific project based on that project's requirements. Before applying any settings mentioned in this document, evaluate the impact with your customer.

What's covered in this guide?

This document provides an overview of the following topics at an expert level and includes links to additional resources for each one:

- Infrastructure requirements
- Configuration best practices
- Google Workspace Migrate settings and their effects
- Limits & quotas
- Reconciliation
- Migration resourcing and project management (time, cost, and milestones)

Terms to know



- **Connection**–Migration source or target environment, including its configuration and authentication details
- **Scoped view**–A view of content in the source environment that has been limited using a mapping to allow it to be scanned more easily or quickly
- **List**–A collection of entities within Google Workspace Migrate that are tied to a specific connection, including:
    - Locations (SharePoint sites)– A list required for migrating data from multiple site collections
    - Users (mailboxes)– A set of mailboxes used to specify what should be migrated (typically all of the users from migration scope) or to restrict migration to only certain users
    - Users (drive sharding)– A list of temporary user accounts used to distribute file creation in Drive to avoid upload limitations and use parallel processing
- **Mapping**–An association of a content source location or user with a target location or user
    - **Identity mapping**–Maps all users that should be included in a particular migration (bridge) as well as permissions, targets and calendar resources
    - **Role mapping**–Used to map SharePoint permission levels to Drive access levels
    - **Data mapping**–Used to determine what content gets migrated and where
- **Settings template**–A collection of migration settings grouped by source data type
- **Scan**–The act of scanning a source connection, used to optimize scheduling
- **Bridge**– A complete set of migration settings, effectively, the “bridge” over which data

is migrated, that Includes connection details, migration configuration, scheduling, reporting, and job run settings

- **Partition**–Migration processing tasks being run on nodes, with up to 6 actions (partitions) running on each node, depending on the source

Introduction

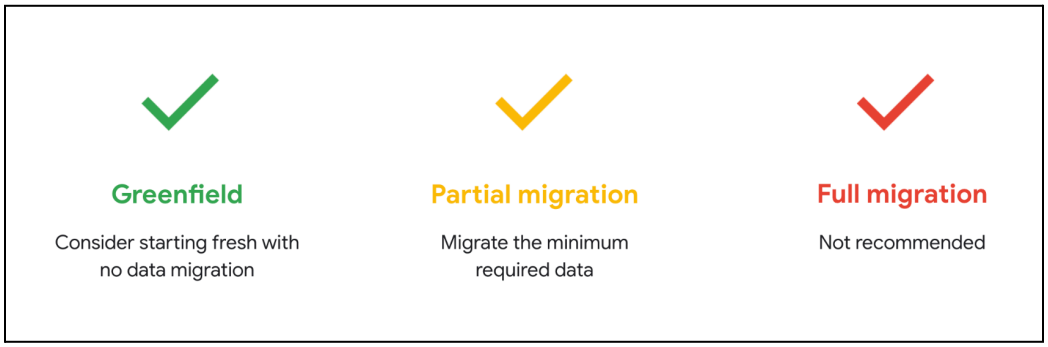
Deploying Google Workspace for enterprise organizations almost always requires existing data, usually mail and files, to be migrated into the new environment. Google Workspace Migrate is a Google built and supported tool that has been used successfully for the largest and most challenging migration projects. In addition to supporting the migration of data from Exchange, Box, SharePoint, and file share systems, it offers advanced features for source scanning and horizontal scaling to increase throughput.

One of the key factors in using Google Workspace Migrate successfully is designing the appropriate architecture to support your migration requirements. This document outlines how to do this based on many years' experience from Google Professional Services, Support, and engineering teams skilled in this area.

Planning a migration

Before you look at migration architecture, we recommend planning the migration in advance, following our standard deployment methodology, and for most large migrations, planning to execute it in multiple phases: core IT, early adopters, and the go-live.

At a high level, the best practice for a migration ahead of the go-live is to identify what data is needed for users to be productive on day one and migrate only that, with users given read-only access to legacy data. In most enterprise organizations, legacy data (for example, files not accessed in the last 12 months) is not actively used, and omitting it can significantly speed up overall migration timelines.



When planning the migration, we recommend conducting detailed discovery for these items:

- **Domains**
  Understand how many domains are in the source environment. Having this information is important to make changes in the DNS panel during the go-live, as well as when onboarding the customer.

- **Users**
  Analyzing user data is crucial in planning because it can help you scope the migration (for example, identifying users who won't be migrated, or users that require some coexistence of new and legacy tools).

- **Shared mailboxes**
  Shared mailboxes must be migrated to Google Workspace as generic users or as groups, so it is important to clarify which option best fits your customer's needs.

- **SharePoint sites and data**
  SharePoint sites can be migrated to Google Drive as shared drives. However, access permissioning in Drive works differently than that in SharePoint. For example, it is not possible to have more restrictive permissions in subfolders. Analyzing the sharing structure in advance is important.

- **Mail data**
  This information is critical in estimating the overall migration timeline because it lets you establish the number of items per user or the amount of data per user, as well as identify how users should be batched for migration.

- **Local or regional regulations or restrictions**
  Different regions might have specific legal or regulatory requirements for handling personally identifiable information and also could have specific limitations on the bandwidth or capacity available for data migration.

- **Network**
  Ensure that customers make the necessary changes to their current network for all the locations to allow users to adopt Google services without interruption.

Although this list covers the main items needing additional discovery, it is not exhaustive. We recommend that you refer to the following presentation for more details (presentation is for

Partners only and is accessible through the Partner Advantage portal):
Google Workspace Data Migration Technical Kickoff | PSO | Y22

Approach to a migration

Every migration is different and brings different complexities.

It is crucial to understand the source data. What's the data type? What's the size and complexity of the data? How is the data distributed? Is there any unsupported data type? Are we exceeding any hard limits in Google Workspace?

Having a clear understanding of the data to migrate results in better planning.

As suggested in the following section, Google Workspace Migrate scans are a powerful way to understand source data. Scans also allow the optimization of the workload distribution across different nodes and correctly prioritize large individual datasets, so that they won't become a bottleneck for the migration.

When approaching a migration, follow the Google Workspace deployment methodology (core-IT, early adopters, go-live) dividing the migration into multiple phases, and leverage delta migrations before the go-live.

Consider what data needs to be migrated for the go-live and what data needs to be migrated at all. It might be not necessary to migrate all the data living at the source. Some data can also be migrated after the go-live to speed up the migration timeline.

Additionally, problems with source data can also affect the duration of a migration. That's another reason to perform a good discovery of source data.

In greater detail, we have listed the following best practices based on running large migrations with Google Workspace Migrate.

Do's and Don'ts

Do

- **Use the latest version of the software.**
  Google regularly updates the Google Workspace Migrate platform and node software and occasionally updates the database software. You should use the latest version of

the software to ensure that you have the most recent Google Workspace Migrate features and fixes. For details, go to Upgrade the software.

- **Estimate the number of nodes needed to complete the migration.**
  Start with fewer nodes, observe the process, and increase the number of nodes as needed. This approach helps you to save budget by allowing you to use only as many virtual machines (VMs) as necessary. Do not exceed 40 nodes per cluster (the maximum number of supported nodes per cluster). If your workload requires more nodes, provision additional clusters. For more information about nodes, go to Install the node servers.

- **Invest time in scans to analyze source data.**
  Scans can be time-consuming, but this upfront investment of time helps to optimize the migration process and is helpful for reconciliation. In general, scans should be performed as close as possible to the start of migration rather than weeks or months in advance. When it's not possible to perform scans, collect the technical findings from the source environment using PowerShell, comma-separated values (CSV) file exports from the interface, or any other method that allows you to collect these values:
    - List of accounts and groups
    - Other resources that are in scope of migration (for example, calendar resources)
    - Number of items per user per service

  These details will help you to distribute loads across different bridges, and calculate the estimated time needed to complete the migration.

- **Separate different services into bridges.**
  For example, migrate Google Calendar, Google Contacts, and Google Tasks on one bridge and Exchange or SharePoint data on another bridge. This approach allows you to troubleshoot each service individually, so if you encounter issues with one service, you don't have to stop the overall process.

- **Avoid bigger databases if possible.**
  Bigger databases are slower in general. When the number of records in the project database is very high (~ 2 billion records) your migration might get stuck. If the project database size is getting closer to 3 TB , do not continue the migration in that project. Splitting the data into more projects when a large amount of data is migrated helps to avoid performance issues.

- **Make sure that you discuss the monitoring points with your customer.**
  The list of the monitoring points depends on the source (for details, go to References).

- **Actively discuss and agree on the migration scope with customers in advance.**
  Although customers often want to migrate all data from their source systems, this is not recommended and could affect migration cost and timings. Similarly, omitting the migration of data that is time-consuming, such as file revisions, can speed up the overall process.

- **Always follow the recommended system requirements.**
  For example, when high throughput is a business critical need, consider using IOPs optimized CPUs and Disks for database VMs.

- **Migrate in phases.**
  Ideally, distribute large data across different phases and use delta migrations before the go-live.

- **Turn off automatic Microsoft Windows Updates & Windows antivirus protection.**
  Turn off the services on all migration servers. Service updates can potentially reset servers in the middle of migration, while antivirus services can scan all data and consume virtual memory.

- **Use Google Cloud to install Google Workspace Migrate.**
  Google has extensively tested Google Workspace Migrate in Google Cloud. Using the product on Google infrastructure offers advantages such as flexibility and direct network connections to Google Workspace infrastructure to reduce latency.

- **Use Google Chrome browser as your default browser.**

- **Make sure to reduce network latency.**

- **Make sure to respect the system requirements.**
  For details, go to System requirements.

- **Create a new project.**
  If you have run a test migration and now you're moving to a production migration, create a new project to migrate the same data to production. Some data might be skipped by Google Workspace Migrate if you're using the same data in the same project as the source.

Don'ts

- **Don’t migrate the full scope of the data.**
  Try to limit it by date. If it’s necessary to migrate everything, identify long tail users (those with significant volumes of data) and run them on a separate bridge. This method allows you to handle the migration of core data for other users separately, using delta migrations to bring over older data afterwards.

- **Don’t move batched users across clusters.**
  The migration process stores data in the database and moving users between clusters can cause data duplication.
- **Don’t have one source user mapped to multiple destination users.**
  Doing so can result in unexpected outcomes (some data might not be migrated, because it’s already in the database marked as migrated). If you need to do this (for example, because you need to migrate the shared mailbox content into multiple accounts), use separate projects and do not run them simultaneously.

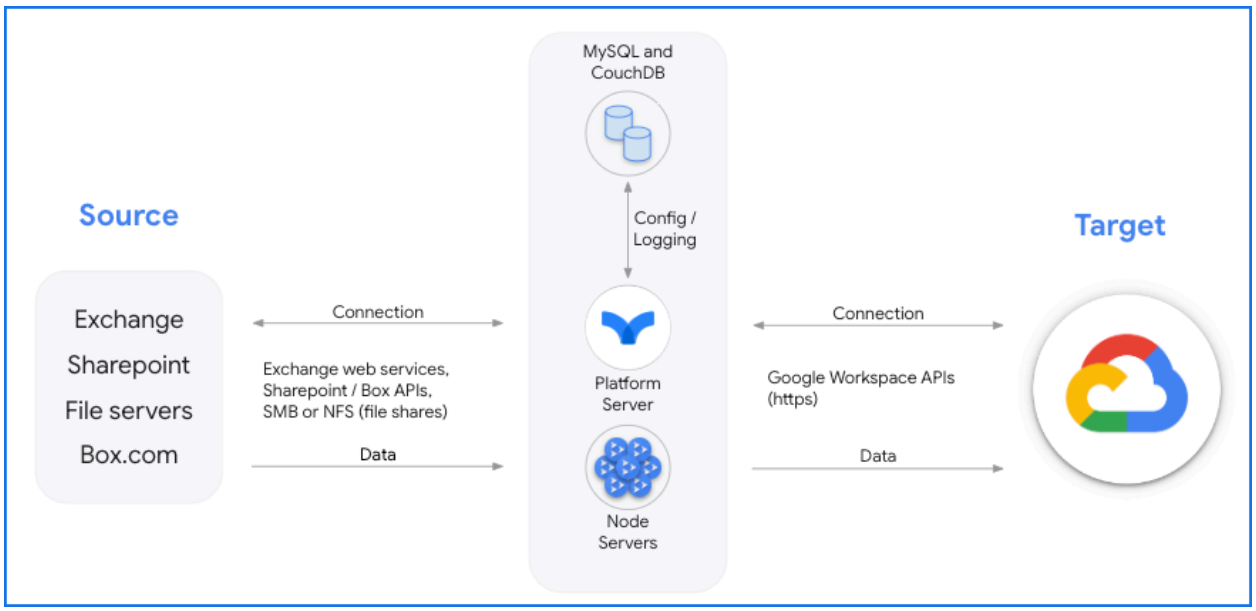
- **Don’t migrate all types of data on a single bridge.**
  Single threading on just one bridge means that, if issues occur, you have to stop everything. It's best practice to use two bridges where one is dedicated for Exchange data migration and the second is for OneDrive or SharePoint migrations.

- **Don't drop off any node server even though its status is not busy.**
  Dropping off might cause some issues when you are in the middle of migration. In some scenarios the status might change in the middle of migration. When the bridge is not running, it’s fine to turn off the nodes to save on costs.

Infrastructure requirements

Google Workspace Migrate consists of several components, including databases, a platform server, and potentially many node servers used for parallelization. The high-level architecture is illustrated in the following diagram.

Architecture



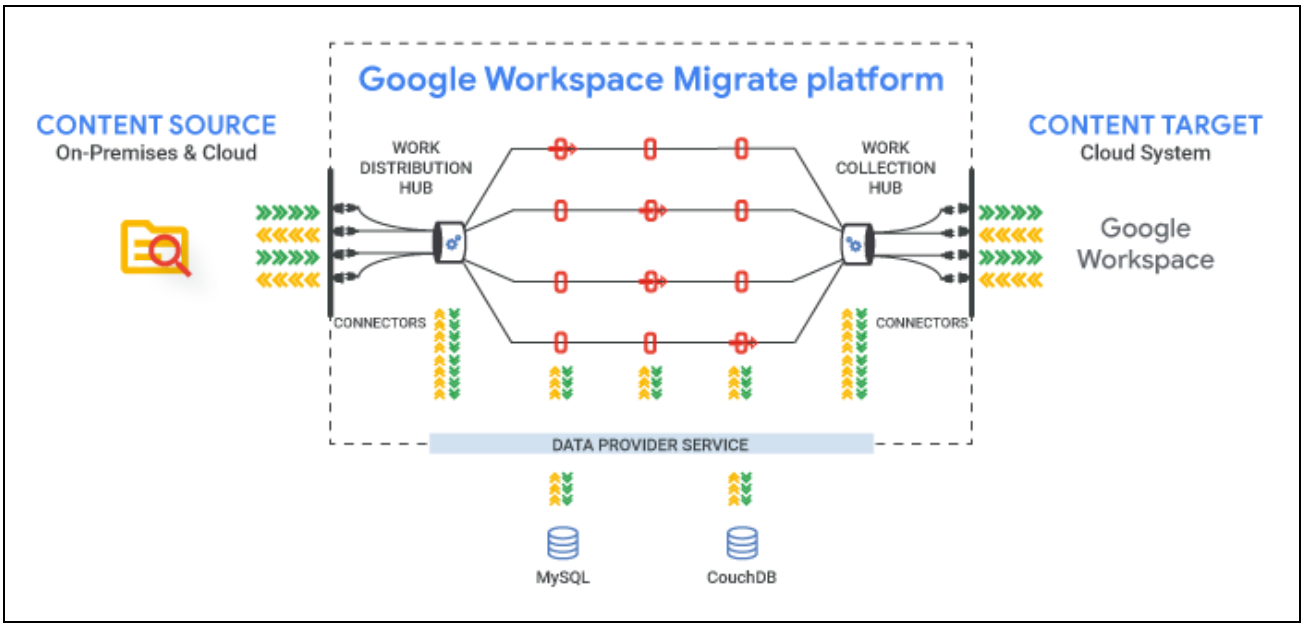
Components

- **Platform server**–Functions as the main control panel of the overall migration system and is used to configure and run the migration bridges
- **MySQL database**–Stores high-level connection and project configuration details, mappings of all objects, metadata for objects, lists and settings templates, scan settings, logs and metadata, project user permissions, and transaction and execution information (except for detailed logging data)
- **CouchDB database**–Stores records of all user mappings and configuration information, including list content, settings templates content, bridge logs and settings, scan settings, report data, and detailed logging information for transactions
- **Nodes**–Run tasks from the queue generated by the platform server

Google Workspace Migrate components run on infrastructure, typically VMs, that must be configured for them. When doing so, always follow the guidance in the system requirements.

Security

The Google Workspace Migrate platform uses multiple asynchronous migration pipelines to achieve maximum throughput. You can set up the platform across several migration servers that are managed and secured from a central location.



Ports

Google Workspace Migrate uses the following ports during a migration:

- **3306**—The default port for the MySQL database
- **5131**—The port that sets the callback address nodes that communicate with the Google Workspace Migrate platform
- **5984**—The default port for the CouchDB database

After you install the node servers, you can optionally set up a port with a TLS certificate. You can then use the port on the node server.

For detailed information about the security of Google Workspace Migrate products, refer to Google Workspace Migrate security overview.

Project setup

Having provisioned the infrastructure to support Google Workspace Migrate, you now need to set up the migration project.

Before configuring migration sources and destinations, you should apply these best practices to the project itself.

- **Do not exceed 10,000 users per project**–If you plan bigger migrations, consider

scaling horizontally (more clusters), or create another project (these can't run simultaneously) and migrate in batches.

- **Monitor disk usage for the MySQL database**–The database requires 1 TB per 200 million migrated objects, and it should be sized based on the scan results or on test migrations. In general, the larger the database size, the slower the performance.
- **If you are planning on running test migrations, use a separate project**–Google Workspace Migrate recognizes the data that has been migrated within a project, so if you plan to migrate users who have already been part of a test migration, set up a new project.

Connections

Establishing the connection is the very first step to configure Google Workspace Migrate. The connection allows you to connect both with the source environment to perform a scan and fetch data and with the destination environment to migrate the data.

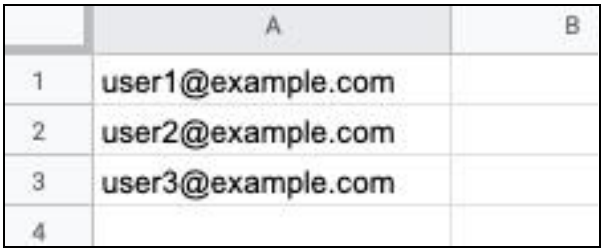
Supported connections

- Google Workspace
- Exchange Server
- Office 365
- Exchange 2010
- Exchange 2010 SP1
- Exchange 2010 SP2
- Exchange 2010 SP3
- Exchange 2013
- Exchange 2016
- Exchange 2019
- SharePoint Server
- Box.com
- Windows file shares

Create a user list

You create a user list to define the data that's available to a connection. For example, you can list the mailboxes in Microsoft Exchange you want to migrate. Make sure your user list includes all the users you're migrating.

The steps to create a user list for Exchange are described in Create a user list.



	A	B
1	user1@example.com	
2	user2@example.com	
3	user3@example.com	
4		

Source connection: Exchange

Help Center quick link - Add or edit an Exchange connection

Important: If you're migrating from Exchange Online (Microsoft 365), use Google Workspace Migrate version 2.4.18.0 or later to make sure you're using a version that supports the authentication process for Exchange Online connections (using a client ID and client secret). For details, go to Deprecation of the ApplicationImpersonation role in Exchange Online.

Google Workspace Migrate connects to Exchange Web Services, typically using **https://<server name>/EWS/Exchange.asmx**.

The default URL for Office 365 is **https://outlook.office365.com/EWS/Exchange.asmx**.

(Optional) You can also use these Powershell snippets to connect to Exchange online from Microsoft environments (for example, when you need to list some content from Exchange, or check any details about certain resources):

```
$UserCredential = Get-Credential
```

to establish a session:

```
$Session = New-PSSession -ConfigurationName Microsoft.Exchange -ConnectionUri
https://outlook.office365.com/powershell-liveid/ -Credential $UserCredential
-Authentication Basic -AllowRedirection
```

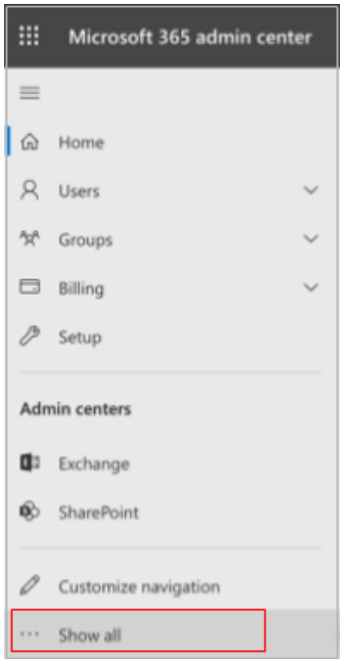
to import the session:

```
Import-PSSession $Session -DisableNameChecking
```

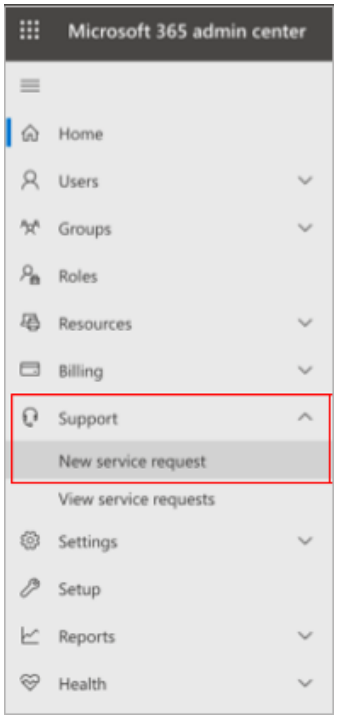
Because Exchange connections use Exchange Web Services (EWS), when the number of connections is high (typically In bigger migration projects) and where the source is a Microsoft 365 environment, you might encounter EWS throttling issues.

To manage throttling, we recommend referring to the most up-to-date Microsoft documentation. However, the following process can serve as a guide:

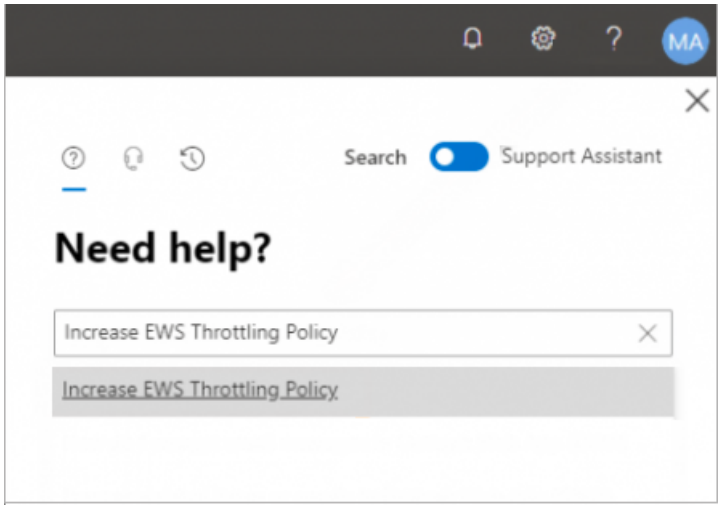
1. Log into the Microsoft 365 admin center as the administrator. On the left side of the navigation menu, under **Admin centers**, click **Show all**.



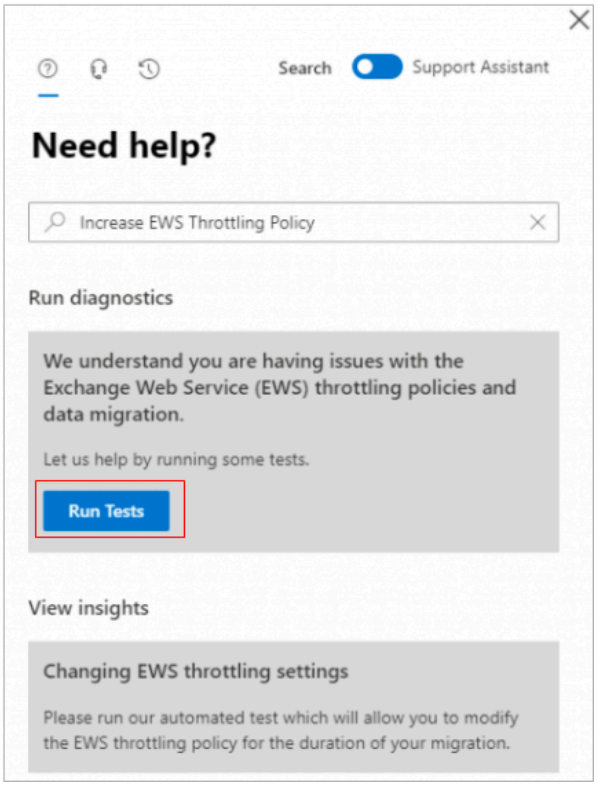
2. Under **Support**, click **New service request**.



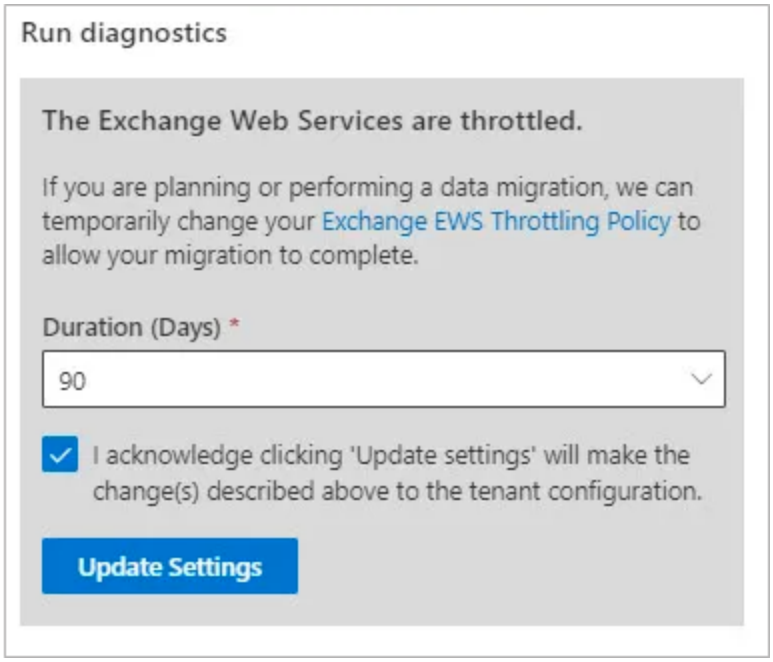
3. Click the Help icon (**?**) in the top right-corner of the page. Search for and select **Increase EWS Throttling Policy**.



4. Click **Run Tests** to determine if EWS is being throttled. It's best to run this option when migration is running. The tests take a couple of minutes to complete.



5. If the diagnostic returns a **The Exchange Web Services are throttled** message, you can change the throttle policy settings. Select your required **Duration (Days)** from the drop-down menu and click **Update Settings**.



You get a confirmation that your EWS throttle policy has been updated. It can take up to 15 minutes for the new setting to take effect.

Source connection: SharePoint & OneDrive

Help Center quick link - Add or edit a SharePoint connection

When connecting to SharePoint, the site URL must be a valid site collection. When you map Mysite root site collection, make sure that you provided a valid location list in the connection settings.

Note that the Google Workspace Migrate SharePoint authentication model changed in version 2.1.18.0 to use role accounts, so you should not attempt to migrate SharePoint data with older versions of Google Workspace Migrate.

Both SharePoint and OneDrive share the same underlying infrastructure, and the Google Workspace Migrate connection settings require an AppID created on the Microsoft side.

The AppID might result in throttling when it is used in connections that are using multiple nodes. Although we recommend using a maximum of 20 nodes per cluster when OneDrive or SharePoint data is migrated, in some situations more than 2-3 nodes per connection (AppID) can result in throttling or even a block on the Microsoft side. The possible workaround for this is to scale the migration horizontally (by adding more clusters, with 2-3 node servers each) or

to use many connections. In this scenario, it is not possible to attach specific node servers to given bridges, so it is important to ensure that a single AppID is in use only by 2 or 3 nodes.

Target connection: Google Workspace

Help Center quick link - Create a Google Workspace target connection

Two service accounts with different scopes are needed for Google Workspace Migrate:

1. A service account for Google Workspace Migrate authentication We recommend that you automate the creation of this service account.
2. A service account with domain wide delegation for accessing target Google Workspace resources

You can also choose to use the same service account for both purposes. Make sure to follow the recommendations in Best practices for using service accounts.

In the Google Cloud project where you created service account #2, make sure that:

- The following Google Workspace APIs are enabled:
  - Admin SDK
  - People API
  - Google Workspace Migrate API
  - Gmail API
  - Google Calendar API
  - Google Drive API
  - Groups Migration API
  - Groups Settings API
  - Google Sheets API
  - Tasks API
- Billing is enabled–This is essential if you need to request a quota increase. You cannot increase quotas for non-billed projects.

Lists

Lists are optional for most migrations. There are two types: user lists and location lists. Lists are attached to the connection configuration.

User list

Help Center quick link - Create a user list (Exchange migrations)

- Contains one email address per line (for example, user@example.com)
- Can be used to:
    - Restrict the mailboxes available for an Exchange scan/migration
    - Expand the mailboxes available for an Exchange scan/migration (used when mailbox search is unavailable or not configured)
    - Specify a list of users for sharding file writes to Drive

Location list

Help Center quick link - Create a location list (SharePoint migrations)

- For SharePoint migrations, this contains one SharePoint site collection per line (for example, https://abc.sharepoint.com/sites/team1).
- These are used to expand the list of URLs accessible to SharePoint.
- You must specify top-level site collections in order to discover the content in lower level collections.

Scoped views

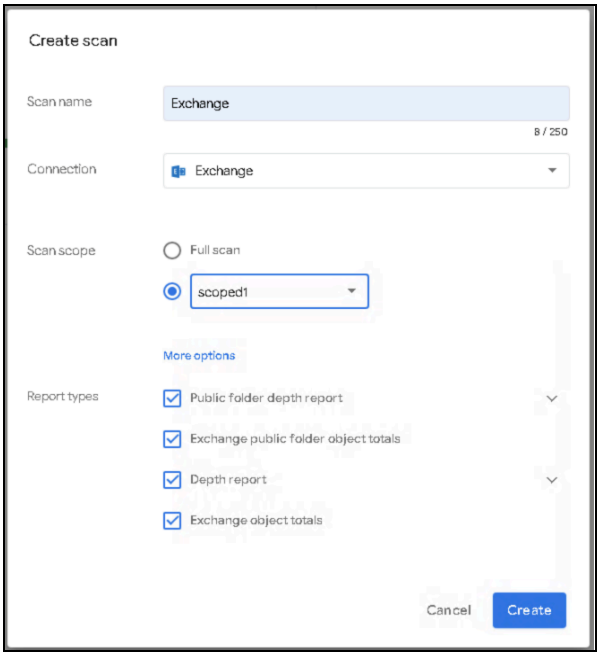
To limit the number of users or drives for a scan, you can use a scoped view to group the data into phases. Doing so enables you to run processes in parallel.

For example, you can scan the source data related to the Core IT phase and also start the Early Adopter scan.

You should set up the users and content in scope for a given migration phase as a scoped view so that the scan can provide accurate counts of how much data needs to be migrated.

You configure scoped views from the connection screen, using the scoped view icon on the right side of the screen:

Go to Scoped view & mapping headers for Exchange for examples of how to create content mapping, including the syntax to use. Once you have configured a scoped view, it will be visible in the **Create scan** screen as shown next.



Scans

Performing scans helps Google Workspace Migrate assess the workload for the partitions during a bridge setup.

Although performing a full scan before the migration can be time-consuming, it provides the following benefits:

- Reports with an estimation of the corpus size, the type of source data, and any issues with the data.
- Optimization of workload distribution across worker nodes. You can set up a bridge directly from the scan to cover the same data.
- Prioritization of large individual datasets that require serial processing (such as users with extensive numbers of emails or files). Doing so ensures that large datasets do not become migration bottlenecks.
- Reconciliation of migration efforts when the migration completes. You can verify that the data you intended to migrate has been copied successfully.

Bridges, nodes & partitions

A way of splitting the load on multiple nodes is to use multiple bridges because each new bridge will consume available nodes.

Up to 6 migration processing tasks (partitions) can be running on each node. During execution you can see how many servers are being used, which can help to reduce the number of nodes (save the cost of Google Cloud usage).

When you migrate from Exchange (email, calendars, contacts, and tasks), you can see that each partition (action) serves about 10 users. In a cluster setup of 40 nodes, this allows for 2400 items (calculated by looking at the number of nodes 40 x 6 partitions x 10 users) simultaneously.

For OneDrive or SharePoint, the load distribution through partitions is different, and one user or mapping is handled by one partition (6 per node).

Mapping

Here are examples of the mapping files that you can use in Google Workspace Migrate while migrating from SharePoint and OneDrive.

Migrate from SharePoint sites to My Drive

Source SPSite	Target GUser	Target GDrive
https://example.sharepoint.com/sites/dev	user1@solarmora.com	GMyDrive

Migrate from SharePoint sites to Shared Drives

Source SPSite	Target GUser	Target GSharedDrive
https://example.sharepoint.com/sites/dev	user1@solarmora.com	BDEz25J3DzFb

Migrate from OneDrive to My Drive

Source SPWeb	Target GUser
https://example-my.sharepoint.com/personal/user_solarmora_com	user1@solarmora.com

Migrate a OneDrive document library to My Drive

In this example, the destination is a subfolder on the user's My Drive.

Source SPLibrary	Target GUser	MapChildrenOnly	Target Sub Path
https://example-my.sharepoint.com/personal/user_solarmora_com/Projects	user1@solarmora.com	TRUE	Projects

Mapping best practices

- Do not map a single-source user to multiple target users within the same project.
- For heavy users with very large amounts of data in OneDrive, consider using folder mapping instead of user mapping to spread the load into more nodes and improve performance.
- When you're migrating from SharePoint and configuring the connection, use *Document libraries (SPLibrary Class)* as a source instead of choosing specific folders. If you select specific folders, a known issue with migrating data hierarchically from SharePoint can generate errors.
- When you're migrating from OneDrive, use SPWeb class in the CSV header and specify **Target GUser** per each line (follow the instructions in Common examples of mappings).
- When you're transferring files to a shared drive, if the original authors of the files don't have access to the target shared drive, their files are transferred using the target GUser account specified in the mapping. This type of transfer can create bottlenecks and slow a migration. We recommend that you do the following:
    - Make sure that the **Migrate folder permissions** setting is turned on in the settings template, or apply permissions later using tools such as GAM.
    - Add additional target GUser accounts to help distribute load. Use at least one target GUser account for each active node server.

Settings

The next sections focus on specific settings in the settings templates that affect migration performance.

Exchange

Email migrations in most cases are more straightforward than file migrations, but take these monitoring points into account:

- Attachments bigger than 25 MB are not supported by Gmail. Google Workspace Migrate offers an option to upload them to the Google Drive of the relevant user with a link into emails. Creating links in emails instead of attaching files is the best practice in Gmail.
- Since in Exchange the same email can be replicated in multiple folders, while in Gmail it's always counted as one email with multiple labels, the number of email messages in the source account might exceed the number of messages migrated to Gmail.
- The maximum number of labels per single user in Gmail is 10,000. However, when a user has more than 500 labels, Gmail can take longer to load. Share the best practices for working with Gmail, including how it handles labels, with users.
  We recommend that you ask the user to reduce the number of folders before migrating.
- Some labels are reserved in Gmail (such as **Spam** or **Trash**) and cannot have sublabels. If a source subfolder has been mapped to Gmail reserved labels, Google Workspace Migrate creates additional labels (not sublabels) where these messages will be inserted.
- A small number of Exchange mail features are unsupported. For the full list, refer to Supported & unsupported features for Exchange.
- If you encounter any EWS throttling issues, refer to the Source Connection: Exchange section for instructions on how to increase EWS throughput.

Shared mailboxes

You can migrate shared mailboxes as groups or to a licensed user account. Each method has pros and cons that you should consider.

Google Group as a destination

If migrating to a group, you must create the group in the target account first. Groups aren't automatically created by the migration process. You can create and set up groups as needed using GAM.

Benefits

Groups:

- Do not require licenses
- Can be configured as a Collaborative inbox

Limitations

Groups:

- Do not support labels (these can be enabled on the group level, but emails are not labeled during the migration process).
- Can be the destination only for inbox mail from the source. Sent items, drafts, junk, spam, archive, and so on are not migrated.
- Might not feel the same as an inbox to users being migrated, meaning that you may need to help users adapt to the changes. Users need to get used to the new interface (they will access a shared mailbox using groups.google.com) and need to understand that to send messages as a group address they add **sendas** to their account.
- Do not support oversize attachments (greater than 25 MB in size). When an email has an oversize attachment and the setting to migrate it to Drive is enabled, the attachment is skipped because the group doesn't have access to Drive. Google Workspace Migrate currently does not allow mapping to a Drive for these attachments.

Additionally:

- Migrated email messages appear as *read only* by the user who sent them. For other group members, the messages appear as unread. This is a known monitoring point when migrating to a group.
- If, after the migration, you allow users to access their mailboxes from the Outlook client, they won't have access to the group from there and will need to access it from a browser.

Generic user mailbox as destination (licensed user)

Migrations to licensed user accounts allow all the email structure to be maintained (all folders are migrated as labels).

Benefits:

- The account has Drive access, meaning that attachments of oversized files are uploaded.
- Such accounts can be delegated (GAM supports bulk actions), and users are able to access them using a browser.

Limitations:

- A single Google account can support up to 1,000 delegates. To avoid impacting account performance, we recommend allowing a maximum of 40 concurrent users.
- Delegates have access to the mailbox, but access to Calendar or Drive must be shared

separately.
- Creating the licensed Gmail account might mean an additional cost for an organization (depending on the assigned edition).

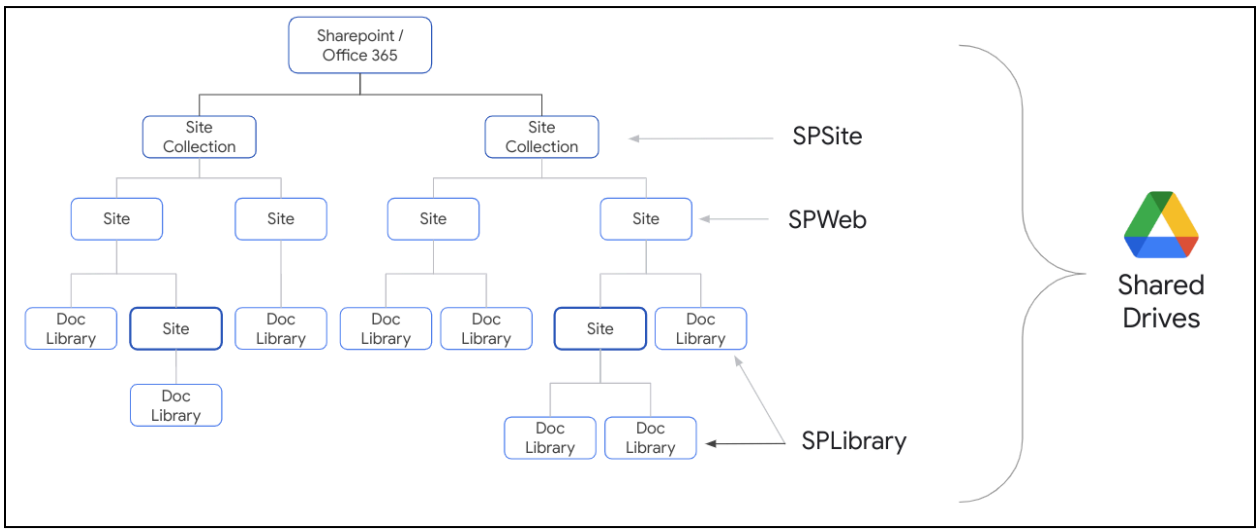
Mail archives

It's a best practice to migrate mail archives only after user migration has completed. Using the Exchange **Archived mail folders/messages** headers, you can apply filters (in the settings templates). You can also disable archives migration in the settings template to exclude this data from the migration.

As a best practice, we recommend that you check for the existence of Microsoft Outlook Data Files (PST files). Google Workspace Migrate does not provide a mechanism for migrating these files. Outlook allows users to create mail archives on their local disk, which are stored in Outlook Data Files (.pst), and you must use Google Workspace Migration for Microsoft Exchange (GWMME), rather than Google Workspace Migrate, to migrate these files. The files are opened for writing when Outlook is active, so moving them to a network drive where GWMME can access them can be difficult.

Note: Outlook also stores data in Offline Outlook Data File (.ost) files, which are local cache files for offline use. Migration from these sources is not possible using any tool that we know.

OneDrive/SharePoint -> Google Drive



High level SharePoint files hierarchy

When planning a migration of OneDrive or SharePoint files, familiarize yourself with Google Drive large migration best practices as well as the following items:

- Be aware of files you can store in Google Drive. Unsupported file types are skipped during migration.
- It is not possible to migrate content to a user's My Drive and to shared drives in a single migration; you must run migrations for these items separately.
- Split the migration into multiple target locations.
- In the settings template for OneDrive, turn off SharePoint related settings. We recommend turning on only settings related to the service being migrated (either OneDrive or SharePoint) to avoid additional API calls, including the following:
    - **Migrate Role Assignments to Google Permissions**–Turn off any items of SharePoint type.
    - **Migrate webs to folders and Site collections to folders**–Both OneDrive and Teams sites are SharePoint Online resources and use the same underlying Microsoft infrastructure. If, after disabling these options, OneDrive migration does not start, turn these on. If you map to Sites URLs, enable the Site collections to folders option.
- Restricted permissions (where a child file or folder has fewer permissions than its parent) are not supported by shared drives. If the file or folder has restricted permissions on the source account, the restricted permissions are not migrated.
- If the **Migrate SharePoint documents with restricted permissions** option is enabled in the settings template, since shared drives do not support restricted permissions, you might see failures in the logs.
- During a migration to a shared drive, use a sharding users list to spread the migration load among many user accounts (up to 10 sharding users per each node). This method allows you to avoid the 750 GB daily upload limit from a single user account, increases migration speed, and reduces the number of API calls by a single user. Although creating sharding users allows you to distribute the number of queries among many users, to avoid hitting quota limits, and to use more concurrent nodes, it does not increase performance directly. Spreading the migration load across a larger group of users like this instead helps you to avoid API bottlenecks and congestion of the Drive service. However, the disadvantage of using this method is that files created in shared drives will have different creators listed in their metadata than those in the source environment. This small user experience (UX) change might be justified by the significant performance increase it supports. We recommend the following:
    - Use one sharding user account per 400,000 files that are being migrated to Drive.
    - Create up to 10 shard accounts per node.
    - Be aware that when you use a sharding users list, files are migrated to the sharding user's drive first, and then the ownership is transferred to the destination user. Because of this, the shard user account is listed as the first modifier of the file. When a shared drive is used as a destination, the metadata of the file identifies the creator as a shard user. Because shard users typically

have system rather than human names, you might need to explain this issue to users

- If you're migrating content to Drive, the target user must have the correct access level before you start the migration.
    - If you're migrating data to a shared drive, the user needs to be a content manager or a manager of the shared drive.
    - If you're migrating data to a folder in a user's My Drive, the user must be an owner of the folder.
- Access to a shared drive in the target location–If the original content authors don't have access to the shared drives on the target account, files are transferred using the target GUser account specified in the mapping. This type of transfer can create bottlenecks and slow a migration. We recommend that you do the following:
    - Make sure that the **Migrate folder permissions setting** is turned on in the settings template.
    - Use the shard technique previously described to help distribute load.
- Leverage shared drives as target locations wherever possible.
- Eliminate or reduce folder hierarchies wherever possible.
- Consider turning off revisions to versions migration (using the **Migrate SharePoint versions to revisions in Drive** setting). This option significantly increases migration duration (by up to 90%) because all revisions are migrated. This is especially significant given that limitations for the numbers of revisions in Microsoft 365 and Google Workspace are different. SharePoint might store 500 revisions of every file by default, while Google Drive can store 200 + 100 temp revisions.
- When Google Workspace Migrate makes an attempt to preserve tabular content in a Google Sheet form, the migration of elements like lists, lists items, or metadata can increase migration duration. Because the result is not always useful to users, we recommend that you do not migrate lists and list items to Sheets.
- If you create a new project (for example, to avoid issues with the database performance), always migrate a new batch of users. Do not remigrate previously migrated data. It will create duplicates.

Limits & quotas

Migration limitations & monitoring points

- Each folder in a user's My Drive has a limit of 500,000 items, and each shared drive is limited to 400,000 items. Because of these limits, do not allow a service account to create more than 400,000 files on behalf of a given user account.
- Make sure to check the limits and types of data you can store in Google Drive.

- The default quota limits for the Drive API are 20,000 calls every 100 seconds, both per user and per project. This limit applies to the sum of read and write calls. The migration of a single file needs to call the API several times (for example, the file is first copied to a temporary location, permissions are applied, then the file is copied to the final destination).
- Avoid nesting more than 20 levels of folders in My Drive or a shared drive, because this can result in issues with the shared drive interface. The interface supports up to 20 levels. Above that, the users might have difficulties with finding the content or navigating through the interface.
- Don't map a single source user to multiple target users within the same project. Doing so can produce unexpected migration outcomes, such as content related to the source user spread across different target users. If you must map a single user to multiple target users, use separate Google Workspace Migrate projects for each mapping.
- Each SharePoint document, document version, file in a document library, list attachment, and list attachment version has a migration size limit of 2 GB. Items more than 2 GB are skipped during a migration.
    - If a document version or list attachment version is skipped, all subsequent versions are also skipped, even if they are smaller than 2 GB. This is a limitation of the Microsoft API used by the current version of Google Workspace Migrate.
    - Because of differences in how the Microsoft and Google platforms calculate the size of files, it's possible for files under the 2 GB limit in the source environment to be actually over the limit. For example, a 2 GB SharePoint file might be shown as a 2.1 GB file in Drive.
    - Because of the limit, it's important to identify oversized files ahead of migration and to use other methods to transfer them, for example, running a Powershell script to export them and then import to Google.
    - SharePoint documents with restricted permissions (where a child file or folder has fewer permissions than its parent) cannot be migrated to a shared drive. If the file or folder has restricted permissions on the source account and it's mapped to a shared drive, the restricted permissions are not migrated. This results in failures visible in the log.
- When you migrate from SharePoint or OneDrive, the connection is handled using a SharePoint AppID that has connection limits and results in throttling or connection closures when more than 2-3 nodes are in use. For more information about this issue, go to Source connection: SharePoint & OneDrive.
- When you migrate from Exchange Online, the connection is handled through EWS, and throttling can also occur. You can adjust EWS limits as described in Source connection: Exchange.
- The Contacts service has a limit of 25,000 contacts per user or 20 MB (depending on which is reached first). This is a hard limit and it is not possible to import a higher number of contacts to a single account. As a workaround, you can consider using a

third-party contact management tool, or importing the list of contacts as a file to a user's Drive.

- Calendar migration involving large amounts of calendar data is also subject to limits:
  - Do not attempt to migrate more than 10,000 events for a given day for a user.
  - Do not migrate more than 100,000 events to a single calendar.
  - For recurring events, if the first instance of a recurring event starts prior to the migration window, instances of it within the window will not migrate.

These articles list known monitoring points and limits:

- Monitoring points for SharePoint
- Supported & unsupported features for SharePoint

API quotas

You should consider default API quotas during all migrations. For large or complex migrations, we recommend planning ahead to both identify Google API quota limits that could be reached and to increase the limits where needed.

Default Google APIs quotas

Gmail API

- Queries per day: 1,000,000,000 per 24 hours (can be increased)
- Queries per 60 seconds per user: 15,000 (cannot be increased)
- Queries per 60 seconds: 1,200,000

Drive API

- Queries per day: 1,000,000,000 per 24 hours
- Queries per 60 seconds per user: 12,000 (can be increased)
- Queries per 60 seconds: 12,000 (can be increased)

Calendar API

- Queries per day: 1,000,000 per 24 hours (can be increased, the formula that helps to calculate this is the average of number of events per user x users x 1.3)
- Queries per 60 seconds per user: 600 (can be increased)
- Queries per 60 seconds: 10,000 (can be increased)

Tasks API

- Queries per day: 50,000 (can be increased)

People API (Contacts)

- Queries per day per user: 125,000 (can be increased)
- Write requests per user per 60 sec: 90 (can be increased)

Groups Migration API

- Queries per day: 2,000,000 (can be increased)
- Queries per minute per user: 600 (can be increased)

IAM API

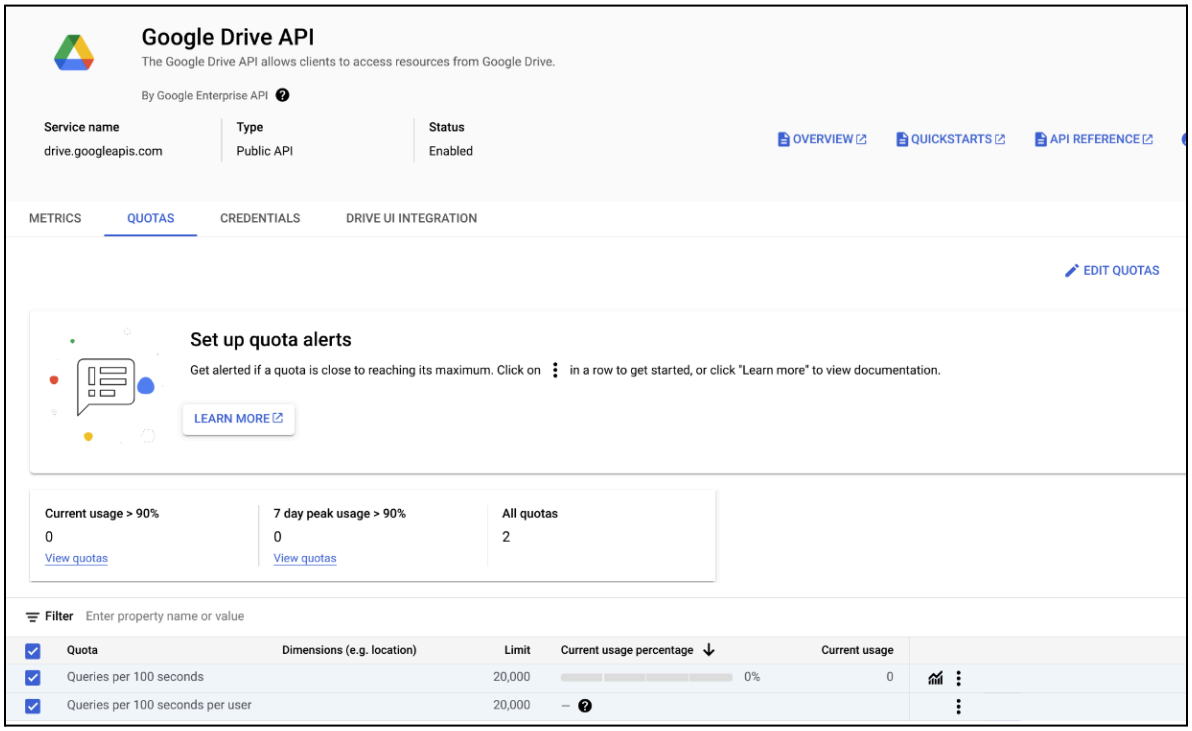
- Read requests (for example, getting a policy): 6,000 per minute
- Write requests (for example, updating a policy): 600 per minute

Service Account Credentials API

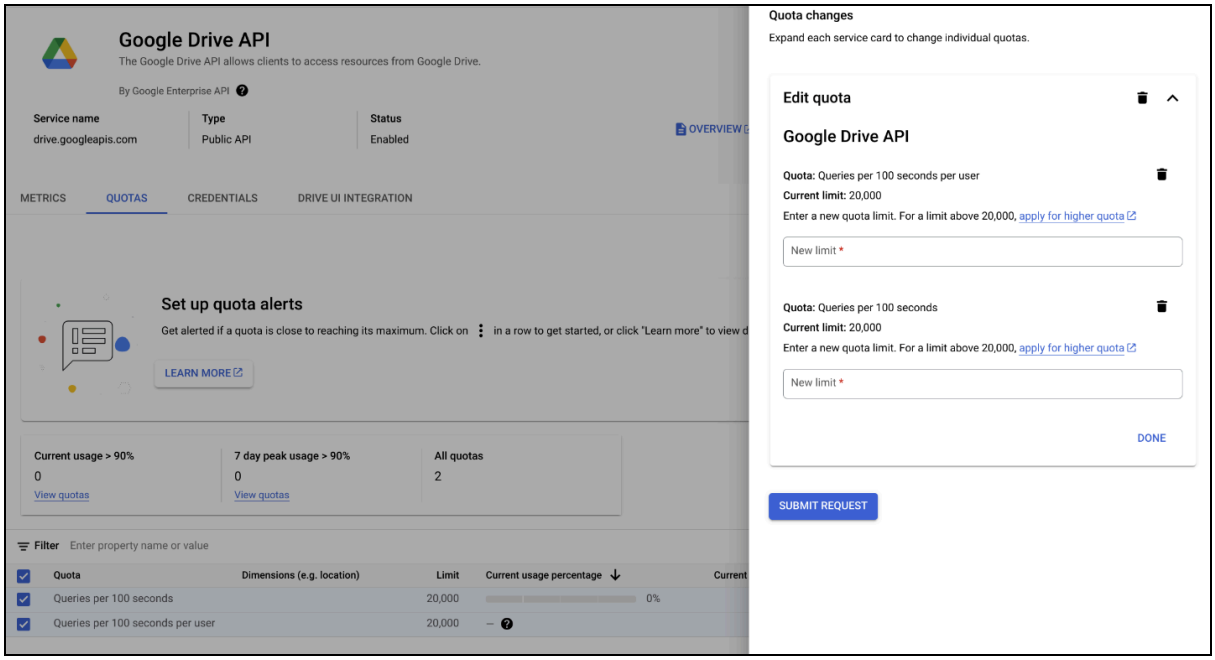
- Requests to generate credentials: 60,000 per minute

Increasing the quota

In the Google Admin console, navigate to the API/Service details (for example, Google Drive API) and select the quota that you want to increase.



Click **Edit quota**.



If you want to increase the quota above a certain limit, you can click **apply for higher quota**. The hyperlink redirects you to the form to fill out.

Another option to increase the API quota is to contact Google Support. Agents are equipped to verify current usage and to increase based on best practices and customer needs.

Monitoring point: Google Workspace Migrate uses some permissions related to APIs to preserve document metadata. During the migration process, files are created in a temporary location, and the SharePoint author is given access to update the content, then they're moved to the correct location. Based on how Google Workspace Migrate works, even if the options to migrate ACL and the permissions are disabled, you might notice Drive permissions related to API calls in the activity logs (and API usage under the Google Cloud project).

Reporting & reconciliation

Google Workspace Migrate provides a number of ways both for reporting on migration status and for logging progress and system events.

These reporting options are critical in managing migrations, because they let you verify that items have been transferred correctly. For example, it's good practice to compare migrated content to the results of the scan executed before the migration. The scan clearly shows the number of objects of different types, which can then be reconciled to the number migrated.

It's also possible to use other methods for reconciliation. For example, because scanning OneDrive is time-consuming, it can be quicker to download a list of resources, including the size and amount of data to be migrated, from SharePoint using a PowerShell script. However, because the Microsoft and Google platforms calculate item numbers differently (for example, Google Workspace Migrate counts ACLs, IMAP configs, and user signatures as items) item counts can differ from those produced by the PowerShell script after a migration.

Scans & reports

Scan results

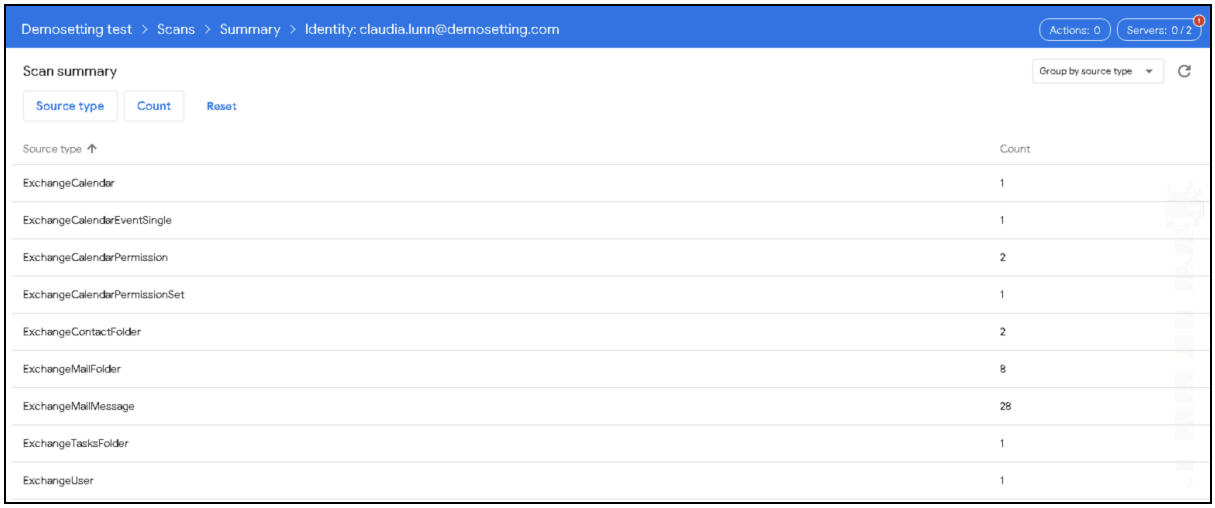
Scan results help you to understand the data that is going to be migrated, both identifying possible issues, such as unsupported document types, and ensuring that connectivity and authorization is functional. Scans can take some time to complete, since they need to crawl all the items in a source to optimize the migration by distributing load across partitions. Doing so allows Google Workspace Migrate to apply optimizations based on the scan results (for example, by increasing the weight of particular partitions).

Scan summary

Scan summaries provide information about the counts per user, and these can be grouped by identity, source type, or error code.

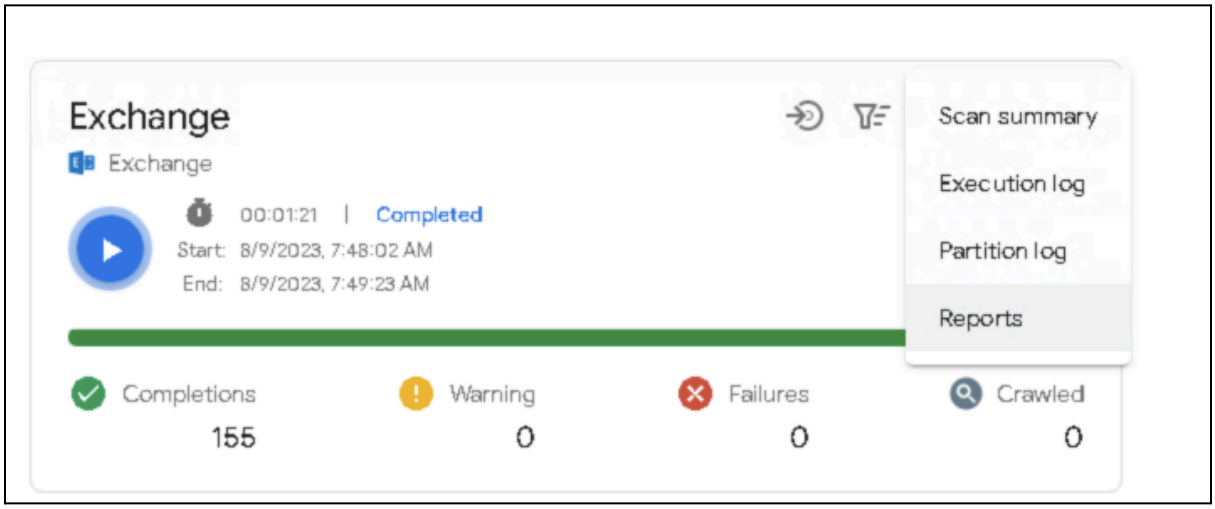
Single user summary

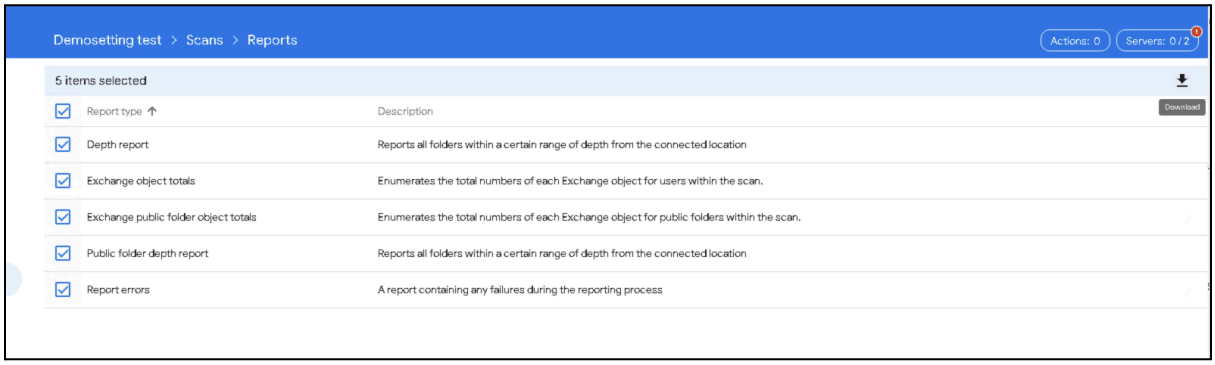
This summary provides information about the data types that can be imported to Google Workspace (such as numbers of files) for a single user.



Scan report export

You can download all scan reports for further analysis (for example, in Sheets).





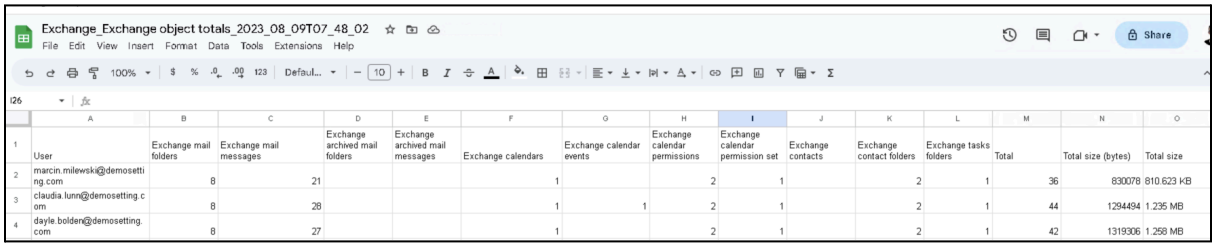
Depth report

This report details all folders within a certain range of depth from the connected location. You can use this report to understand if it's possible to either put these into Drive or split them into

multiple locations because of the 20 subfolder Drive limit. The report is also useful for understanding whether you should use individual mappings or folder mappings for file migration.

Exchange object totals

This report provides information about the number and the size of items to be migrated from Exchange.



Exchange_Exchange object totals_2023_08_09T07_48_02

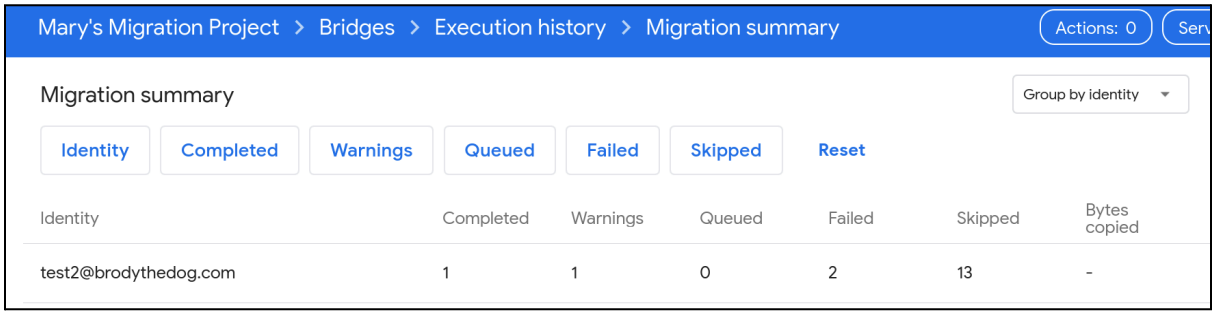
	User	Exchange mail folders	Exchange mail messages	Exchange archived mail folders	Exchange archived mail messages	Exchange calendars	Exchange calendar events	Exchange calendar permissions	Exchange calendar permission set	Exchange contacts	Exchange contact folders	Exchange tasks folders	Total	Total size (bytes)	Total size
2	marcin.milewski@demosetting.com	8	21			1		2	1		2	1	36	830078	810.623 kB
3	claudia.lunn@demosetting.com	8	28			1	1	2	1		2	1	44	1294494	1.235 MB
4	dayle.bolden@demosetting.com	8	27			1		2	1		2	1	42	1319306	1.258 MB

Shared drive candidate report

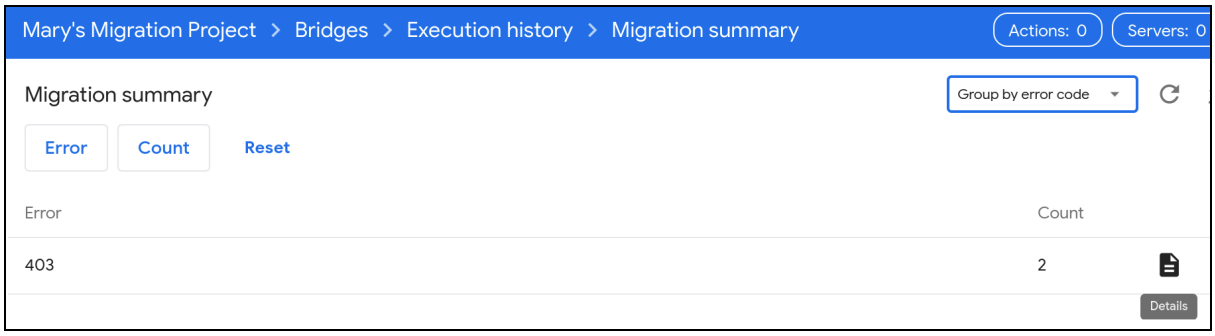
This report assesses the suitability of folders for migration to shared drives.

Migration summary & logs

Scans and reports are useful when you're planning a migration. The migration summary overview is useful when testing or managing it. This report provides an overview of the migration status for a given bridge, with counts for completed, failed, and skipped items grouped either by identity by default, or by whatever is specified in the mapping used for the migration run.



In addition to grouping by identity, you can choose to group by error code to drill into failures:

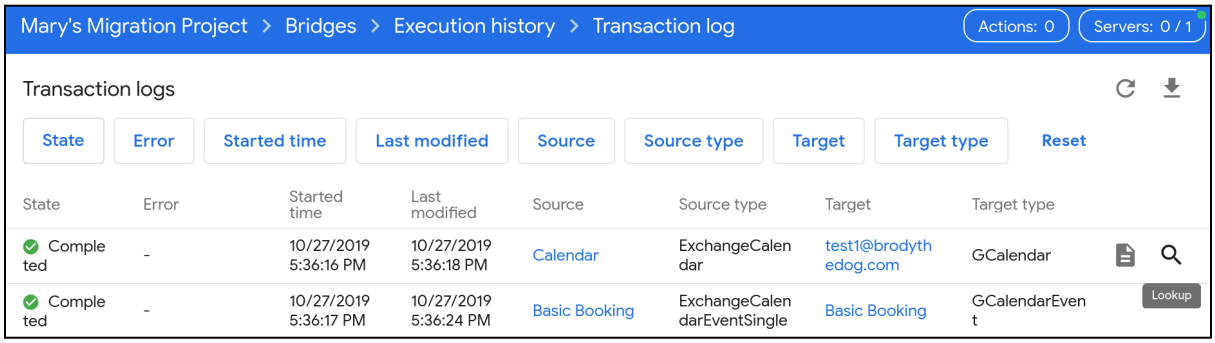


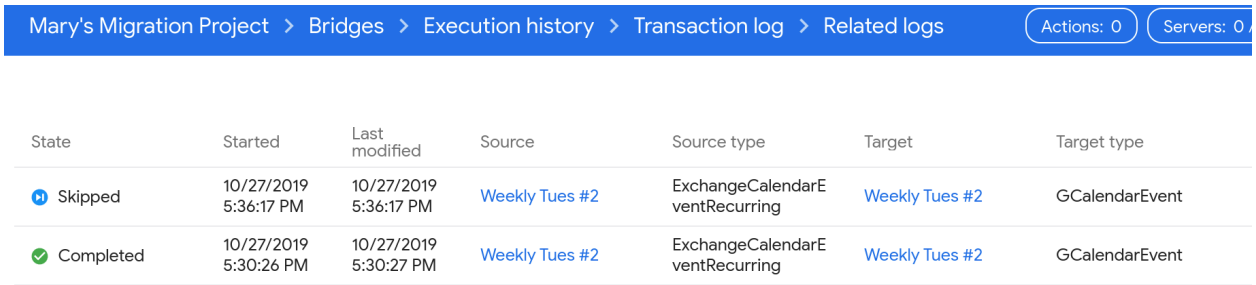
Partition log

This log provides a summary of the items migrated, their start and end times, and the processing node for each partition (roughly equivalent to one entry in a mapping file). If you're seeing crashes or other issues, the time stamp and node file tells you where and when to look in the node server appbridgeservicehost.log files for more information. It's also possible to drill down into a per-partition execution or transaction log, which is smaller and therefore easier to access, since it is limited to a single partition.

Transaction log

The transaction log is the most comprehensive log, providing detailed transaction-level information on every item migrated. Transaction logs also provide the ability to look at an object across multiple migration runs, using the lookup button. After running a bridge, you can filter the transaction log based on the type of data.





For example, we can see in the preceding image that this item was skipped on the most recent run, because it was migrated successfully on the initial run.

Appbridgeservicehost.log

These logs are useful for troubleshooting unexpected issues. The platform server and each of the node servers keep their own log files, written locally to disk. To find the logs:

- On a platform installation: C:\Program Files\Google Workspace Migrate\Google Workspace Migrate Platform\AppBridgeServiceHost.log
- On a node installation: C:\Program Files\Google Workspace Migrate\Google Workspace Migrate Platform Node\AppBridgeServiceHost.log

Google Workspace Activity logs

Admin.google.com stores transaction logs for all activity in services like Drive, Calendar, and so on. The logs don't show failed API responses, but they can be a useful resource for troubleshooting what the migration tool has done.

Troubleshooting

In migrations where a very large amount of data is migrated and a high number of writes to the database is expected, it can take significant time to download reports. A report might also fail to be exported because the database is processing. In such situations, we recommend filtering results before the export (for example, by the completion status).

To avoid issues like this, it's good practice both to limit a single Google Workspace Migrate project to about 10,000 users and to create multiple projects if the workload is bigger. With larger numbers of users or items, the MySQL database size increases, resulting in slower database performance because the size of the tables is larger. Here, using performance optimized CPUs and disks (as described in Configure disks to meet performance requirements) might be necessary.

Estimation & resourcing

Estimating the cost and time needed to complete a migration is of critical importance not just for managing delivery team resources, but also for meeting deployment project milestones.

Cost

Maintaining migration infrastructure comes with costs. We therefore recommend keeping VMs turned on only when migrating or when you need to access resources for troubleshooting.

When you turn off the VM, you pay only for storage space, but note that migration processes must be fully complete before you turn off VMs.

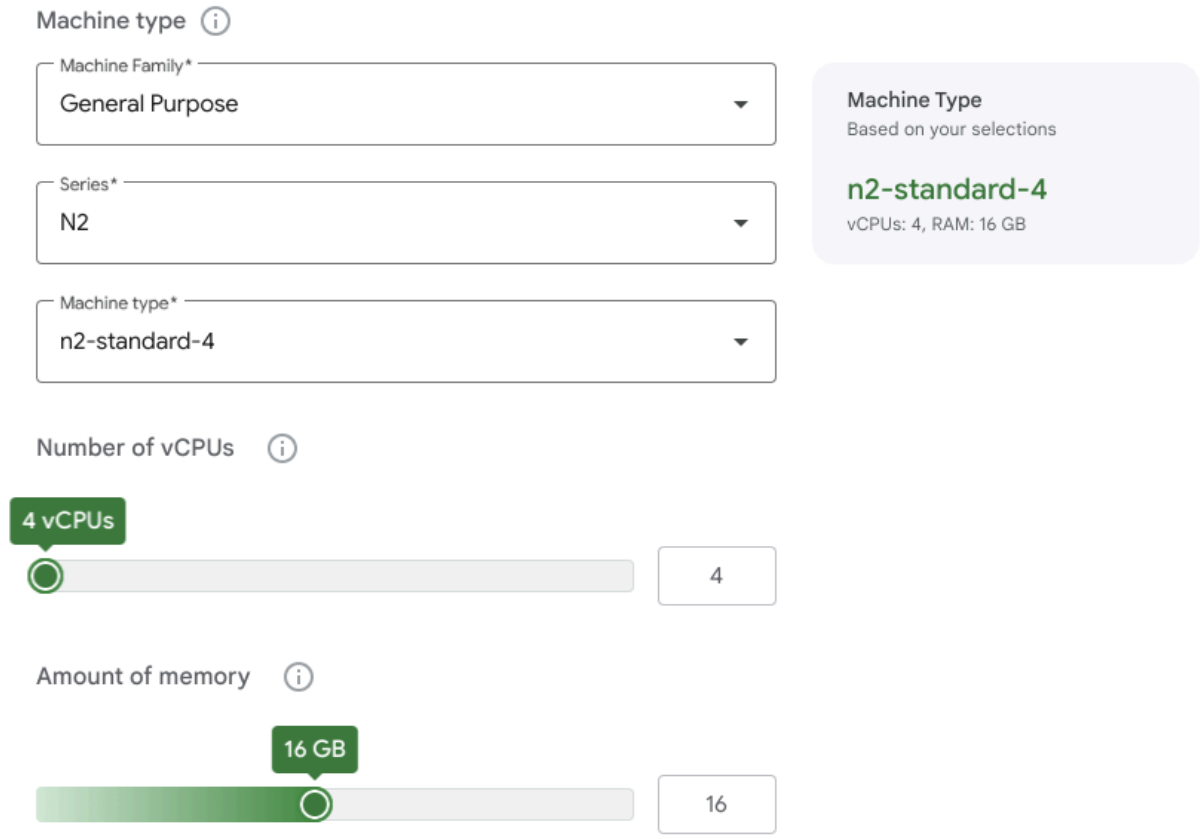
To better understand the costs for the migration infrastructure, refer to Compute Engine pricing. You can also use the Google Cloud's Pricing Calculator to calculate the overall cost of the migration infrastructure over a given time.

The elements of cost are:

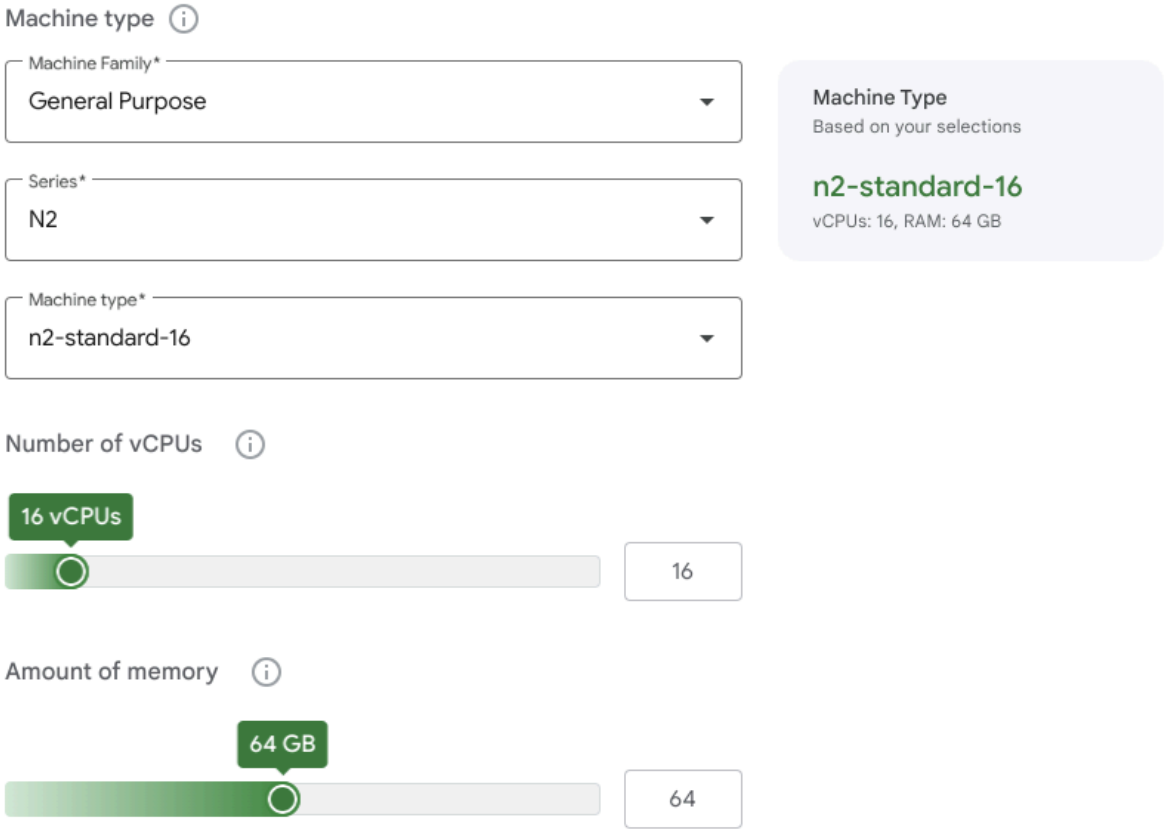
- Platform–Compute Engine instance (4 core, 16 GB RAM, 200 GB SSD)
- 2 x Compute Engine Instance (16 cores, 64 GB of RAM, and SSD drive)
    - MySQL—We recommend approximately 1 TB per 100 million objects that you want to scan and migrate.
    - CouchDB—We recommend approximately 1 TB per 40 million objects that you want to scan and migrate.
- Up to 40 node servers per cluster (4 cores, 32 GB of RAM, and 200 GB SSD each)
- Network ingress cost
- Geographic location of the servers (cost depends on the location)
- Windows service license costs for platform, node, and database servers

Example: The cost of 1 migration platform (cluster) that includes one platform server, one CouchDB database, one MySQL database, and 40 nodes (calculated using the Google Cloud's Pricing Calculator based on the recommended system requirements, located in the EU (Frankfurt), without Google Cloud sustained discount taken into account) would be as follows:

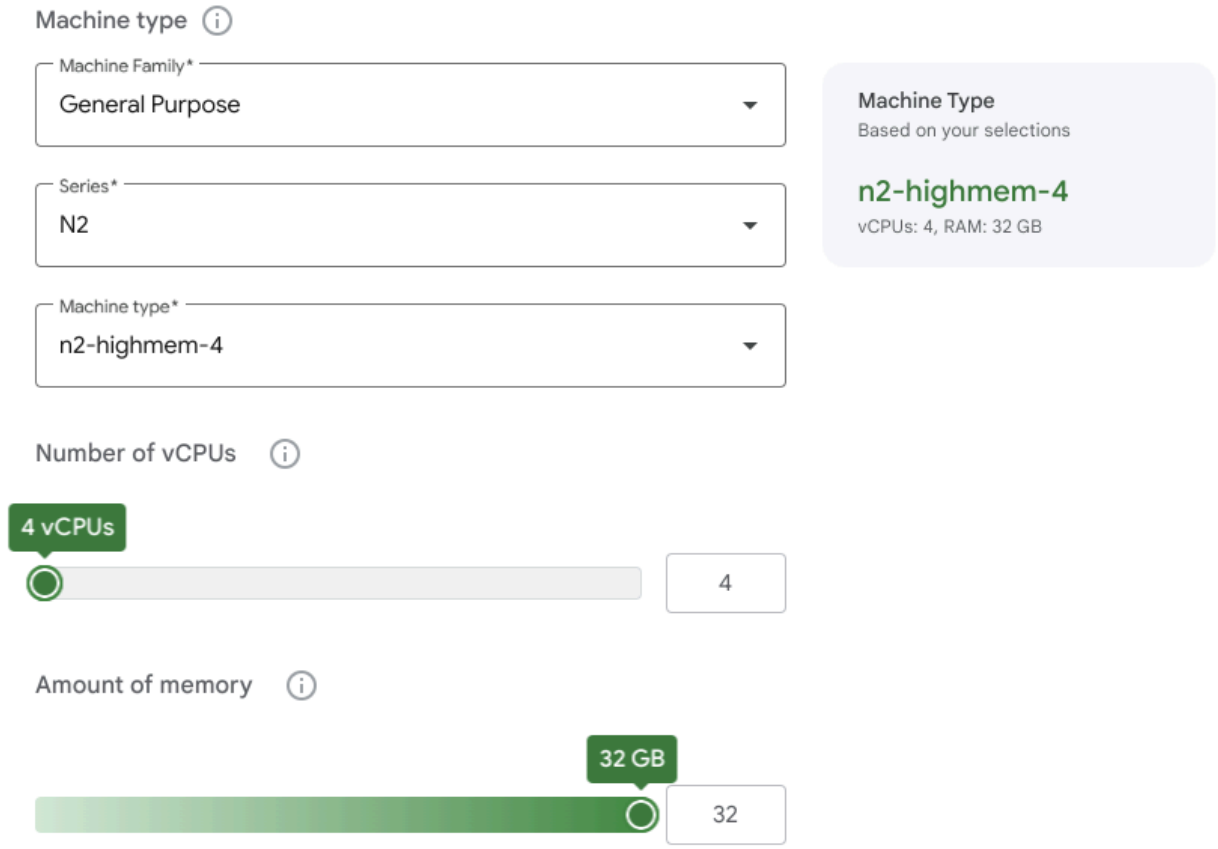
Platform (one platform per cluster of 40 nodes)



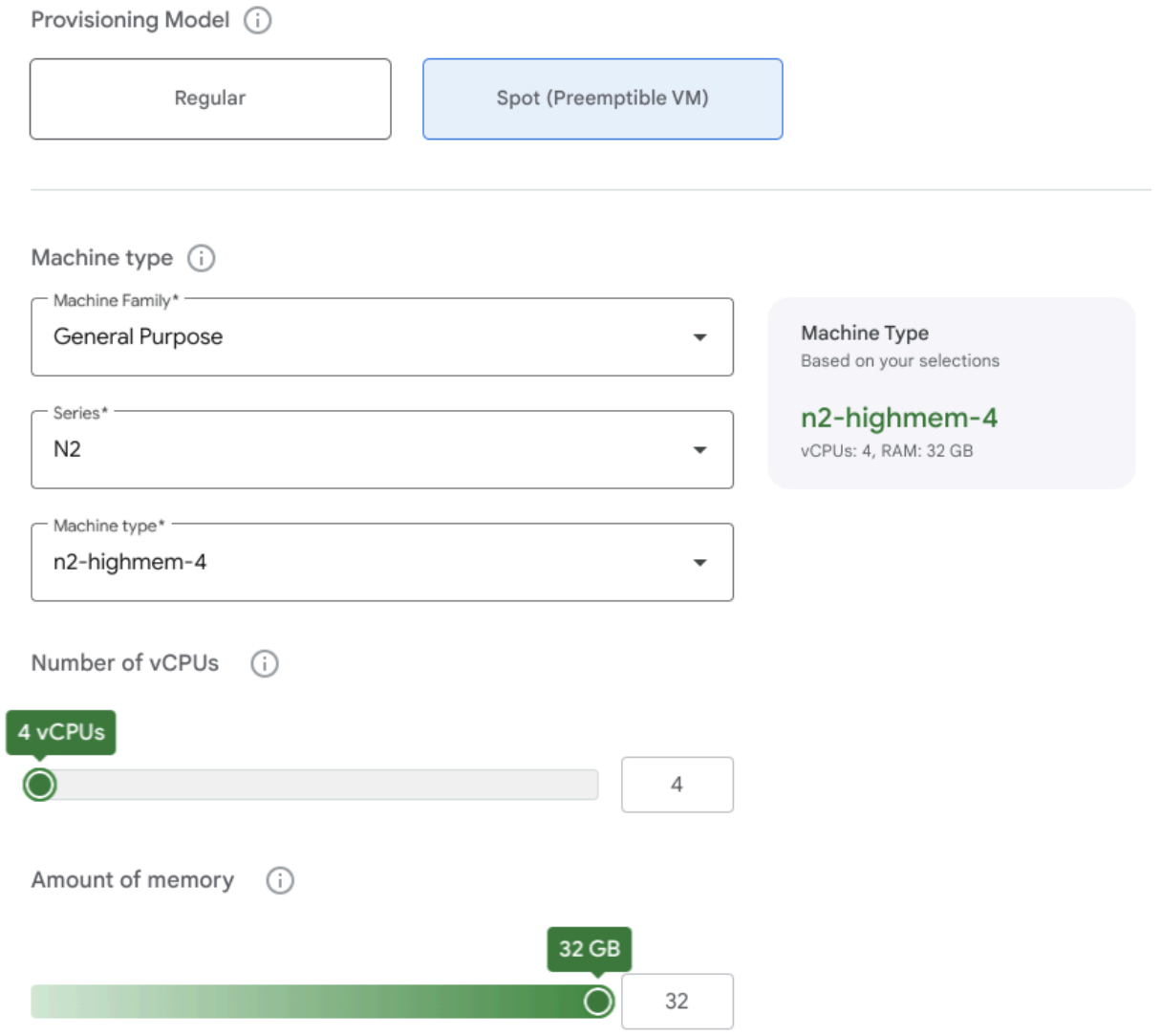
Databases: 2 x (16 cores, 64 GB RAM, 1TB SSD)



Node servers: up to 40



Preemptible node servers: up to 40 x (4 core, 32 GB, 200 GB SSD)



Total cost/month with standard nodes: $20164

Total cost/month with preemptible (spot) nodes: $13622

Time to complete the migration

The time taken to complete a migration depends on many factors including API limits, quota limits, settings applied, and potential issues encountered. Although scans allow you to estimate migration times with a high degree of confidence, we strongly recommend building contingency time into migration estimates to allow for unforeseen or unplanned events. The formulas provided in this section assume that no technical difficulties, such as database, connectivity issues, or API quota, occur during the migration process.

Exchange

Calculate the estimates for Exchange based on the number of items (mail, permissions, and files) and not on the amounts in GB or TB. You can obtain the item count using a migration scan, or a survey of the source environment.

As a rule of thumb, you can calculate Exchange mail estimations based on the following:

(Average mail items x users / clusters x nodes x 6 (partitions) + the maximum item count per user (the user with the maximum number of email items))/ 60 x 60 x 1

The result of this formula is the estimated number of hours to complete the migration.

Example: 15,000 users, 2 clusters (40 nodes each), the average item count per user is 10,000 and the maximum number of users is 80,000 = ~4.5 days. Divide the result of the calculation by 24 to estimate the number of days.

OneDrive & SharePoint

File migrations can be more time-consuming than mail migrations because the relevant APIs have different limits and the files tend to be accessed and modified frequently (versus messages that are sent once and then only read afterwards).

The time required to complete a OneDrive or SharePoint migration depends on many factors including whether you are migrating versions and using sharding users. Because revisions might be made to source files after an initial scan, the actual migration time can be difficult to estimate precisely.

The rule of thumb to calculate the time needed to perform a OneDrive or SharePoint migration to Drive or shared drives (when the **Migrate revisions** option is disabled) is the following:

Average item count per user x number of users / number of clusters x number of nodes x 6 (partitions) + max number of items (user with the highest number of files) / 60 x 60 x 0.5

When you are migrating revisions, the equivalent formula is the following:

50 (an assumption that there are about 50 revisions of a file on average) x average item count per user x number of users / number of clusters x number of nodes x 6 (partitions) + max number of items / 60 x 60 x 0.5

The result of the formula is the estimated amount of hours needed to complete the given migration.

Examples:

- OneDrive -> Drive migration for 15,000 users, 2 clusters (40 nodes each), with an average number of items per user of 1,000 = 2 days
- With the **Migrate revisions** option turned on, the equivalent number can be as large as ~94 days. Remember that a single file may have up to 500 versions on the source side and separate API calls will be needed for each version.

References

Below you can find the list of additional helpful resources

- Supported & unsupported features for Exchange
- About Exchange settings template options
- Monitoring points for Exchange
- Monitoring points for SharePoint
- Monitoring points for file shares
- Monitoring points for Box
- Supported & unsupported features (Google to Google migrations)
- Monitoring points for Google to Google migrations
- Google Workspace Migrate best practices
- Google Workspace Migrate security overview