Proventeq Content Productivity Suite

User Guide

Version: 9.1

April 2025

Authored by: Proventeq

Status: Published

Proventeq Ltd, Reading Enterprise Centre, Earley Gate, Whiteknights Road, Reading RG6 6BU, United Kingdom

1. GETTING STARTED

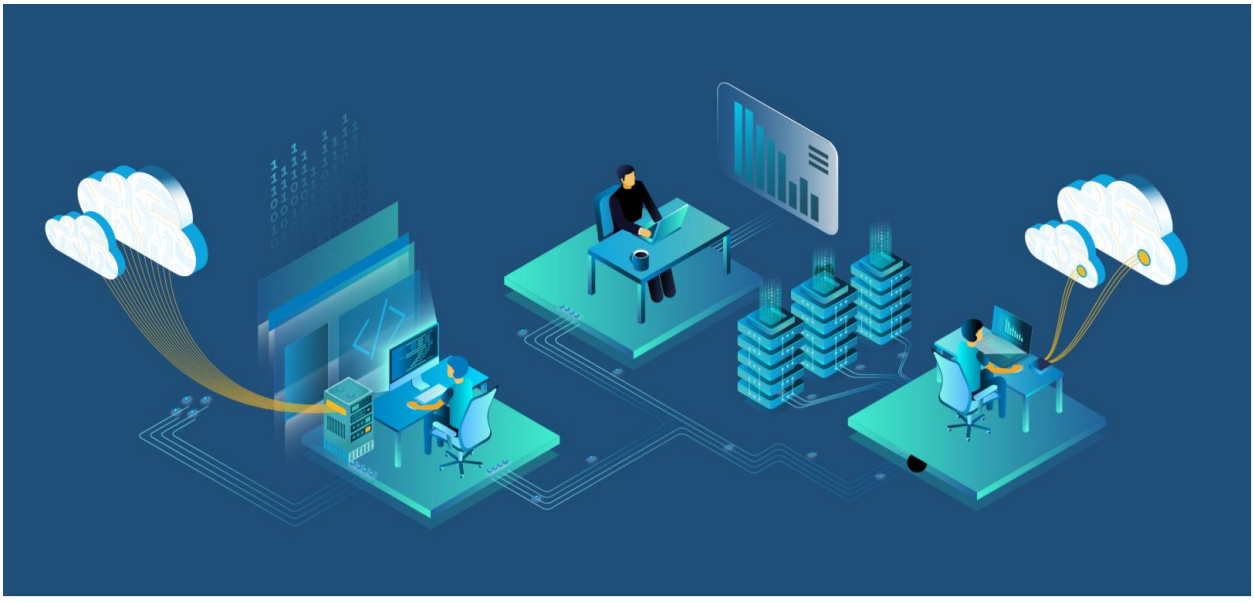
Content Productivity Suite (CPS) is a versatile enterprise grade content management engine that is deployed as a Windows application. It allows you to configure the different source and target system connections, perform content analysis, set up advanced content transformation rules and execute and monitor the migration.

This release supports source systems like Citrix ShareFile, Windows File System, Documentum (SQL and Oracle), iManage, BOX, FileNet P8, Alfresco, OpenText CS (SQL and Oracle), OpenText EDocs(SQL),M-Files, Oracle UCM(SQL, Oracle, DB2). Valid targets are SharePoint Online, File System, M-Files and BOX. Note: Not all sources support migrations to File System, M-Files or BOX. Contact Technical Support for details.

CPS has several separately licensed components which may be available depending upon the product license.

- Migration Accelerator – Content Management Migration also known as PMA
- Insights – Source system analysis and reporting
- Orchestrator – Automate the migration
- Monitor – SharePoint security reports

CPS has been designed with flexibility in mind, so it can be configured and/or extended to address any kind of content analysis, orchestration, monitoring and migration requirement. It provides an interactive user interface to configure, migrate, verify, and enhance the quality of migrated content.



Start the product from the desktop icon or by selecting Start and navigating to Proventeq Content Suite. When the product starts without a license, all features will be disabled, and it will display the license activation screen.

Produce License Activation

To activate a licence online the server must be able to successfully connect to https://api.proventeq.com. If the connection is successful the browser will display “Hello”, if this fails you may need to change server/firewall configure or use Offline Activation.

1.1. 1.1.1. Online Activation

The License key will have been provided to you along with the software download link. Enter the Licence key and click **Activate Online**. This will authenticate the License key and register the product.

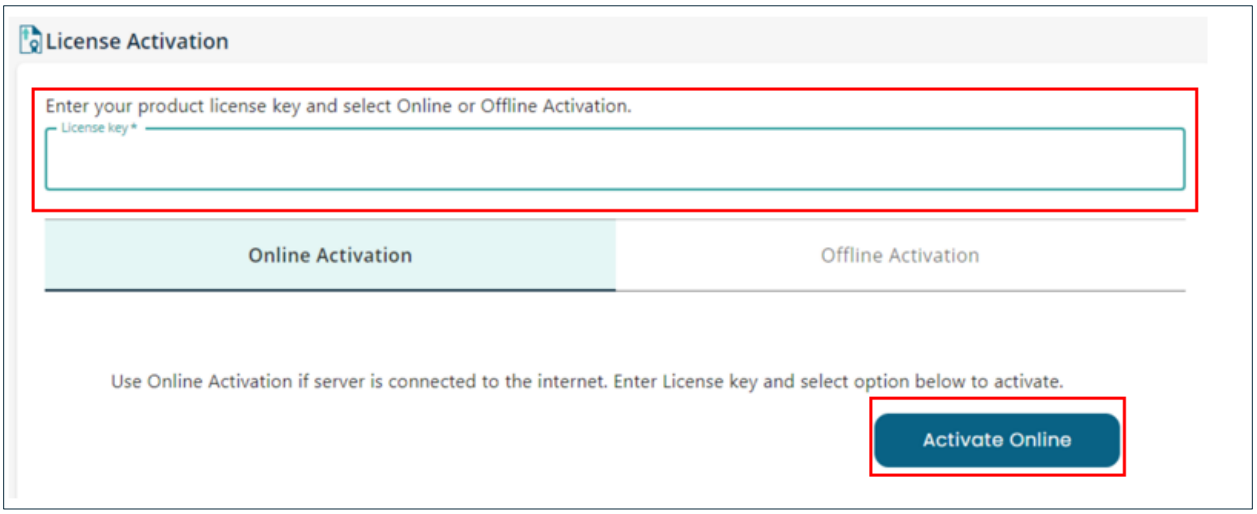


FIGURE 1 - ONLINE LICENSE ACTIVATION

1.1.2. Offline Activation

If it’s not possible to activate the licence via the internet it is possible to Activate Offline. Enter the License key provided and select **Offline Activation**. Click **Generate Activation Request file.** It will generate a file .REQ file that should be sent to license@proventeq.com **.** On receiving the Activation response file(.RESP) from Proventeq, click **Upload Activation Response** to activate the product.

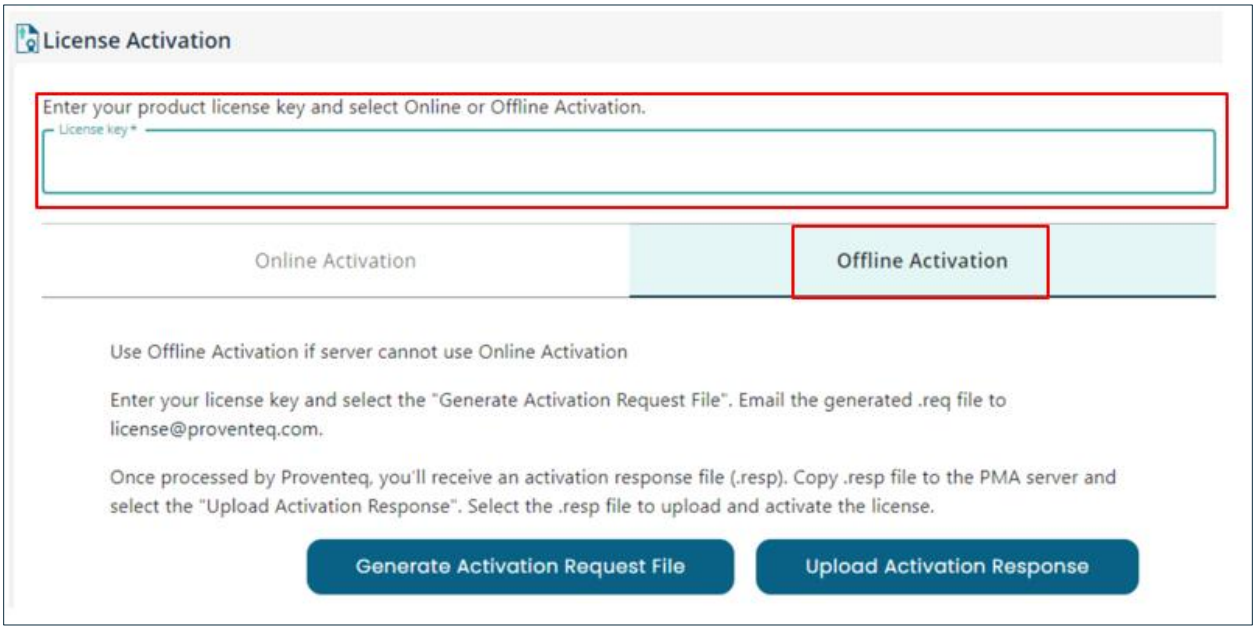


FIGURE 2 - OFFLINE LICENSE ACTIVATION

Once licence is activated, select the Home button to goto the CPS Content Suite Home page.

1.2. Log into Proventeq Support system (Zendesk)

Proventeq uses the popular support system Zendesk for storing knowledge base, documentation as well as ticket tracking. The CPS client contains helpful Learn more… links which link to relevant KB articles and errors messages that may occur in migration may also refer to KB articles. To view these articles, you are required to be logged into Zendesk.

We strongly recommend that you log into your Zendesk account from the CPS system in order to make the most of these features. If you do not have an account, please email pcs.support@proventeq.com. If you are delivering the project on behalf of a customer, include the customer's name in the request.

NOTE: When creating a support ticket, we strongly recommend that you attach screenshots, CPS log files along with the steps to reproduce the issue as this will help us to resolve issues more quickly.

2. UI OVERVIEW

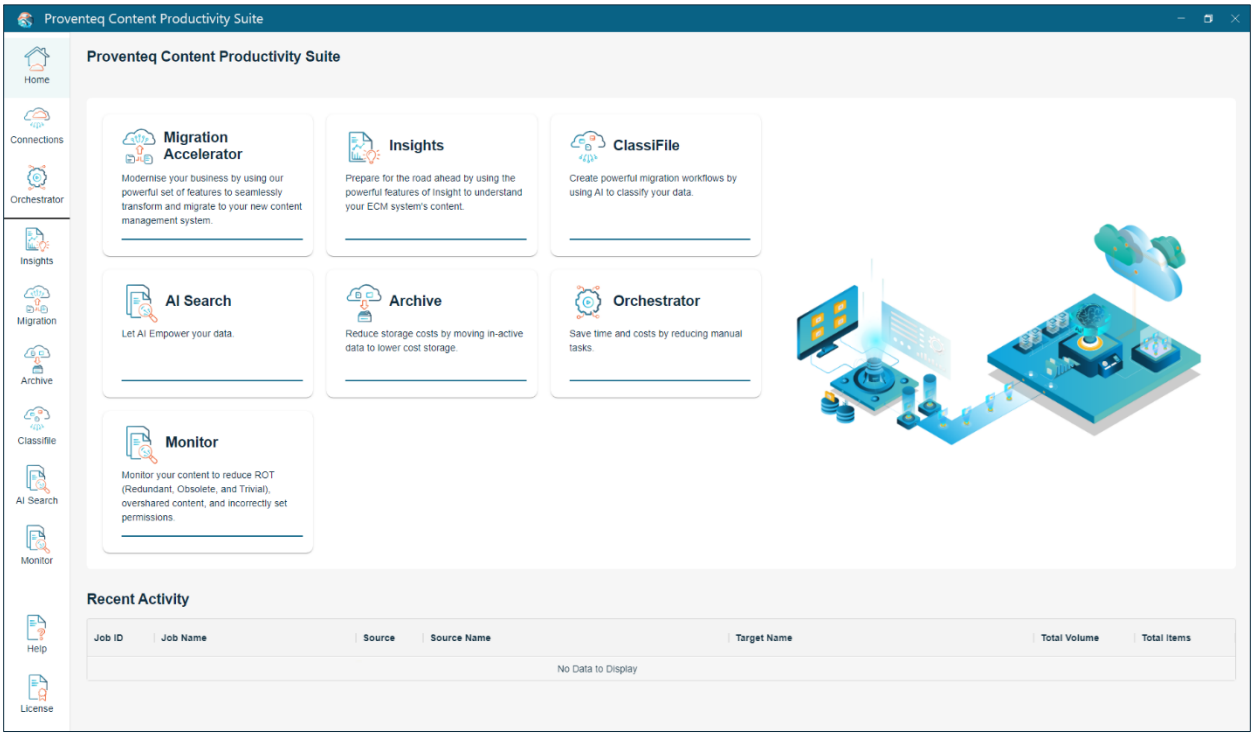


FIGURE 3 - PROVENTEQ CONTENT SUITE HOME SCREEN

Throughout the product there are *Learn more* links which can assist you by opening the relevant section of this online User Guide.

Note: The CPS desktop client is a single user only client. This means that only one user can operate the client on a particular server, irrespective of what user account them may be logged into on that system.

2.1. Navigation Panel

On the left-hand side is the navigation panel. These options are used to carry out tasks that are not related to a specific job. It consists of: -

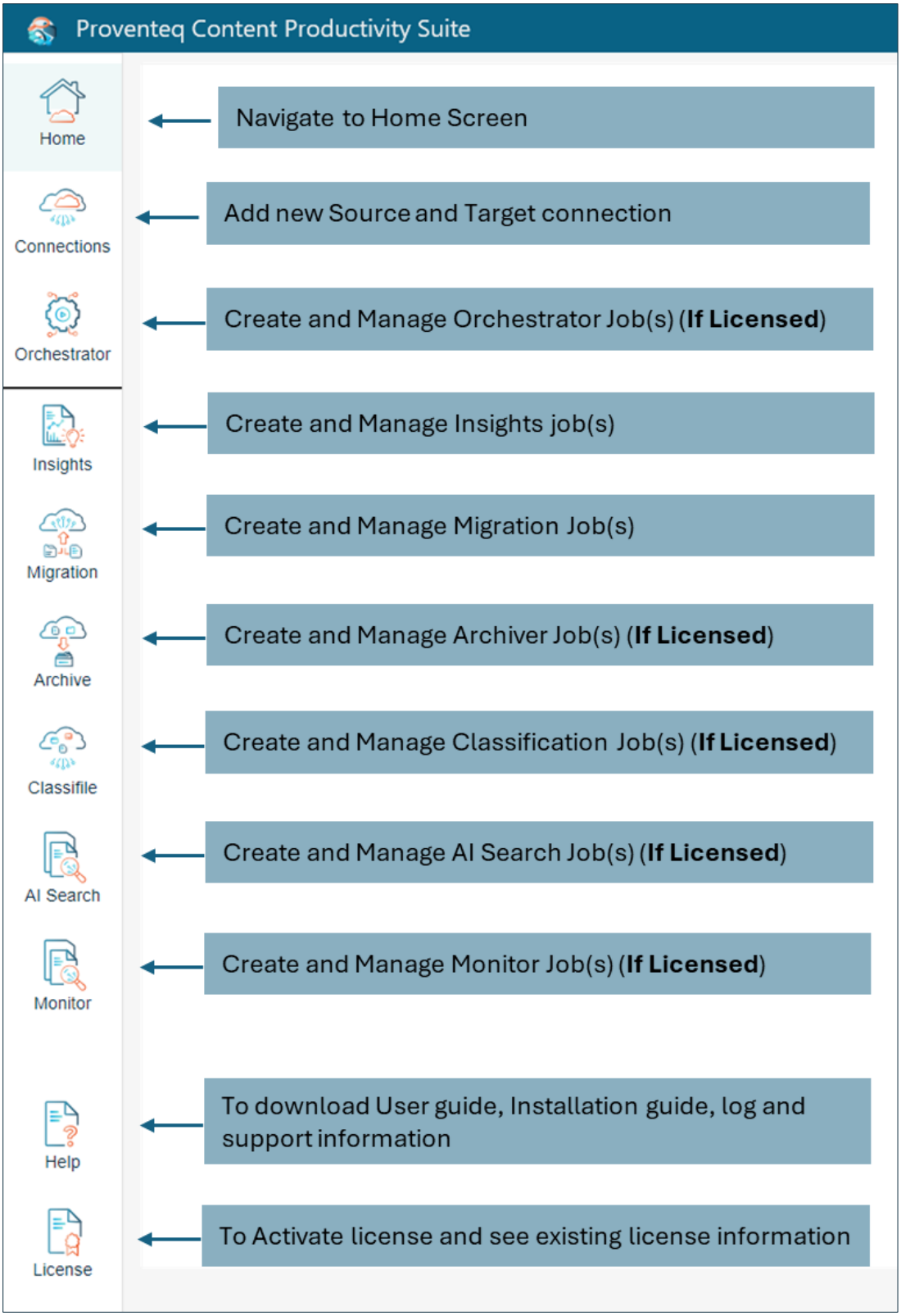


FIGURE 4 - PROVENTEQ CONTENT SUITE NAVIGATION

2.1.1. License Information

License activation (online or off-line) as well as license information and consumption can be found here.

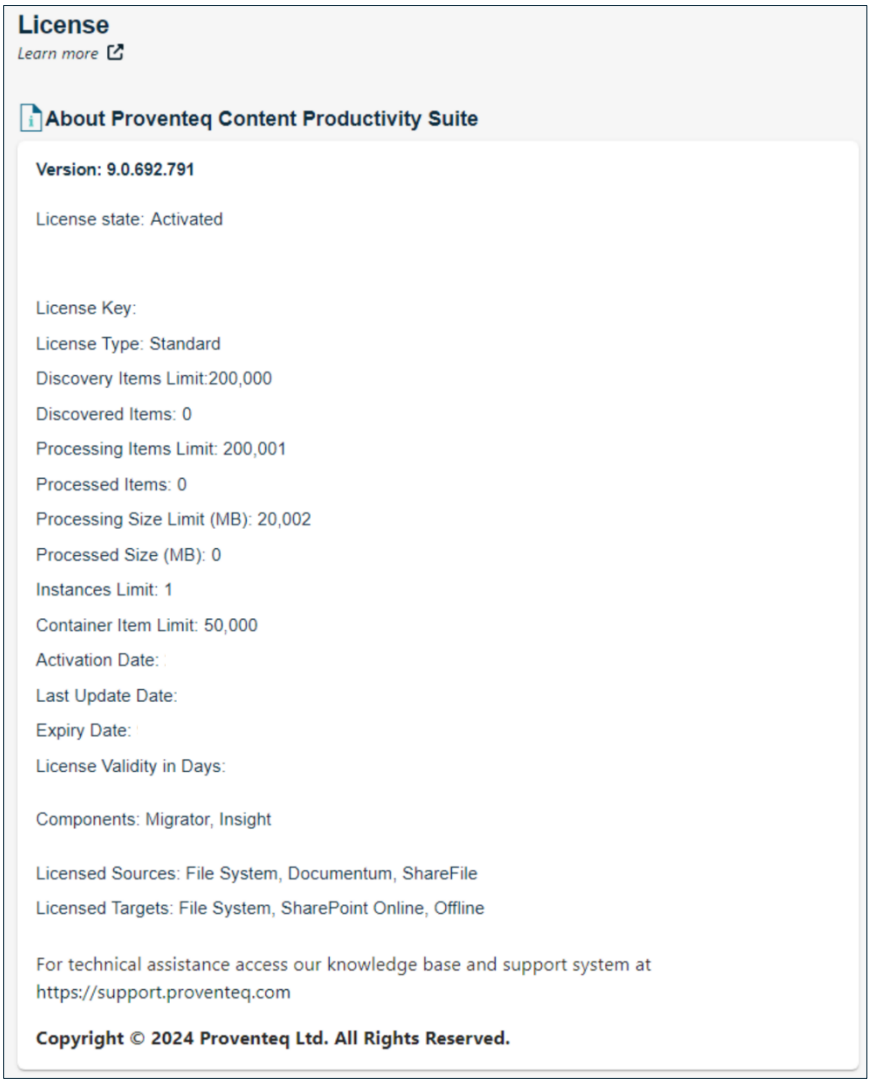


FIGURE 5 - LICENCE INFORMATION

3. CONNECTIONS

Source and Target Connections can be created from the Connections option displayed in the Navigation Panel.

3.1 Managing Connections

A connection is a location that can be migrated from, which is called a 'Source Connection' or a location to be migrated to which is called a 'Target Connection'. They will contain settings related to the relevant system and these connections can be reused across migration jobs.

These connections can either be added to CPS during job creation wizard (as shown earlier) or separately in Manage Connections.

To create a connection using the Manage Connections option, click on the icon in the navigation panel from where you can create or edit Source and Target connections.

Click on to add a new connection or to edit existing connection.

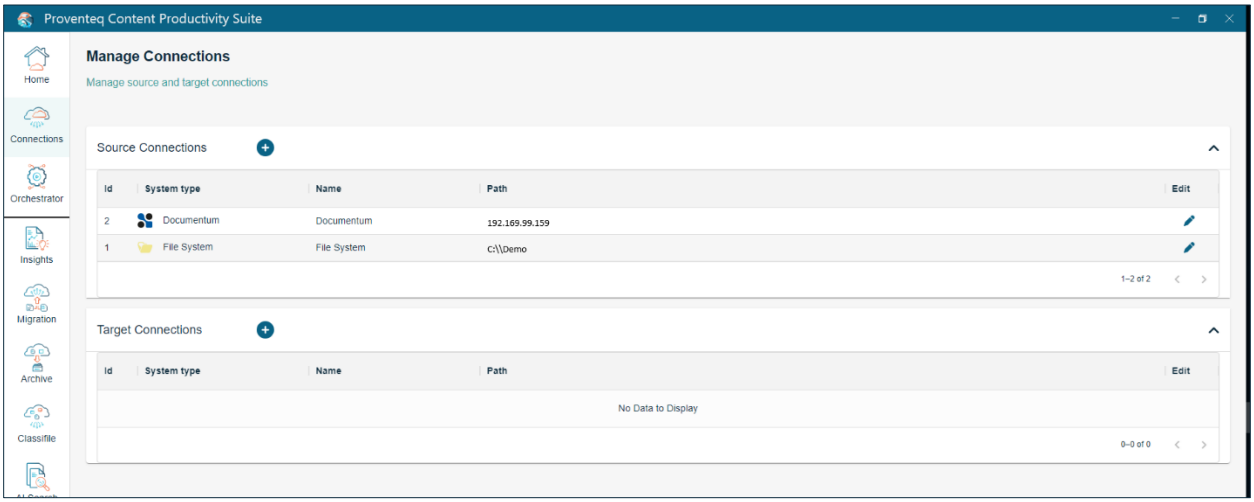


FIGURE 6 - ADD/EDIT CONNECTIONS

3.2 Add a New Source Connection

Click on Plus icon display next to Source Connection header to add new source connection into the system.

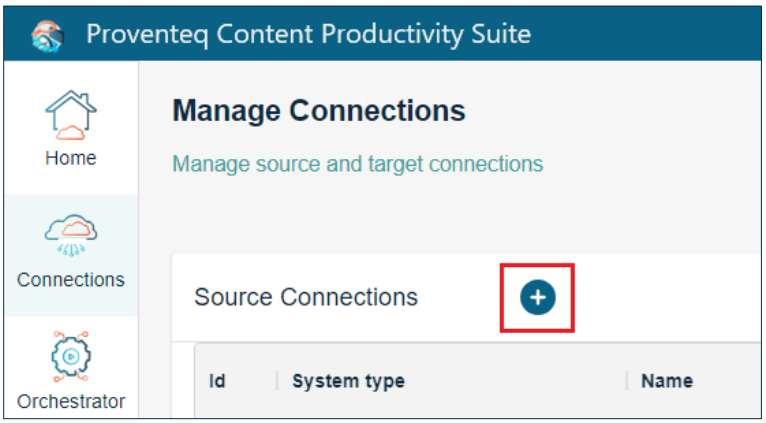


FIGURE 7 - ADD NEW SOURCE CONNECTION

When click on icon, it will open following screen to Select the source platform to add connection from

NOTE - Only platforms that are licensed for migration will be listed in Source System dropdown.

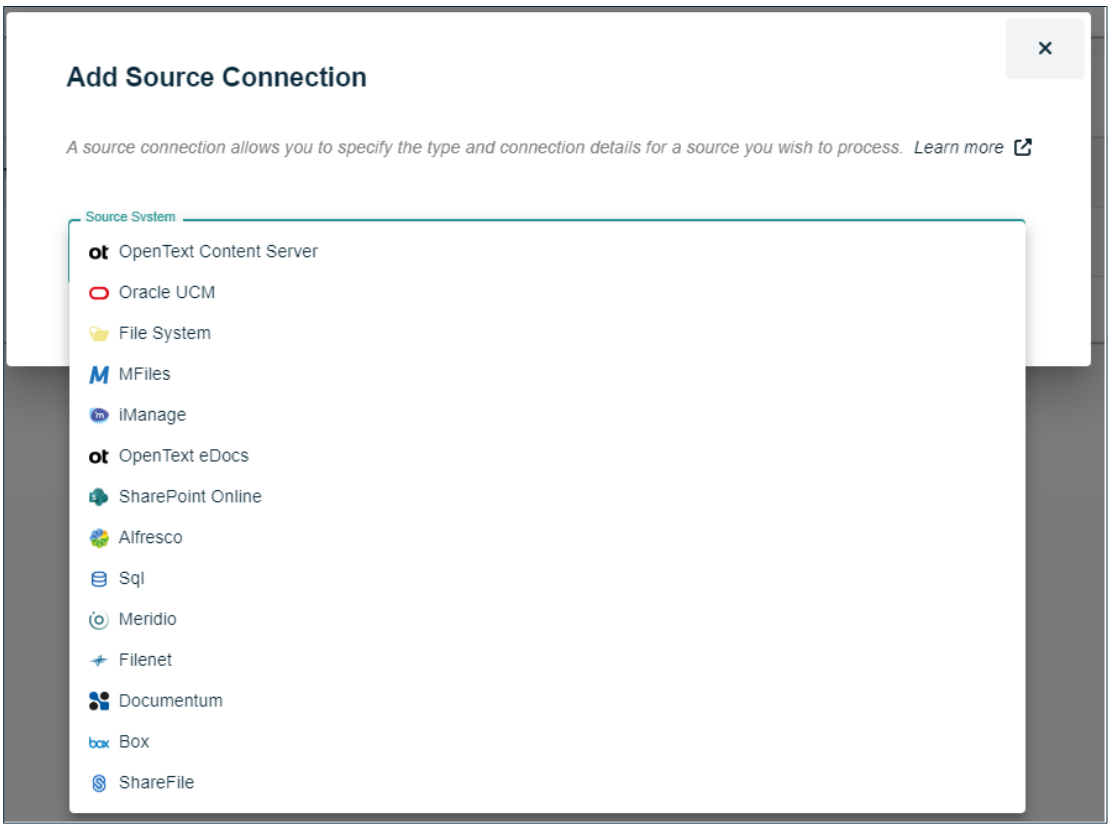


FIGURE 8 - ADD SOURCE CONNECTION SCREEN

3.2.1. Adding a File Share as a Source

Once select File System option in Source System dropdown, it will show following fields to add details related to it

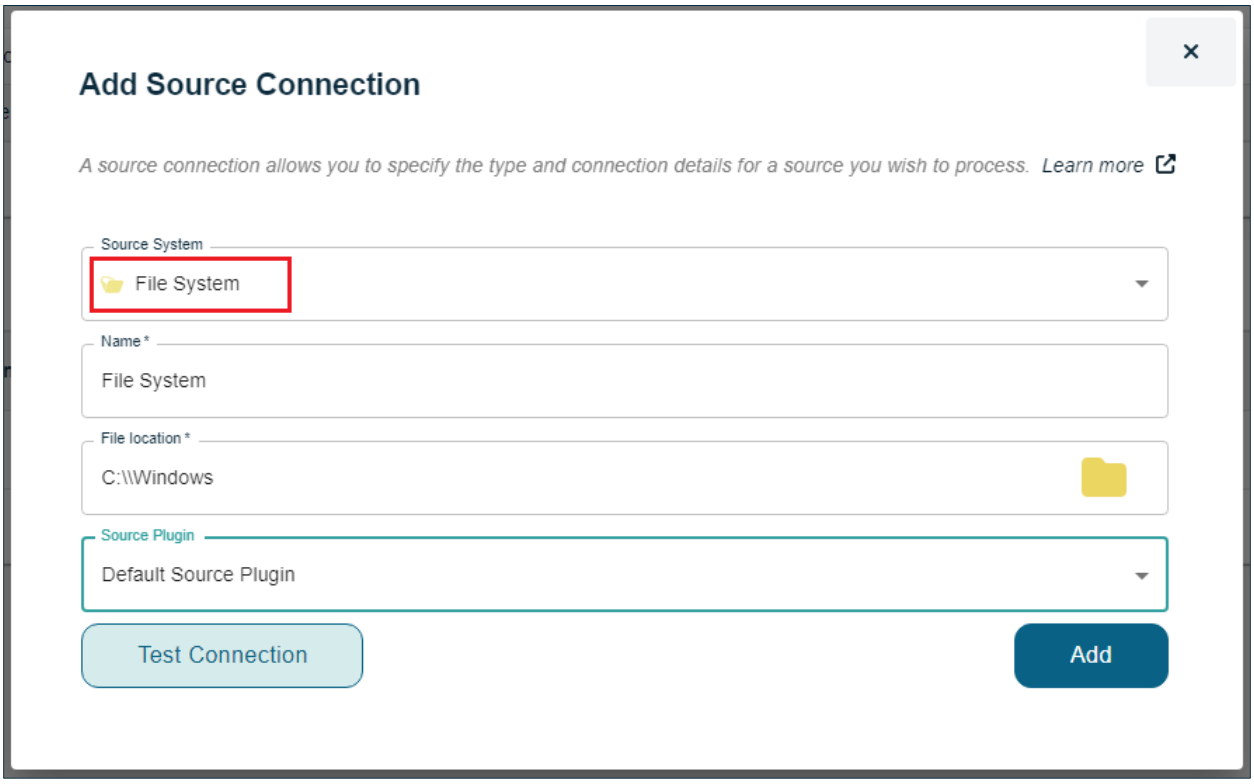


FIGURE 9 - SCREEN TO ADD FILE SYSTEM SOUCE CONNECTION

Name: Name of connection, to uniquely identify the connection within the CPS

File Location: Provide the location from where to migrate the content

Note: By default, mapped or Network Drives will not be shown in the Browse window, to resolve this follow steps in this article - https://support.proventeq.com/hc/en-us/articles/4409959298321

Source Plugin: Select "Default Source Plugin" for normal migration. In case any specific plug-in needs to use, choose JSON Import option in the dropdown

Test the Connection using Test Connection button before adding the Source Connection. Once test connection succeeded, Click on Add button to add connection.

3.2.2. Adding ShareFile as a Source

Once select ShareFile option in Source System dropdown, it will show following fields to add details related to it

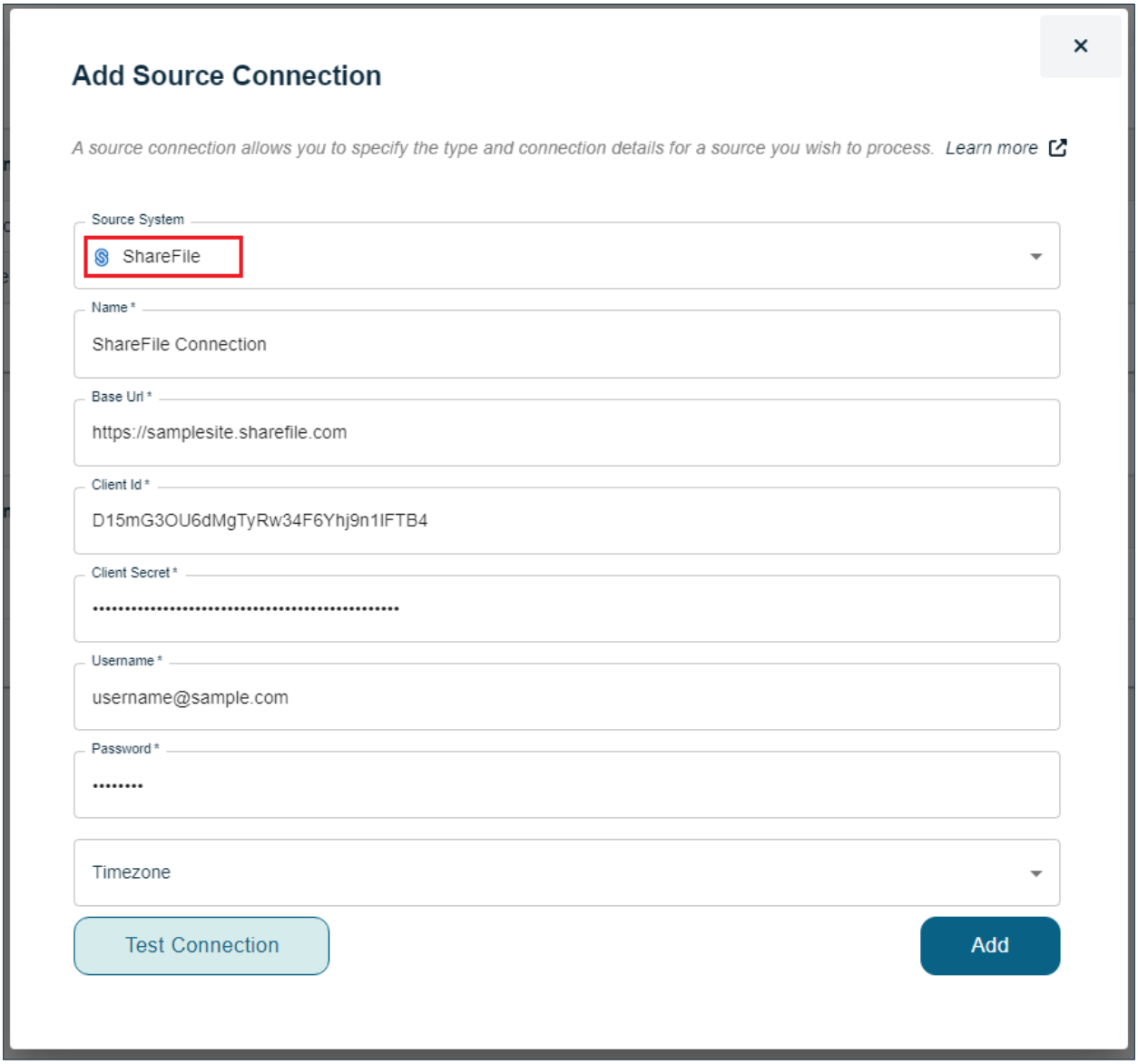


FIGURE 10 - SCREEN TO ADD SHAREFILE SOURCE CONNECTION

Name: Name of connection, to uniquely identify the connection within the CPS.

Base URL: URL of the system ShareFile. E.g. https://samplesite.sharefile.com

Client id: Client ID of the Source ShareFile.

Client Secret: Client Secret of the Source ShareFile.

NOTE: The Client id and client secret below can be obtained by clicking on Get an API Key on the following site https://api.sharefile.com

Username: Username for connecting to ShareFile. The user should be an administrator.

Password: Password for the above-mentioned user or application specific password if using MFA.

Note: If system is using Multi Factor Authentication (MFA) then it's necessary to generate an application specific password and use the password it generates here.

Steps:

Log into customer domain e.g. https://sfacme.sharefile.com

Go to Personal Settings > Personal Security > Two-Step Verification > scroll down to Application-Specific Passwords and click Generate Password.

Full details here https://support.citrix.com/article/CTX277723#generate

Time Zone: If Source and Target are in different time zones, choose appropriate time zone to convert date/time for different metadata fields.

Test the Connection using Test Connection button before adding the Source Connection. Once test connection succeeded, Click on Add button to add connection.

3.2.3. Adding Documentum Source

Documentum has a database and a file storage component so access details to both may be required. The exact information required depends upon several factors.

- Whether Documentum is using an Oracle database or a Microsoft SQL database.
- Whether using retrieval method of File Store or API

Retrieval Method controls how items should be retrieved from Documentum. The API retrieval method means that CPS will call the Documentum API to retrieve items. The File Store method means that CPS will access the Documentum files on disk directly to retrieve content. API method provides improved compatibility/reliability in some scenarios, but File Store can provide faster retrieval because it will bypass the Documentum system.

Once select Documentum option in Source System dropdown, it will show following fields to add details related to it

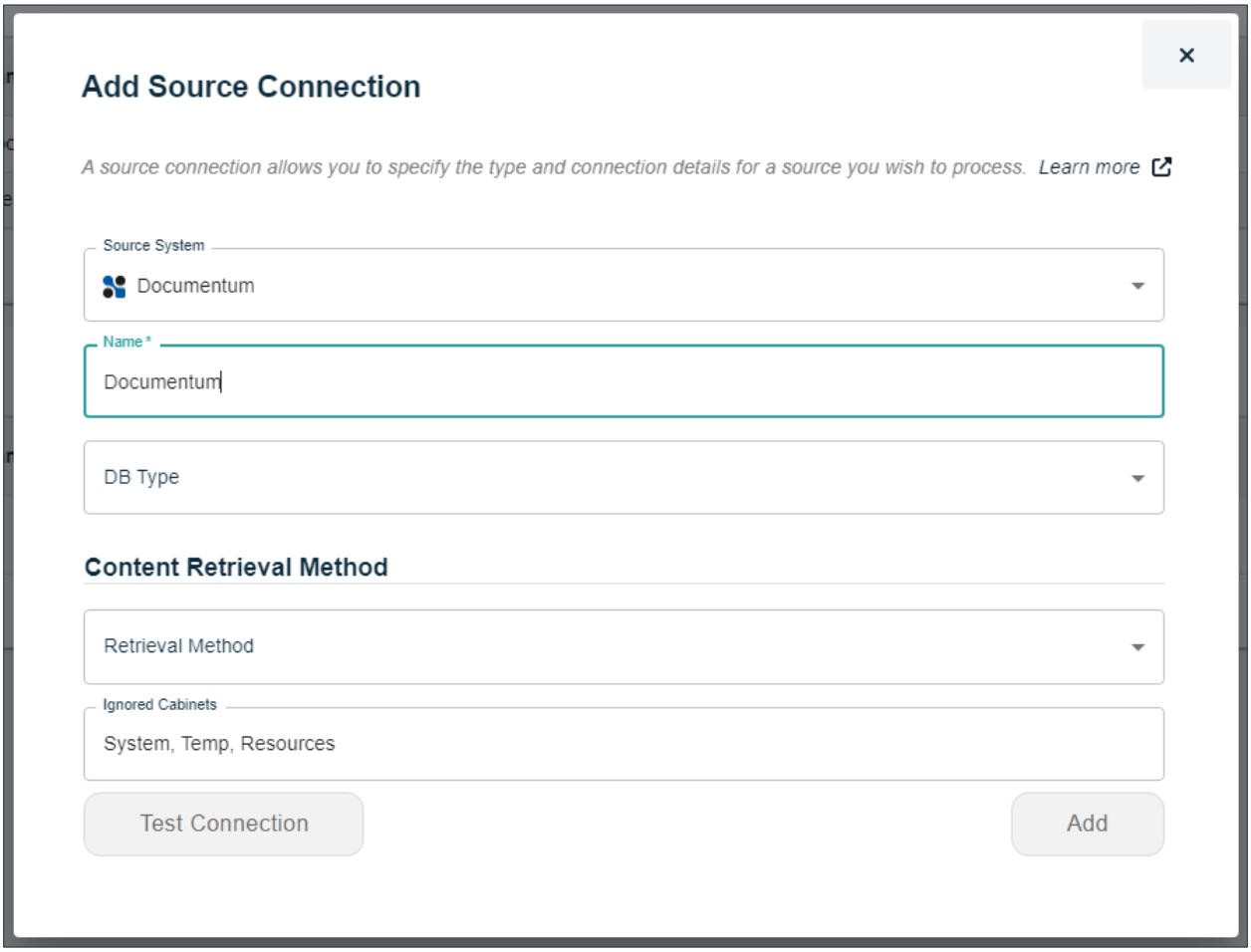


FIGURE 11 - SCREEN TO ADD DOCUMENTUM SOURCE CONNECTION

Name: Name of connection, to uniquely identify the connection within the CPS Application.

DB Type: This dropdown has two options SQL or Oracle (type of database used by Documentum)

If select Oracle, following additional fields will be displayed on screen

Host Database Server: Name of Oracle database server hosting Documentum database.

Port: Port number for Oracle database system

Radio button to select Whether to use SID or Service Name

SID/Service Name: Enter the SID or Service name based on selection of Radio button

Schema: Enter valid Oracle Schema name

Username & Password- Details of account with access to Documentum database.

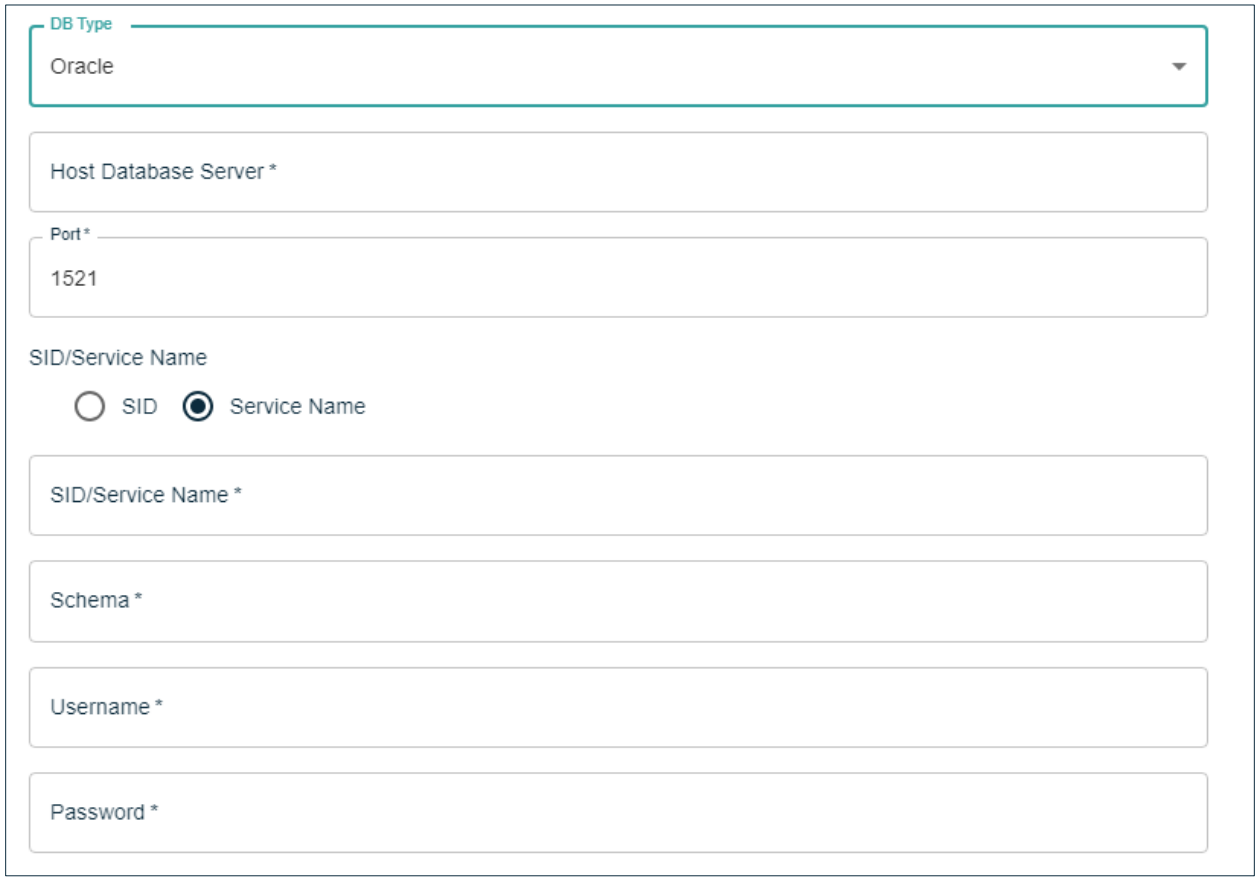


FIGURE 12 - SCREEN WHEN SELECTING ORACLE DB TYPE

If using **SQL**

Server: Name of SQL database server hosting Documentum database.

Port: Port number for Microsoft SQL server database system

Name of Documentum database: Documentum database name

Option to use Integrated Security. It selected the CPS service account (i.e. the account it us under will be used to access SQL. If not selected, then required to specify details username and password of a SQL login account with access.

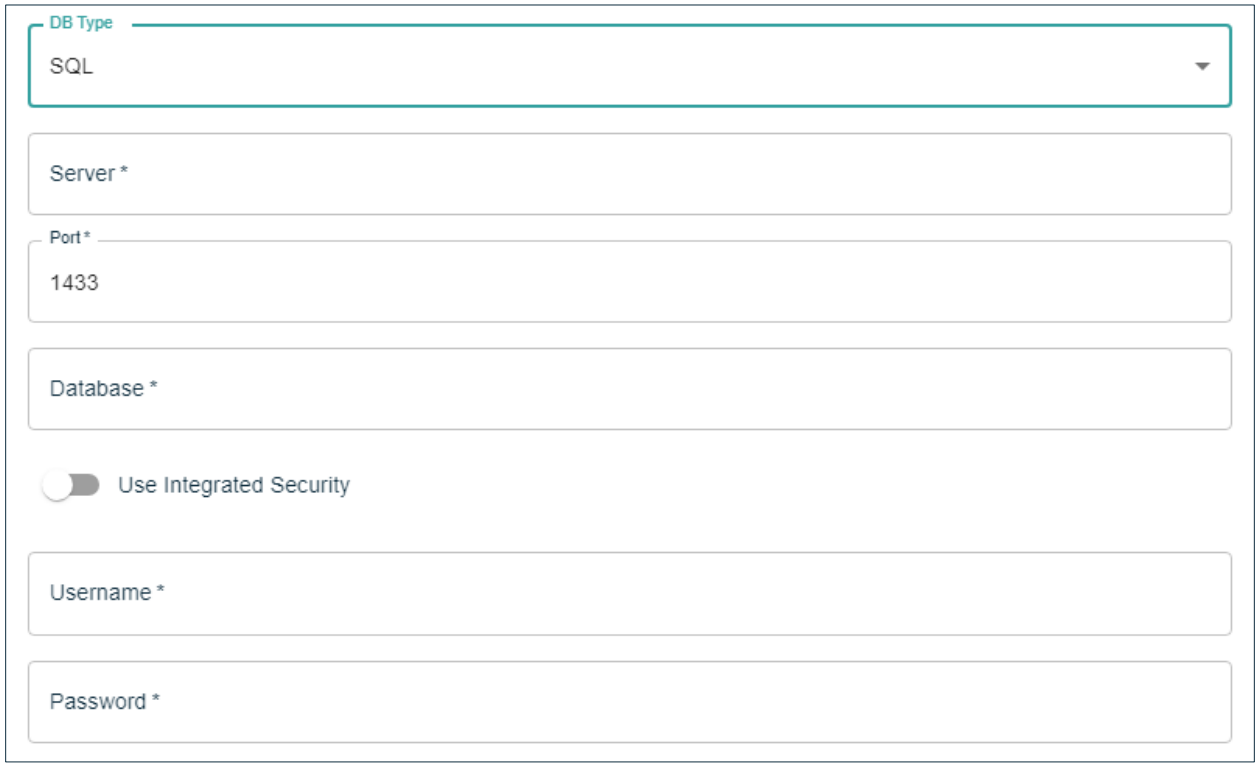


FIGURE 13 - SCREEN WHEN SELECTING SQL DB TYPE

Content Retrieval Method

This setting controls whether items are retrieved using requests to the Documentum API or by CPS accessing the information directly.

When Select **API**

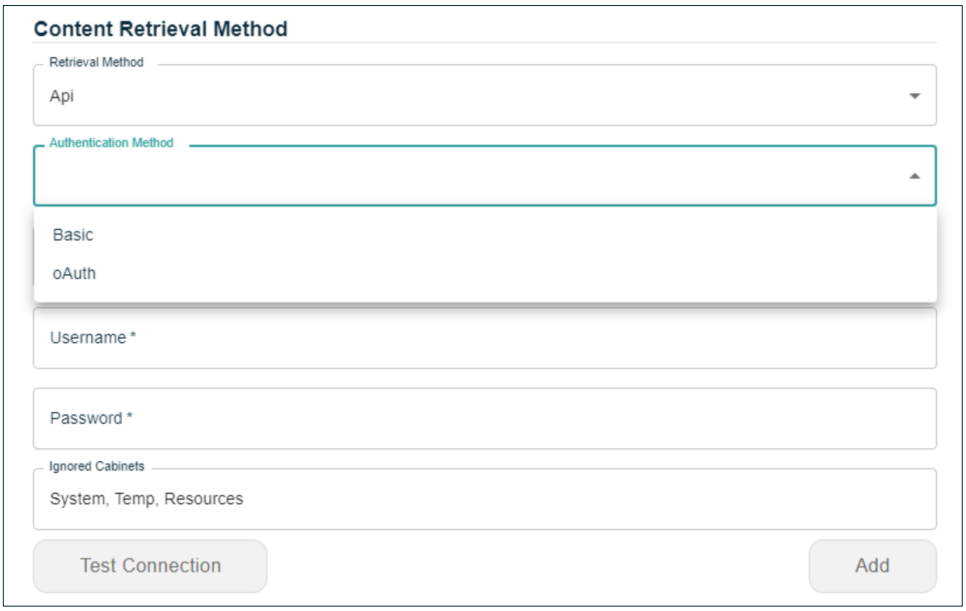


FIGURE 14 - SCREEN WHEN SELECTING API CONTENT RETRIEVAL METHOD

Authentication Method: Either Basic or OAuth can be used

Server URL: REST URL of the target Documentum Content Server. This is normally the hostname/IP address of the Documentum server followed by ":8080/dctm-rest/services" e.g. http://192.162.19.29:8080/dctm-rest/services

Username: : This is the Username for connecting to Documentum. The user should be an Administrator of Documentum.
Password: Password for the above user

Ignored Cabinets: Cabinets not to be processed e.g. System,Temp,Resources

When Select **File Store**

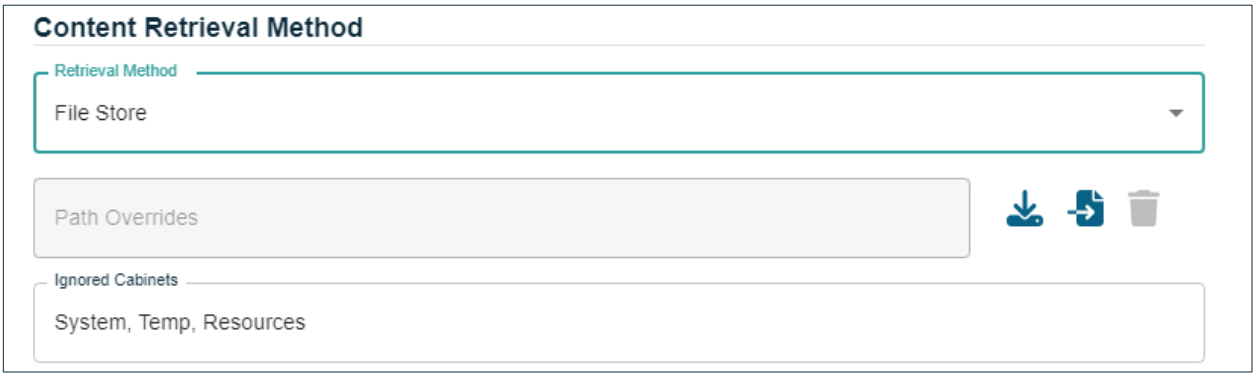


FIGURE 15 - SCREEN WHEN SELECTING FILE STORE AS CONTENT RETRIEVAL METHOD

Path Overrides: If the path in the database cannot be resolved from the CPS server, this feature can be used to 'map' that entry to an accessible location. Can also be used when source data is copied to another location.

The details of path overrides can also be specified in a .CSV file and imported. Below is example of formatting required.

```
fileStore,overrideFilepathValue
fileloc_1,\\fileserv1.acmecorp.com\Fin\US\storage_1
fileloc_19,\\fileserv1.acmecorp\Fin\UK\storage_19
```

NOTE: The override information is used during the discovery process to create source path URI for items. If the path override is edited after discovery, it may be necessary to reset and rerun discovery for these changes to be processed.

Ignored Cabinets: Cabinets not to be processed e.g. System,Temp,Resources

Test the Connection using Test Connection button before adding the Source Connection. Once test connection succeeded, Click on Add button to add connection.

NOTE: Virtual documents will be migrated as Document Sets in SharePoint so it is necessary to enable this feature on the SharePoint site collection.

3.2.4. Adding BOX Source

Once select BOX option in Source System dropdown, it will show following fields to add details related to it

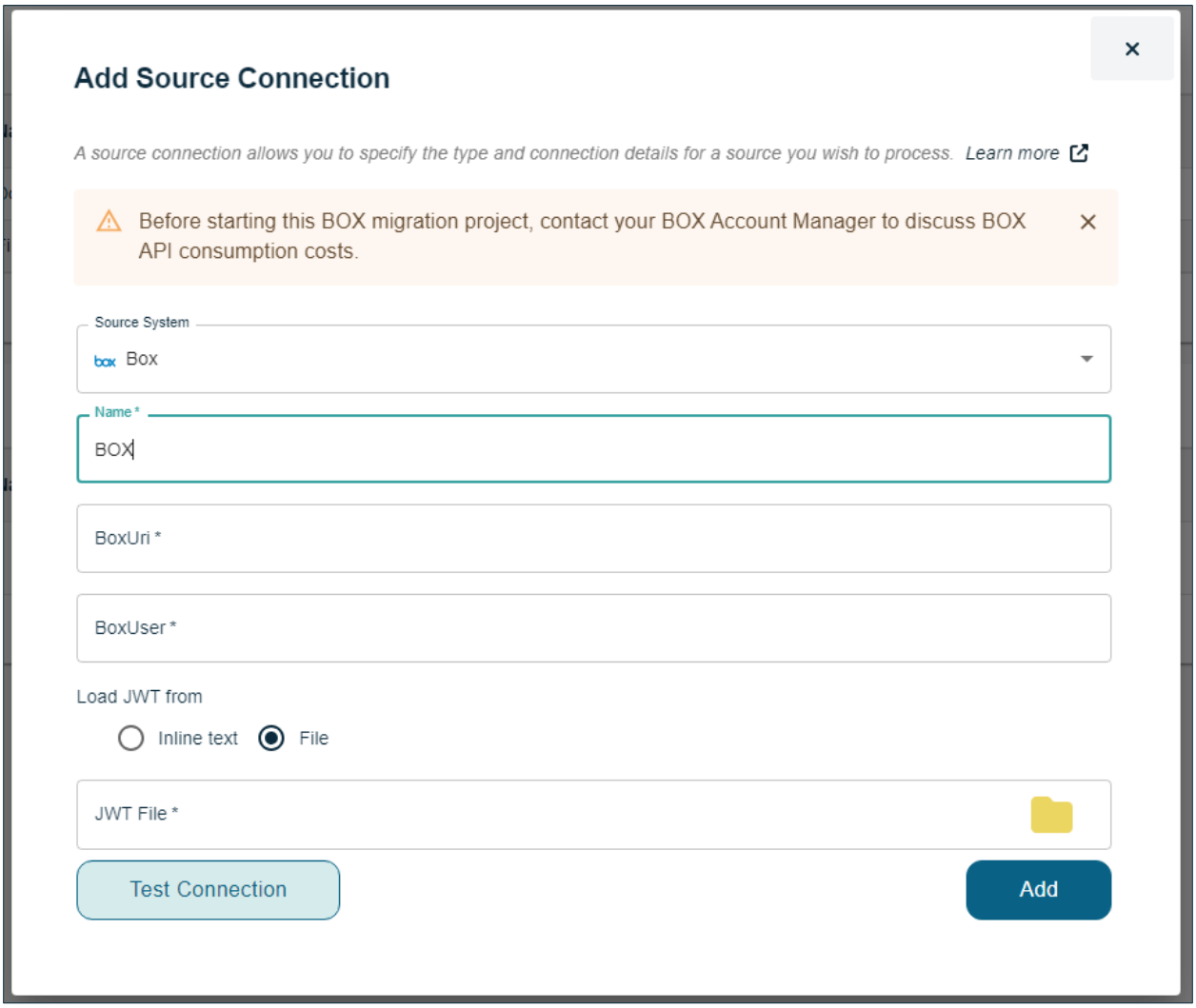


FIGURE 16 - SCREEN TO ADD BOX SOURCE CONNECTION

Name: Name of connection, to uniquely identify the connection within the CPS Application.

Box URL: URL of the Box system. Usually this will be https://app.box.com

Box User: Username of Box enterprise admin account

Load JWT from: The CPS Box connector supports the JWT (JSON web token) method of authentication. The token can either be located in a .JSON file or select 'Inline text' and enter the token in the specified field.

Inline text: String containing JWT token

File: Location of a file containing JWT token. Click on Folder icon to Browse the file location

Test the Connection using Test Connection button before adding the Source Connection. Once test connection succeeded, Click on Add button to add connection.

> NOTE: When migrating from BOX, API calls are used to discover and retrieve data. This is likely to have an impact on the API limits associated with your BOX system. Therefore it's essential to discuss your API call limits with your BOX Account Manager. See 19.2.2. for details of how migration consumes BOX API

3.2.5. Adding iManage Source

When select iManage option in Source System dropdown, it will show following fields to add details related to it

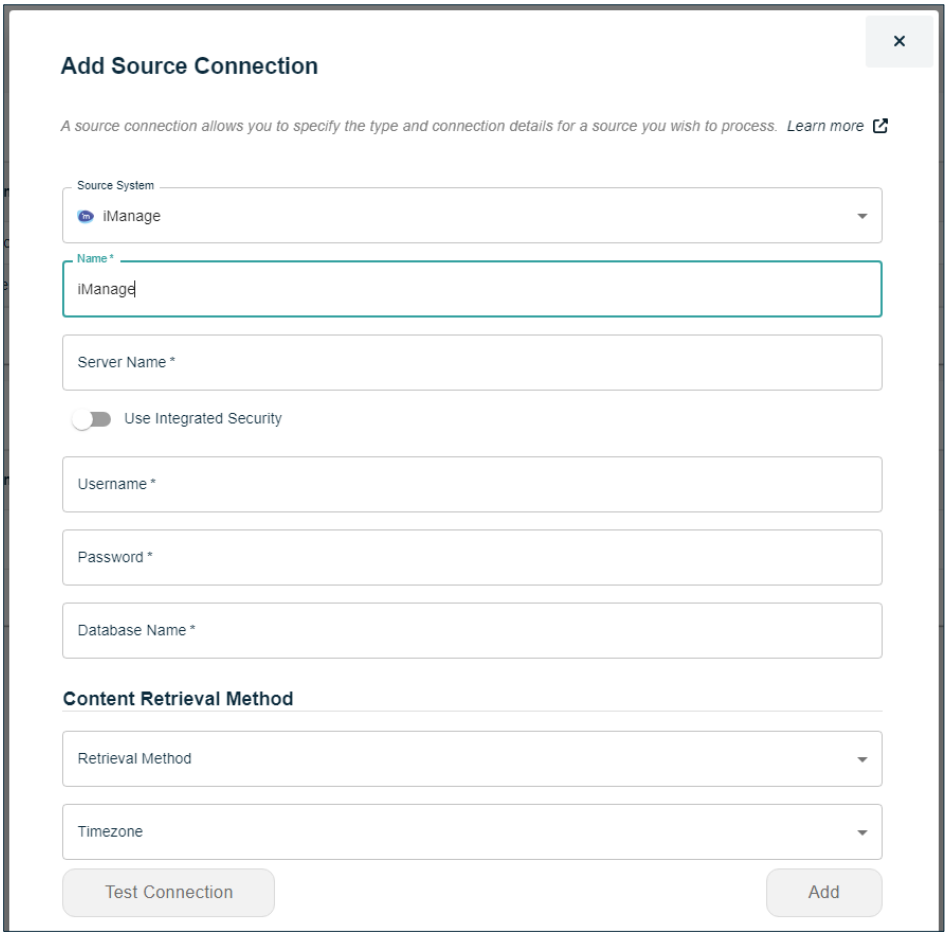


FIGURE 17 - SCREEN TO ADD IMANAGE SOURCE CONNECTION

Name: Name of connection, to uniquely identify the connection within the CPS application.

Server Name: Name of SQL server hosting iManage database.

Use Integrated Security – If selected then the account currently logged into will be used to access the iManage database.

Username and password: If **Use Integrated Security** is unselected then specify the username and password of an SQL Login account with access to the iManage database.

NOTE: To discovery and therefore migrate from iManage it's necessary to place a copy of the iManage database onto the same SQL server that the CPS Staging database is hosted on. If a hosted iManage system is being used, then contact iManage to request a copy of your database.

Content Retrieval Method:

There are two methods that can be used to retrieve iManage data, "API" or "File Store". If the iManage system is hosted or encrypted, then you MUST use the API method. The File Store method means that CPS will retrieve the underlying iManage files rather than use the iManage API. You may choose to copy the files to the CPS server to reduce network impact and improve performance.

If Select **File Store**

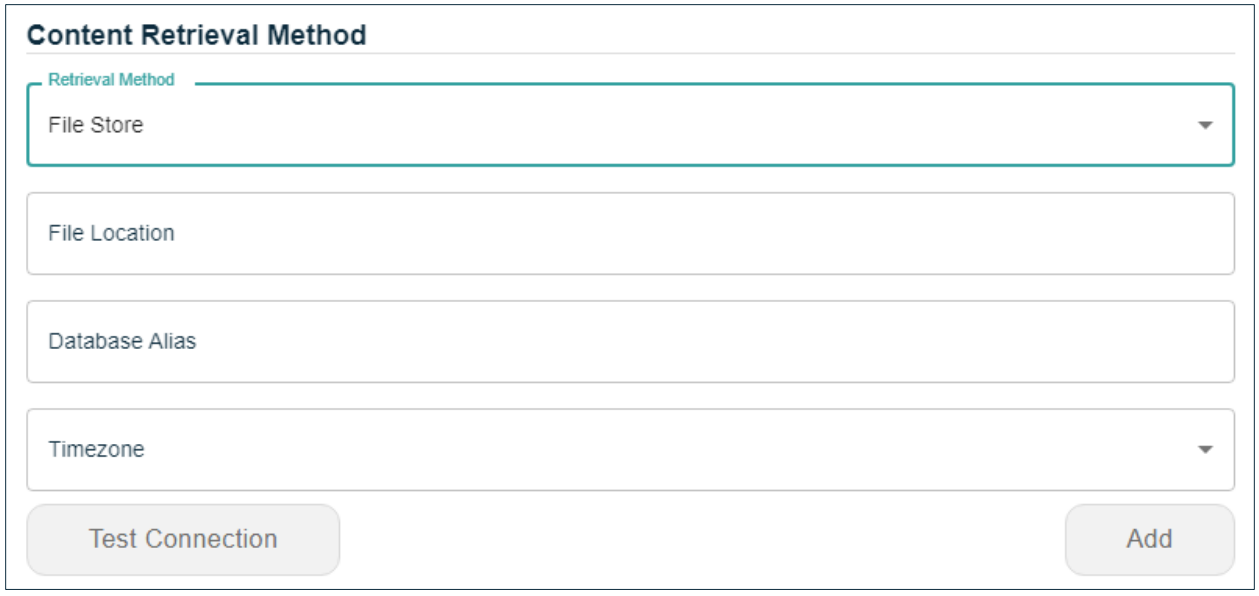


FIGURE 18 - SCREEN WHEN SELECTING FILE STORE AS CONTENT RETRIEVAL METHOD

File Location: If the iManage files have been copied from the iManage server to another location for migration, specify the location here.

Database Alias: Can be used to identify a specific iManage instance.

If Select **API**

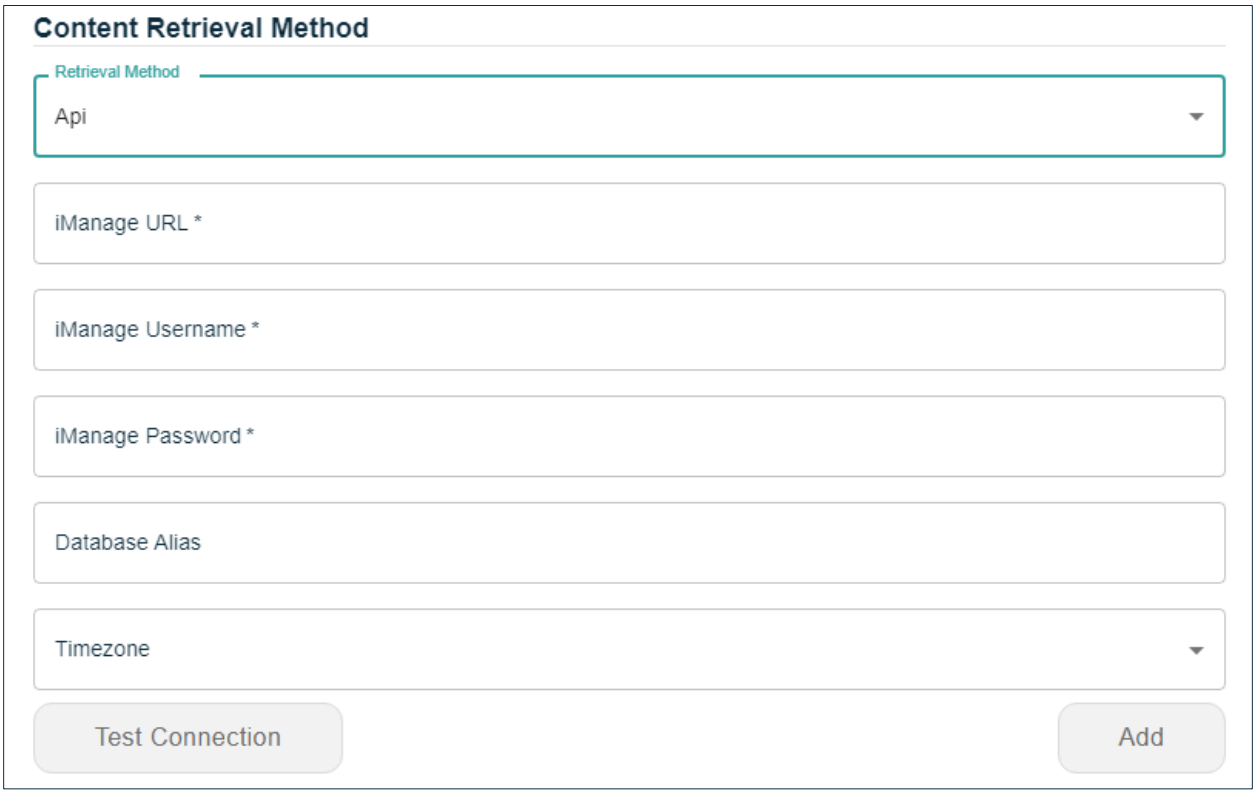


FIGURE 19 - SCREEN WHEN SELECTING API AS CONTENT RETRIEVAL METHOD

iManage URL: IManage URL

Username: Username to connect iManage system

Password: Password of above iManage connection user

Database Alias: Can be used to identify a specific iManage instance.

Test the Connection using Test Connection button before adding the Source Connection. Once test connection succeeded, Click on Add button to add connection.

NOTE: Migration of Workflows from iManage is not supported

3.2.6. Adding FileNet Source

When select FileNet option in Source System dropdown, it will show following fields to add details related to it

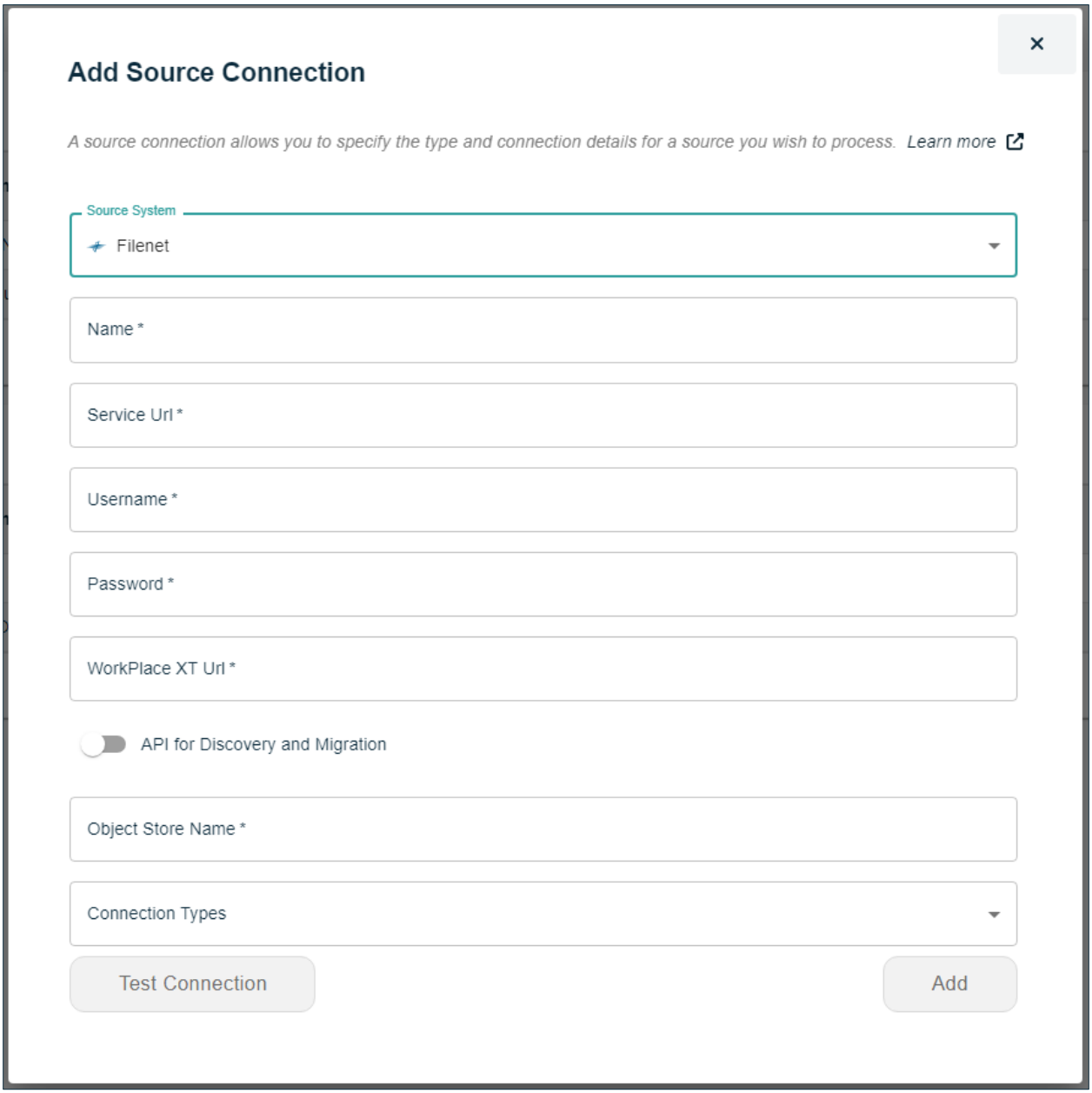


FIGURE 20 - SCREEN TO ADD FILENET SOURCE CONNECTION

Name: Name of connection, to uniquely identify the connection within the workspace.

Service URL: URL of the FileNet Web API service hosted on Content Engine server.

Username: FileNet Username for connecting to FileNet Content Engine. The user should have Administrator privileges. Domain accounts are not supported.

Password: Password for the above user.

WorkPlace XT URL: Normally of the form, e.g. "http://acmecorpfilenet:9080/WorkplaceXT". The FileNet WorkplaceXT is normally hosted on FileNet XT Server.

API for Discovery and Migration: With this toggle off(default), CPS will access the FileNet database directly during discovery and use the FileNet API for retrieval. If toggle selected, the FileNet API will be used for both discovery and retrieval.

Object Store name: Name of the FileNet Object store

Connection Types: Specify the database product used by FileNet. It can either be SQL, Oracle or DB2.

NOTE- CPS can either use the FileNet API for both discovery and retrieval or be configured to access the FileNet database directly. It supports FileNet using Microsoft, Oracle or DB2 databases.

When Select **Oracle** option in Connection Type drop down

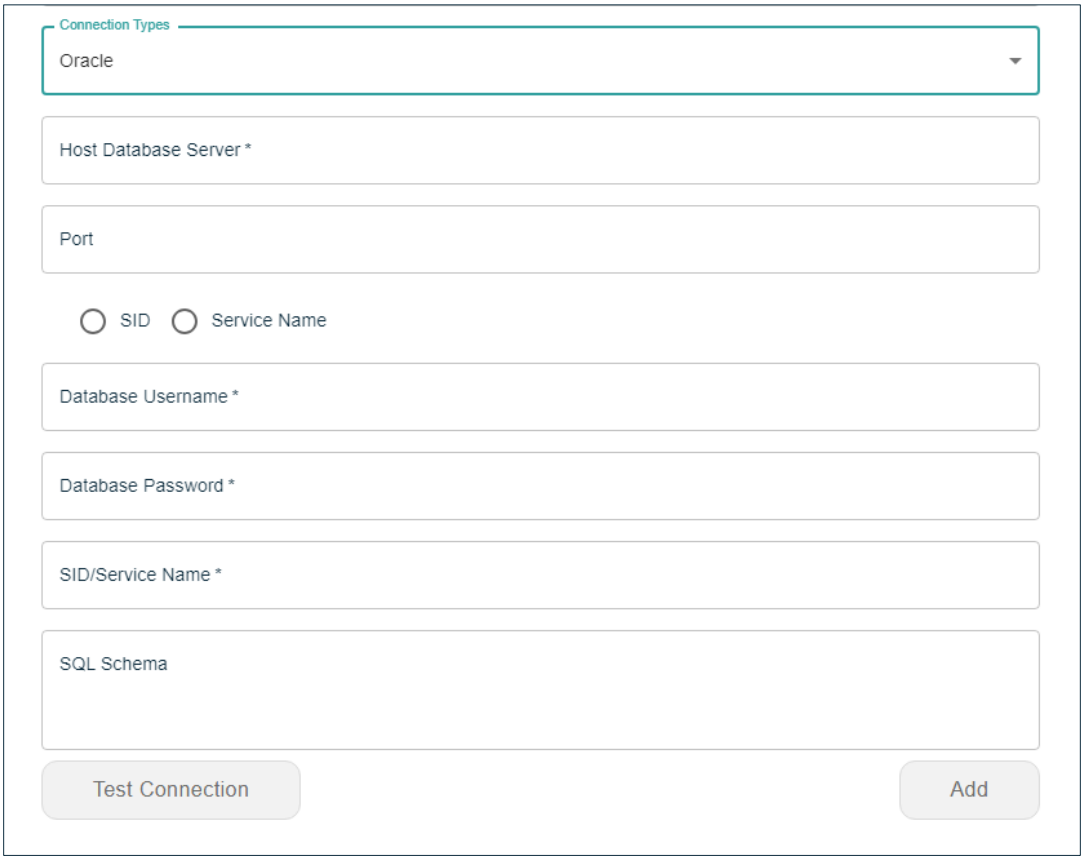


FIGURE 21 - SCREEN WHEN SELECT ORACLE AS CONNECTION TYPE

Host Database Server: Name of Oracle database server hosting FileNet database.

Port: Port number for Oracle database system

Radio button to select Whether to use SID or Service Name

Username and Password: Credentials of account with access to Oracle database
SID/Service Name: Name of SID/Service

SQL Schema: Oracle SQL schema

When Select **SQL** option in Connection Type drop down

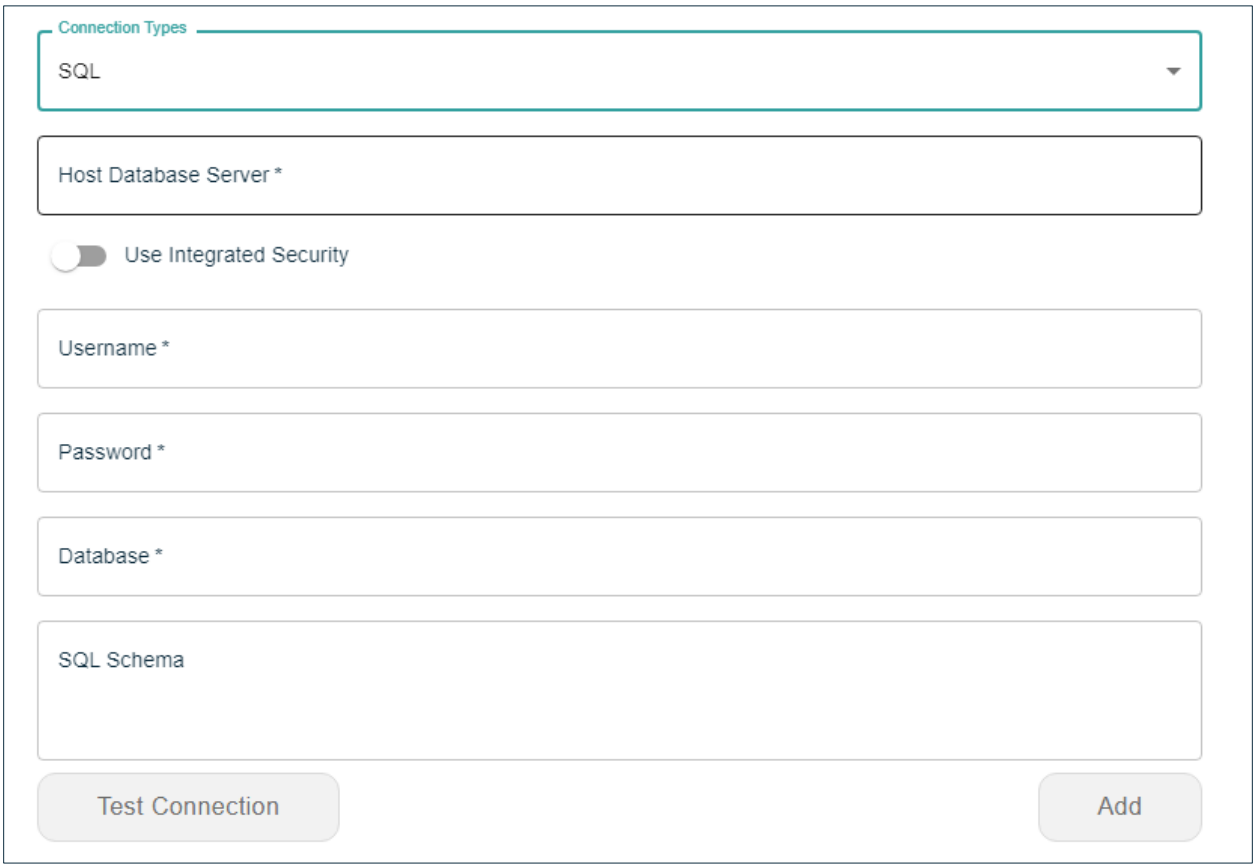


FIGURE 22 - SCREEN WHEN SELECT SQL AS CONNECTION TYPE

Host Database Server: Name of SQL database server hosting FileNet database.

Option to use Integrated Security. It selected the CPS service account (i.e. the account it runs under will be used to access SQL. If not selected, then required to specify details username and password of a SQL login account with access.

Database: Name of FileNet SQL database.

SQL Schema: Name if SQL schema e.g. dbo

When Select **DB2** option in Connection Type drop down

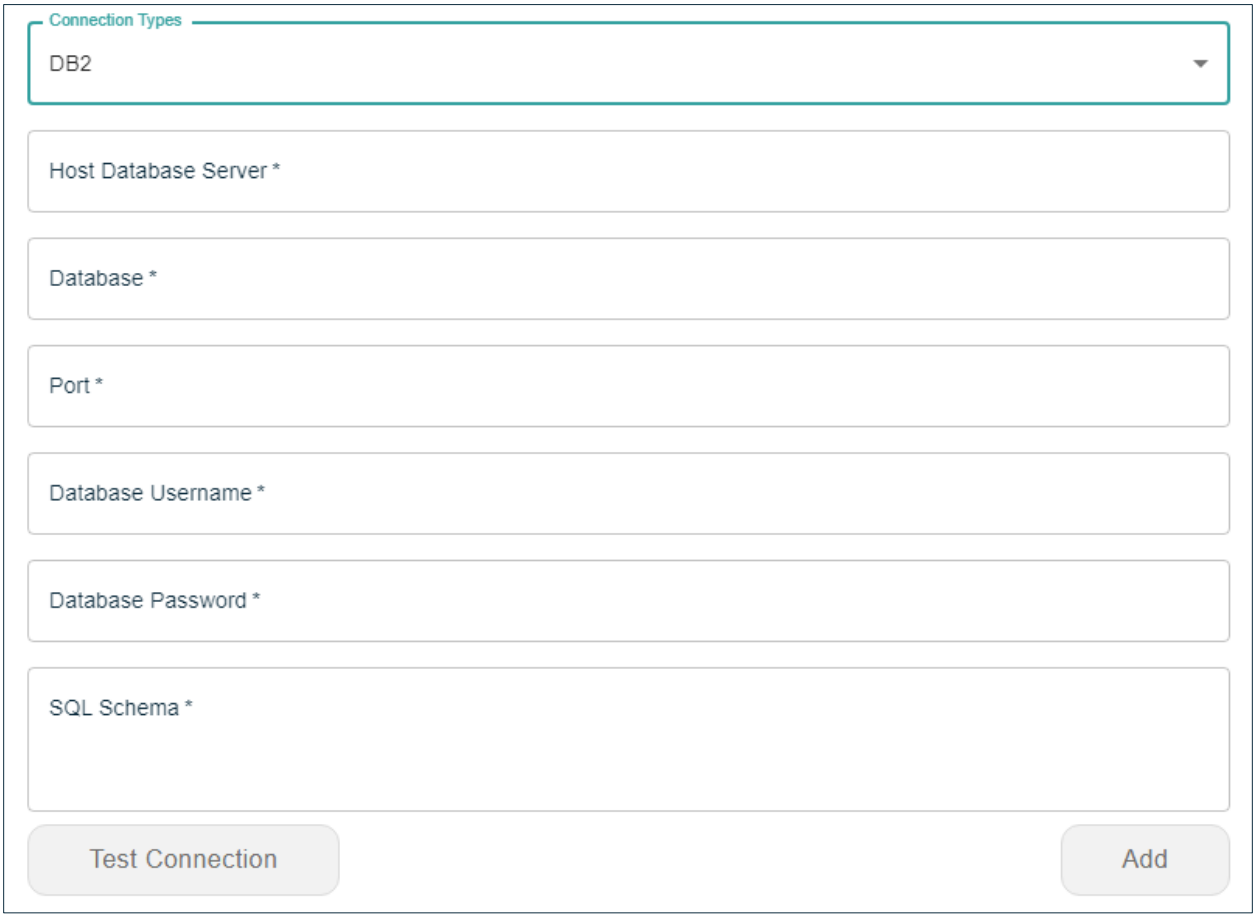


FIGURE 23 - SCREEN WHEN SELECT ORACLE AS CONNECTION TYPE

Host Database Server: Name of DB2 database server hosting FileNet database.

Database: Name of FileNet DB2 database.

Port: Port used to connect to DB2 database.

Username and password: Credentials to access DB2 database.

Test the Connection using Test Connection button before adding the Source Connection. Once test connection succeeded, Click on Add button to add connection.

3.2.7. Adding Alfresco Source

When select Alfresco option in Source System dropdown, it will show following fields to add details related to it

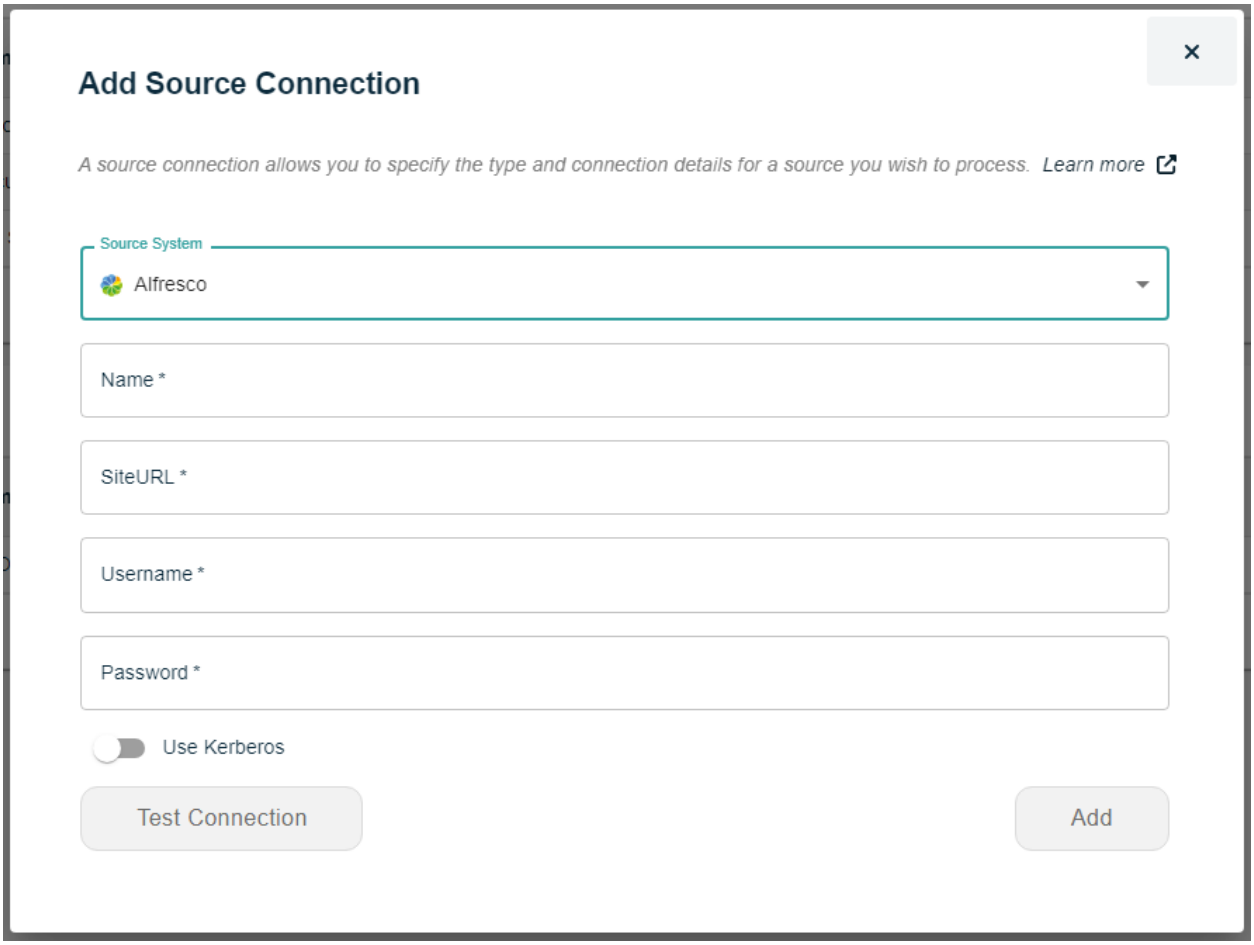


FIGURE 24 - SCREEN TO ADD ALFRESCO SOURCE CONNECTION

Name: Name of connection, to uniquely identify the connection within the CPS Application.

SiteURL: URL for Alfresco server. This is either in the format

http://<Alfresco server>:<port>/alfresco/api/-default-/public/cmis/versions/1.1/browser
e.g. http://prodcms:8080/alfresco/api/-default-/public/cmis/versions/1.1/browser
or
http://<Alfresco server>:<port>/alfresco/api/-default-/public/cmis/versions/1.0/atom

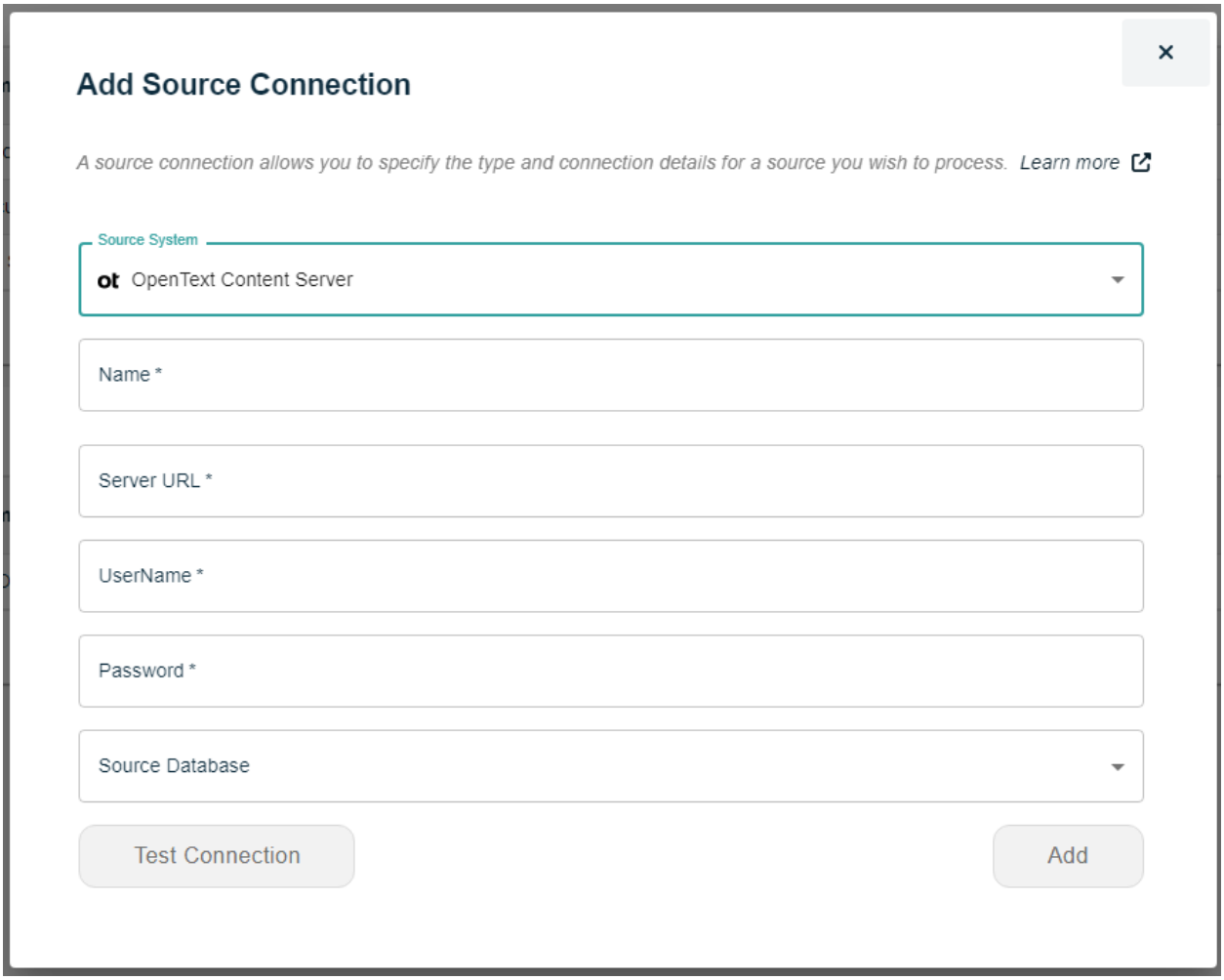
Username and Password: Details of account with access to Alfresco server

Use Kerberos: Select to use Kerberos

Test the Connection using Test Connection button before adding the Source Connection. Once test connection succeeded, Click on Add button to add connection.

3.2.8. Adding OpenText CS Source

When select OpenText Content Server option in Source System dropdown, it will show following fields to add details related to it



N**ame:** Name of connection, to uniquely identify the connection within the workspace.

Server URL: URL of OpenText Content System. Eg. http://<opentextserver>/otcs/cs.exe/

Username and Password: Details of account with access to OpenText CS server

Source Database: Select whether OpenText is using SQL Server or Oracle as it's backend database.

When Select **SQL** in Source Database option

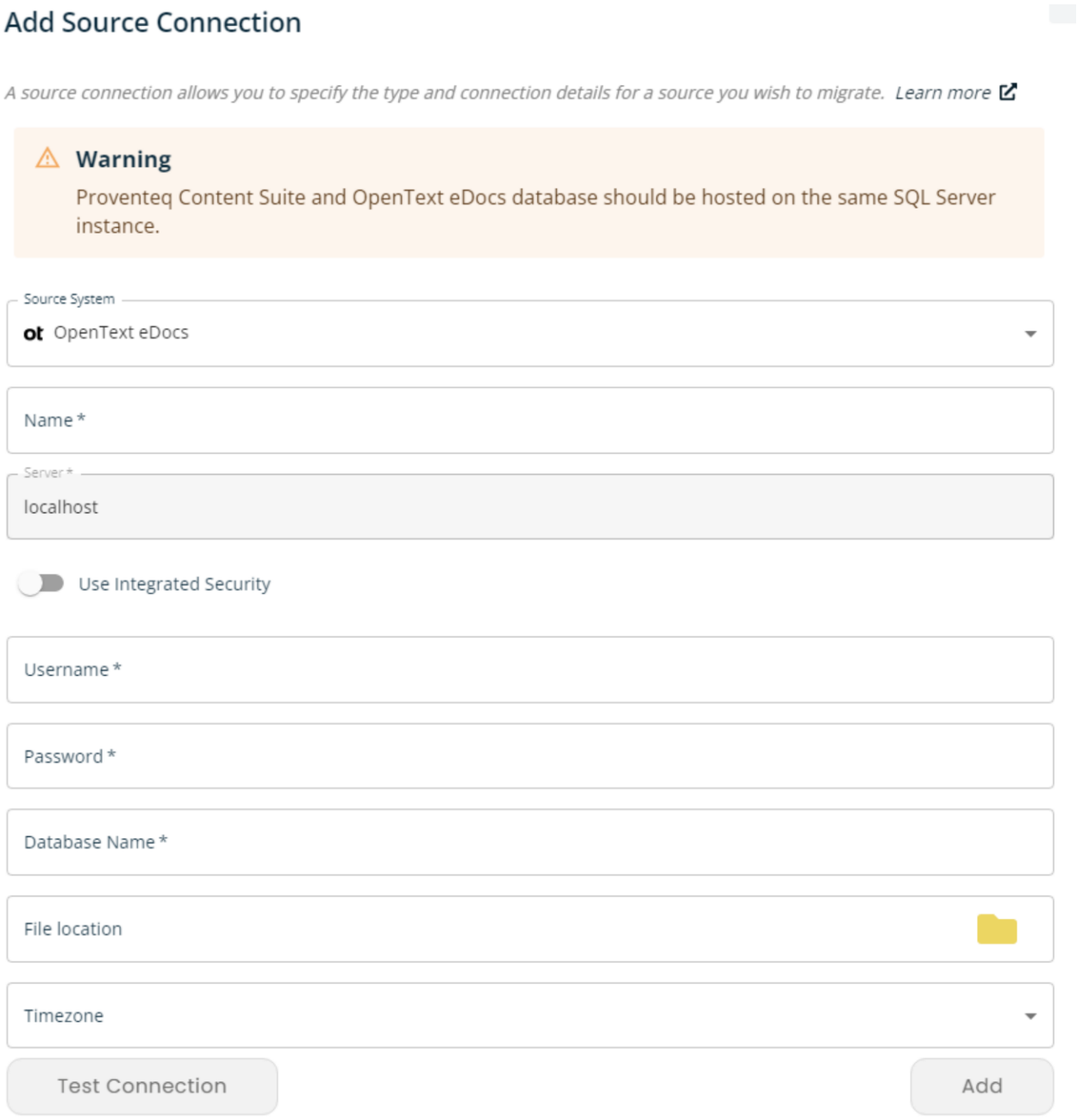
Database server name: Name of Microsoft SQL server storing OpenText database.
Database name: Name of OpenText SQL database
Use Integrated Security – If selected then the account currently logged into will be used to access the database.
Username and password: If **Use Integrated Security** is unselected then specify the username and password of an SQL Login account with access to the database.
SQL Schema: SQL schema name of OpenText Database.

When Select **Oracle** in Source Database option

Test the Connection before adding the Source Connection

3.2.9. Adding OpenText Edocs Source

To discover and therefore migrate from EDocs, it's necessary to place a copy of the EDocs database onto the same SQL server that the CPS Staging database is hosted on.

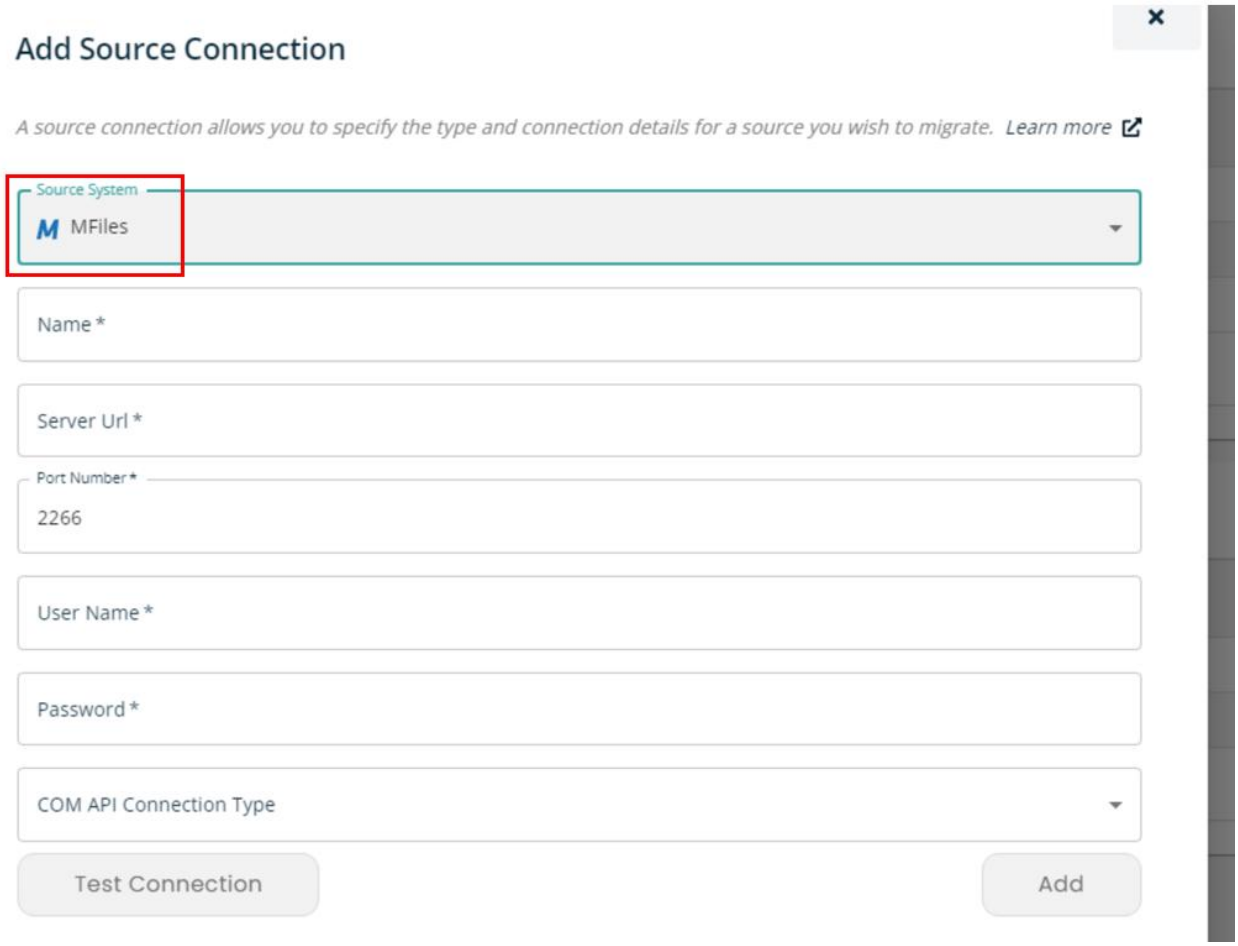


Name: Name of connection, to uniquely identify the connection within the workspace.
Server: Name of SQL server hosting EDocs database.
Use Integrated Security – If selected then the account currently logged into will be used to access the EDocs database.
Username and password: If **Use Integrated Security** is unselected then specify the username and password of an SQL Login account with access to the EDocs database.

Database Name: Name of EDocs SQL database
File location: Location where EDocs files are stored. If the documents to be discovered are located across multiple file locations this can be customised by Proventeq with Premium Support or Professional Services.

Test the Connection before adding the Source Connection

3.2.10. Adding M-Files Source

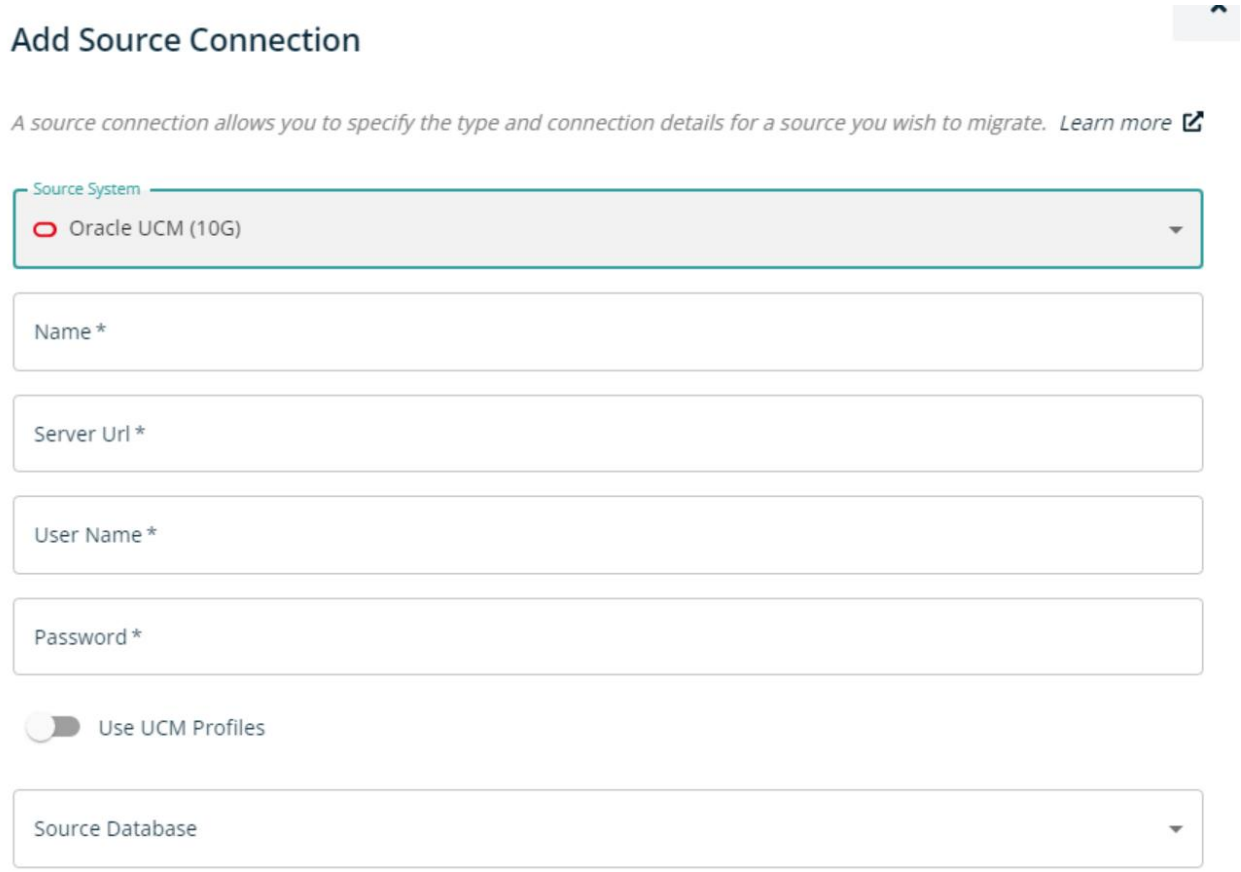


Name: Name of connection to uniquely identify the connection within the workspace.
ServerUrl: URL of the source M-Files server.
Port Number: Port number for the COM API utilization
Username: Username for connecting to M-Files server. The user should be an administrator for the source M-Files system.
Password: Password for the above-mentioned user.
COM API Connection Type: HTTPS or TCP connection.

Test the Connection before adding the Source Connection

3.2.11. Adding Oracle UCM/WCC Source

CPS support migrations from Oracle UCM/WCC that use either Microsoft SQL or Oracle as their underlying database.

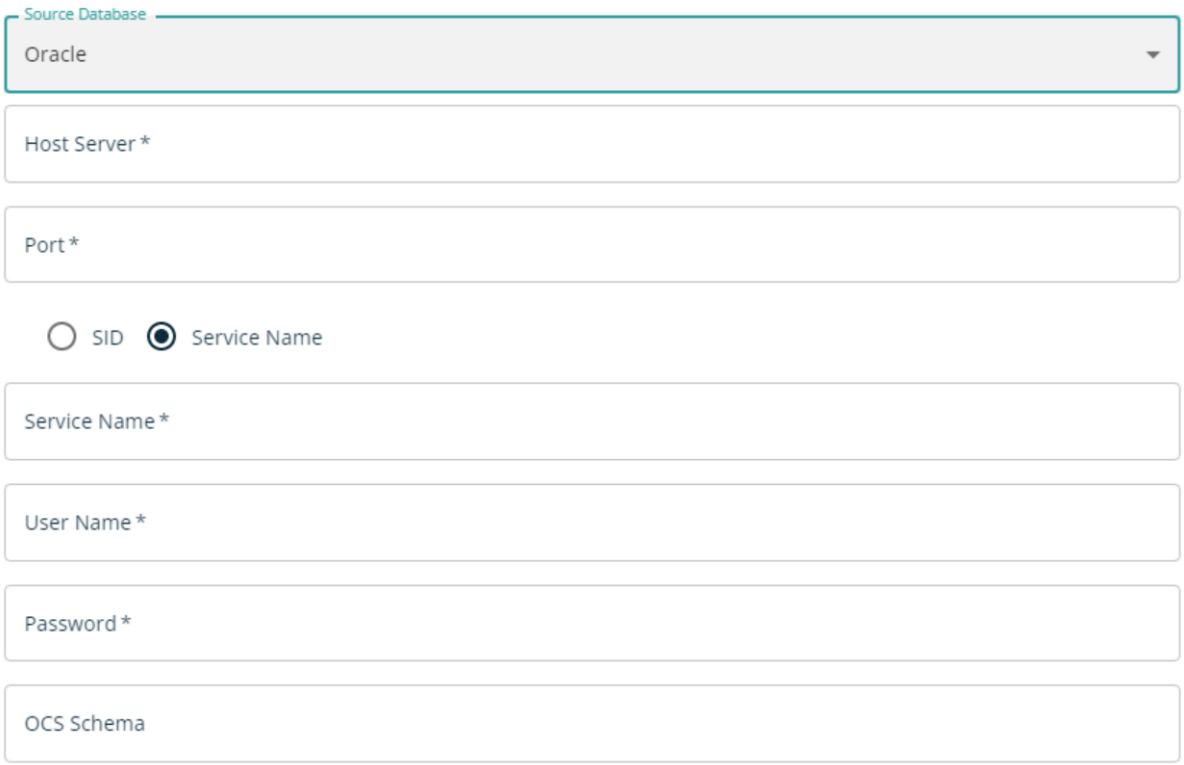


Name: Name of connection to uniquely identify the connection within the workspace.
Server Url: URL of the source Oracle UCM server. e.g. http://ucmsys.acmecorp.com/16200/_dav/cs/idcplg
Username & Password: Credentials of account with Oracle UCM access
Use UCM Profiles –

Oracle UCM using SQL Server

Host Server: Name of SQL database server hosting Oracle UCM database.
Name of Oracle UCM database:
Port: Port number used by SQL server
Option to use Integrated Security. It selected the CPS service account (i.e. the account it runs under will be used to access SQL. If not selected, then required to specify details username and password of a SQL login account with access.
SQL Schema: e.g. dbo

Oracle UCM using Oracle Database

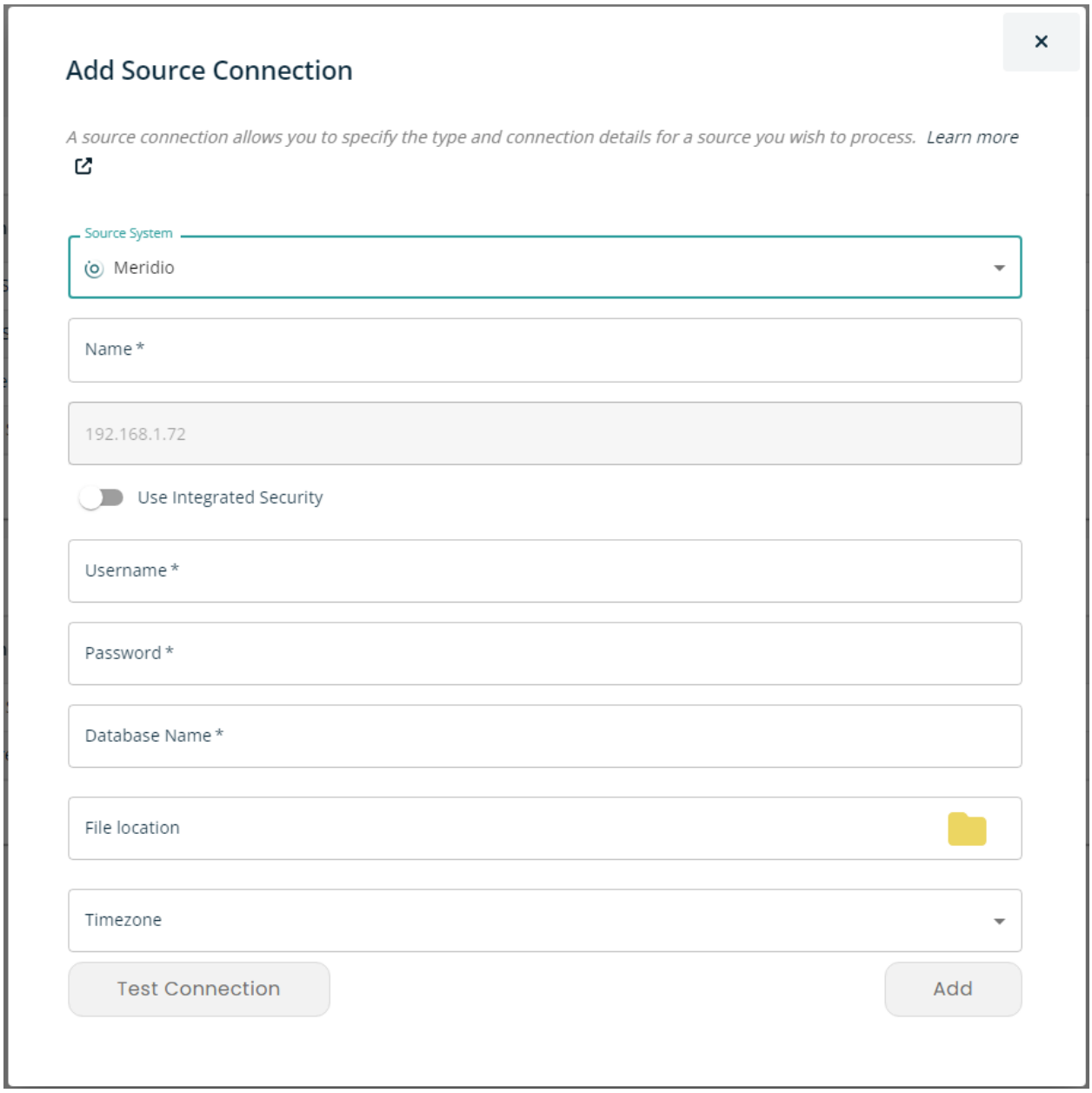


Host Database Server: Name of Oracle database server hosting Oracle UCM database.
Port: Port used by Oracle database server
Whether to use SID or Service Name
SID/Service Name:
Username & Password- Details of account with access to Oracle UCM database.
OCS Schema:

Test the Connection before adding the Source Connection

3.2.12. Adding Meridio Source

CPS support migrations from Meridio that use Microsoft SQL as their underlying database. Once select Meridio option in Source System dropdown, it will show following fields to add details related to it



The following fields need to be configured for a Meridio connection:

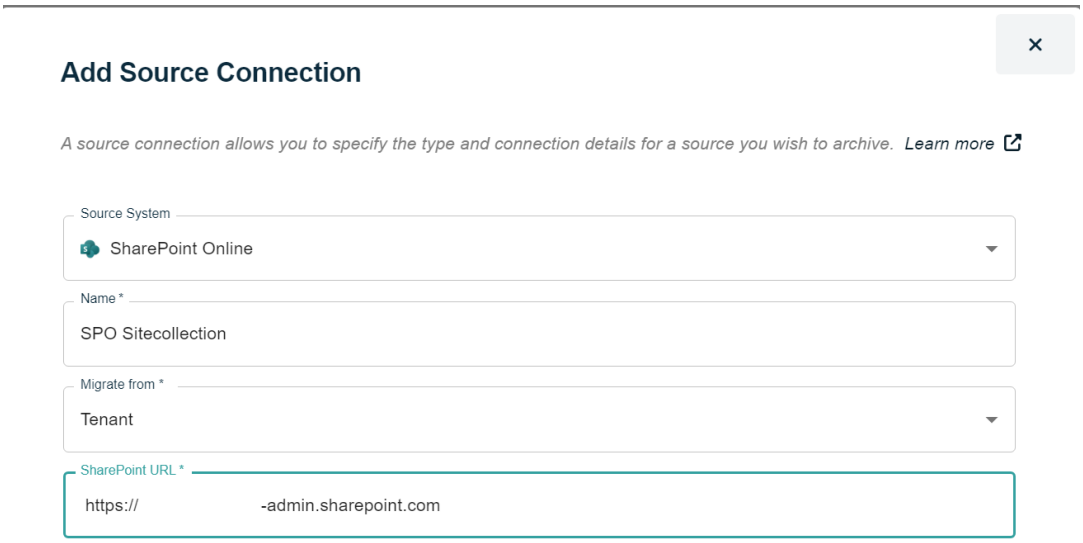
- **Name**: Name of connection, to uniquely identify the connection within the workspace.
- **Server Name:** Name of the database instance hosting the Meridio database. **Proventeq Migration Accelerator staging database and Meridio database should be hosted on the same SQL Server instance**. e.g. DBServer\Instance.
- **Use Integrated Security:** Checked the checked box to use the Integrated Security.
- **UserName**: Username for connecting to Meridio database. The user should be an Administrator for this database.
- **Password**: Password for the above-mentioned user.
- **Database Name**: Database Name of the Meridio.

- **File Location**: Full path to the folder storing the Meridio Files.
- **Time zone**: Dropdown to select relevant time zone in case Meridio data need conversion due to time zone difference.

Test the Connection before adding the Source Connection

3.2.13. Adding SharePoint Online as Source

Select SPO from the dropdown list and add the details in the source connection screen as stated below



The following fields needs to be configured for a SharePoint Connection:

- **Name:** Name of connection to uniquely identify the connection
- **Connect To:** Select **Tenant** or **Site Collection**.
- **SharePoint URL:** The SharePoint Tenant, Site collection or Team URL

Connect To: Tenant or Site Collection

- **Site Collection** – To archive documents from a particular site collection.
- **Tenant** – A Tenant is the parent of all Site Collections. This Tenant option can be used when there is a need to archive data into more than one Site Collection. Note the tenant role SharePoint Administrator is required when using this option to allow it to see across all Site Collections.

SharePoint Authentication Methods

There are several authentication methods that can be used to connect to SharePoint

- **Error! Reference source not found.** (recommended)
- **Error! Reference source not found.**
- **Error! Reference source not found.**

SharePoint Online – App Based Authentication

The general recommendation is to use App based Authentication whenever possible.

You will need to provide the URL below to your Global Administrator to provide admin consent and return to the application after they have granted consent.

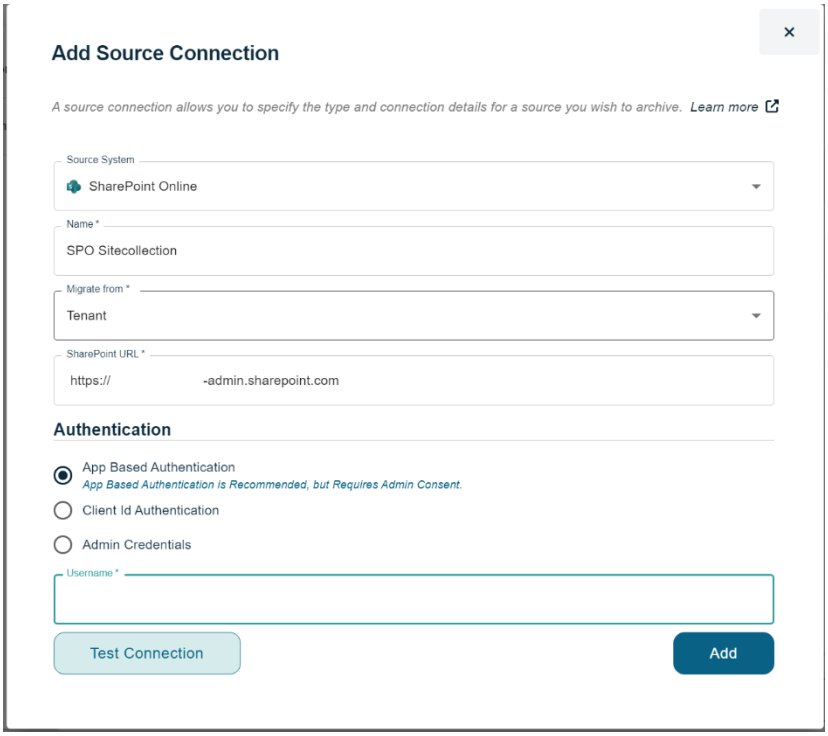


FIGURE 117 - USING APP BASED AUTHENTICATION

App Based Authentication Settings:

- **Username**: The Username for connecting to SharePoint. If connect is to **Tenant** option, then user should be SharePoint Online Admin or Global Admin of the Office365 tenant.

If connection is to **Site Collection,** then the user should be a site collection administrator.

Click on the "Test Connection" button, and Connection successful message will be displayed on top of the screen.

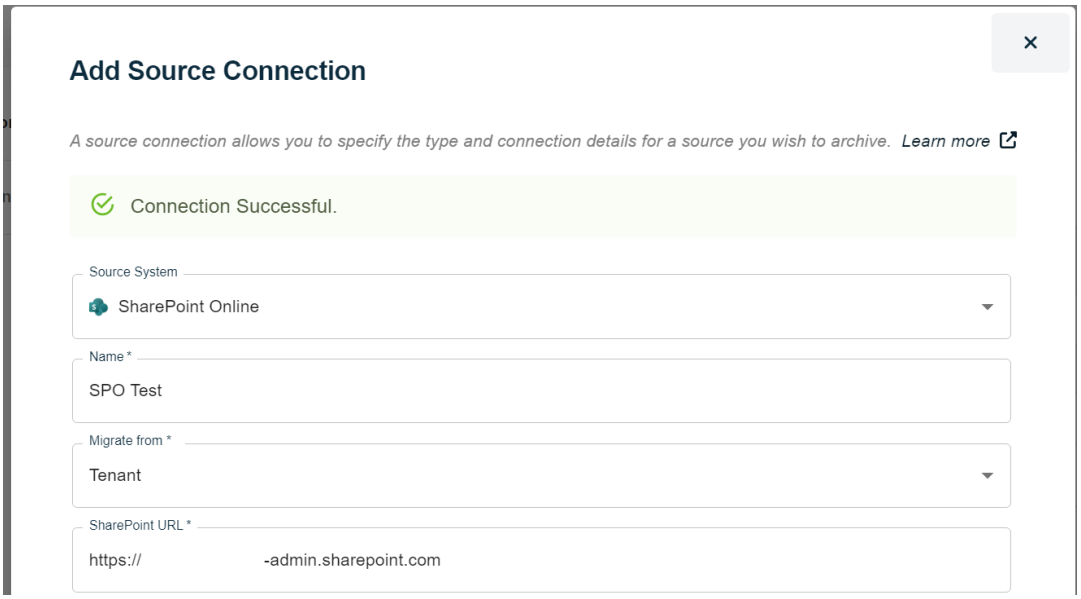


FIGURE 118 -CONNECTION SUCCESSFUL

Click on the “Add” button to add the SharePoint connection.

SharePoint Online - Client Id Authentication

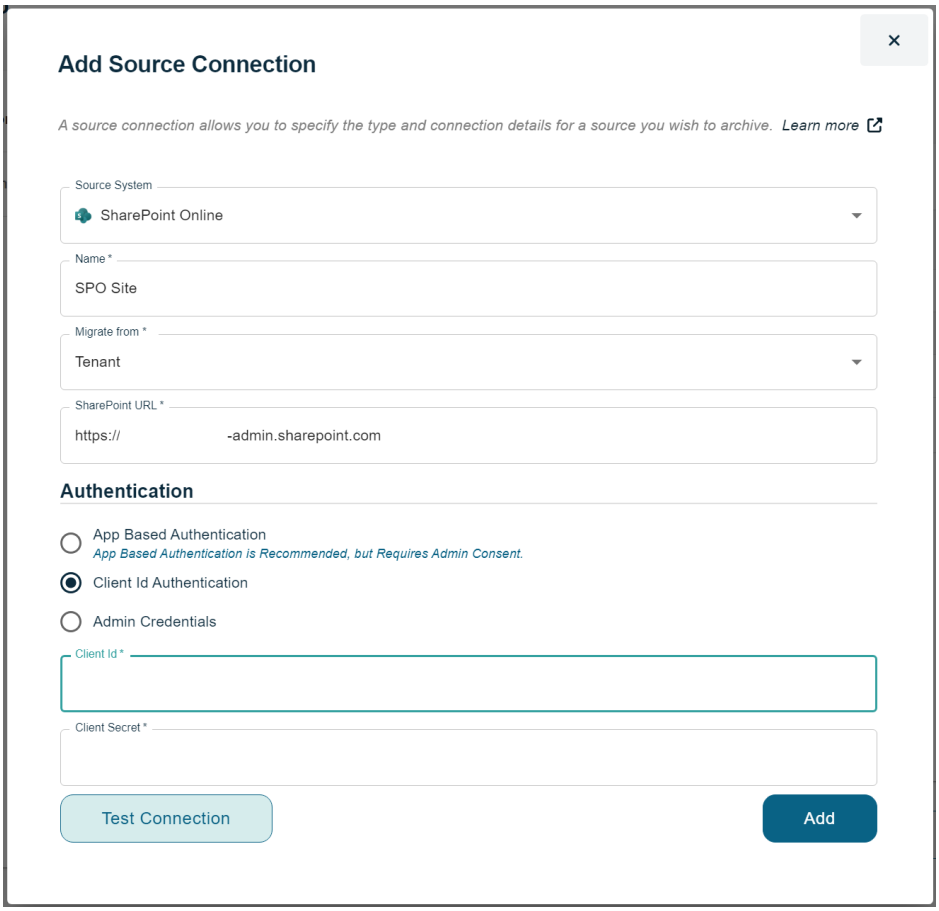


FIGURE 119 -USING CLIENT ID BASED AUTHENICATION

Client Id Authentication Settings:

- **Client ID**: The app-only principal Client ID.
- **Client Secret**: The app-only principal Client Secret.

Click on the "Test Connection" button to verify that the details entered are correct.

Click on the “Add” button to add the SharePoint connection.

SharePoint Online - Admin Credentials Authentication

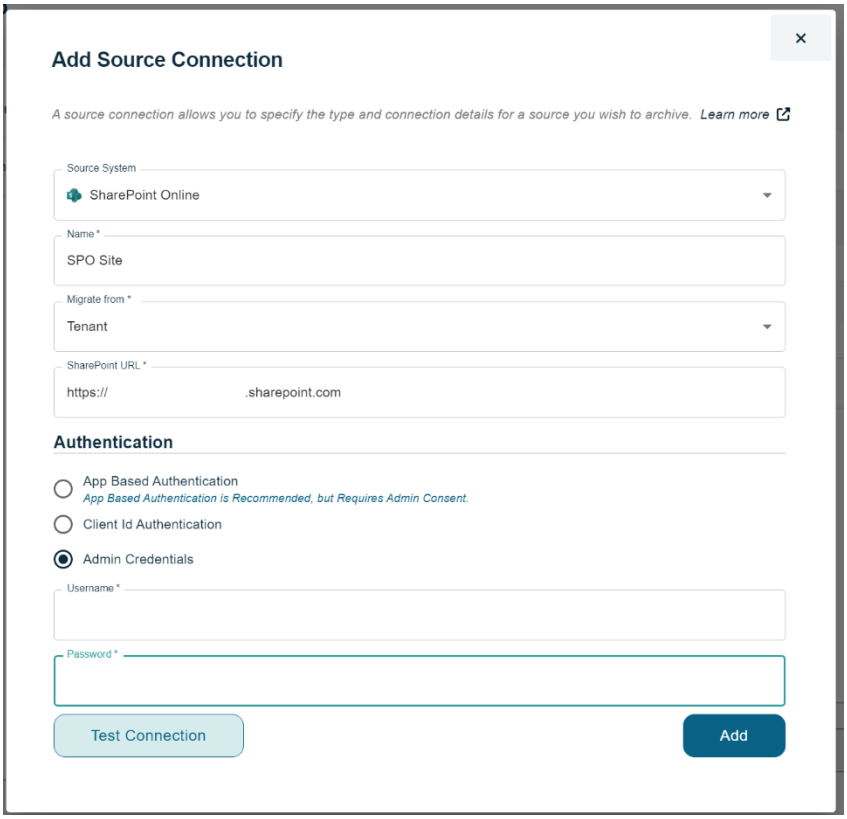


FIGURE 120 -USING ADMIN CREDENTIALS AUTHENICATION

Admin Credentials Authentication Settings:

- **Username**: The Username for connecting to SharePoint. The user should be a site collection administrator for the target Site Collection or SharePoint Online Admin for the tenant level connection.
- **Password**: The Password for the above-mentioned user.

3.3 Add a New Target Connection

Adding a new target connection can be created from the Connections option displayed in the Navigation Panel. The target systems shown is determined by the CPS licence that has been applied. Often it will by a single target.

FIGURE 25 - SELECTING TARGET SYSTEM TYPE

3.3.1. Adding SharePoint Online Target

For more information on authentication options see Accessing Microsoft 365 from PMA – Proventeq Migration Accelerator.

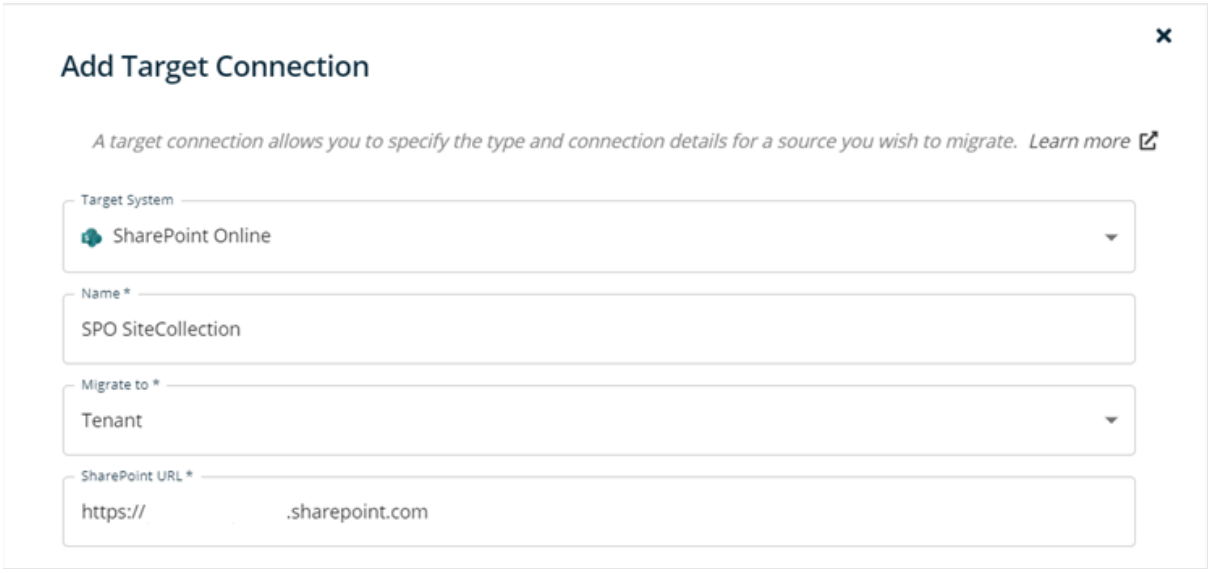


FIGURE 26 - SHAREPOINT ONLINE CONNECTION DETAILS

The following fields needs to be configured for a SharePoint Connection:

- **Name:** Name of connection to uniquely identify the connection
- **Connect To:** Select **Tenant** or **Site Collection**.

- **SharePoint URL:** The SharePoint Tenant, Site collection or Team URL

Connect To: Tenant or Site Collection

- **Site Collection** – To migrate to a particular site collection.
- **Tenant** – A Tenant is the parent of all Site Collections. This Tenant option can be used when there is a need to migrate data into more than one Site Collection or when migrating data to OneDrive. Details of the specific Site Collection within the Tenant to migrate to can be specified in the Migration Job Creation wizard or later, using Structure Mapping feature. Note the tenant role SharePoint Administrator (SPAdmin) is required when using this option to allow it to see across all Site Collections.

SharePoint URL:
When connecting to SharePoint the URL should be as follows

Target	Connection URL format
Root Site Collection	https://<tenantname>. sharepoint.com
Specific Site Collection	https://<tenantname>. sharepoint.com/sites/<Site Name>
Team Site	https://<tenantname>/teams/<Site Name>
Tenant	https://<tenantname>-admin.sharepoint.com

SharePoint Authentication Methods

There are several authentication methods that can be used to connect to SharePoint

- **Error! Reference source not found.** (recommended)
- **Error! Reference source not found.**
- **Error! Reference source not found.**

SharePoint Online – App Based Authentication

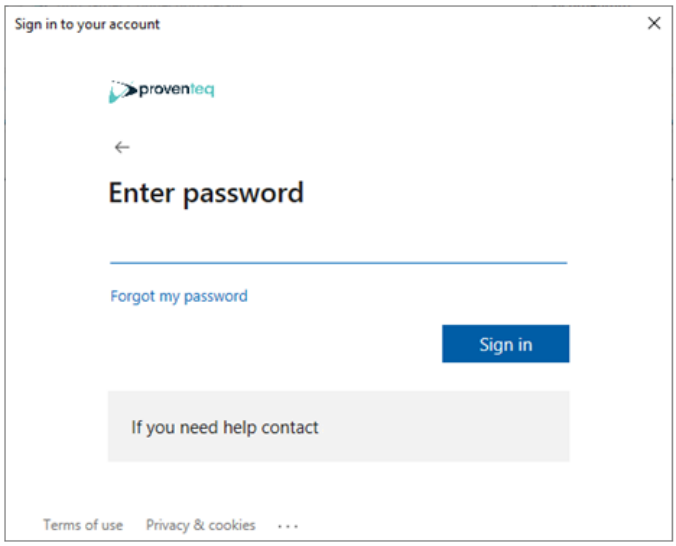
The general recommendation is to use App based Authentication whenever possible.

You will need to provide the URL below to your Global Administrator to provide admin consent and return to the application after they have granted consent.

Standard
https://login.microsoftonline.com/<TenantID>/adminconsent?client_id=d1e5e128-9660-4380-aabb-3e0061c1047c

If using Sensitivity Labels
https://login.microsoftonline.com/<TenantID>/adminconsent?client_id=0a8da96c-bf61-4839-9ede-2d0b6647414c

For information on how to get TenantID for your organisation see
https://docs.microsoft.com/en-us/onedrive/find-your-office-365-tenant-id



If user is Global admin user, then the user will get a consent screen and user will need to accept the consent.

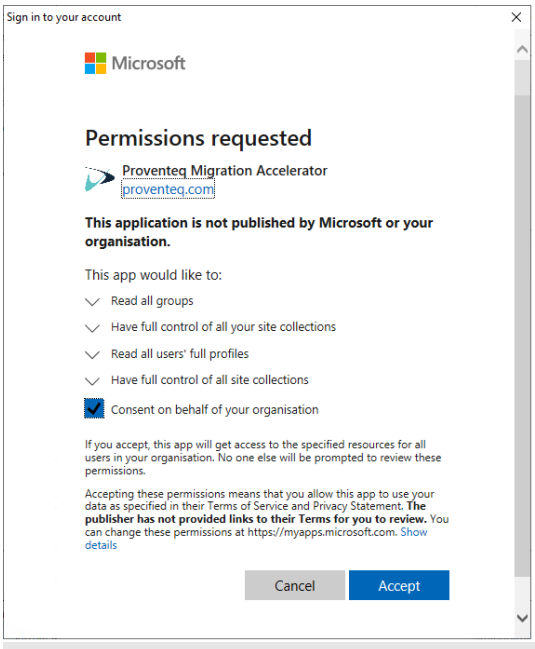
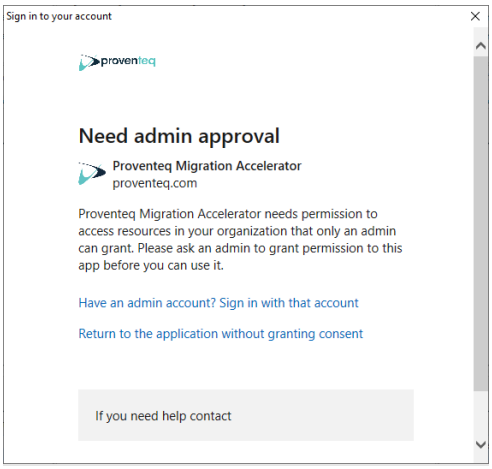


FIGURE 27 - ACCEPTING APP BASED PERMISSION REQUEST

Tick the checkbox for giving the consent on behalf of your organisation and click on the “Accept” button. This will redirect you back to Add Target Connection screen.

If user is not Global Admin, then user will need to request Admin user to login and accept the consent for Proventeq Migration Accelerator application.



Note: To know more about Azure application permissions and admin consent in Microsoft identity platform endpoint. Please refer the below link:

https://docs.microsoft.com/en-us/azure/active-directory/develop/v2-permissions-and-consent

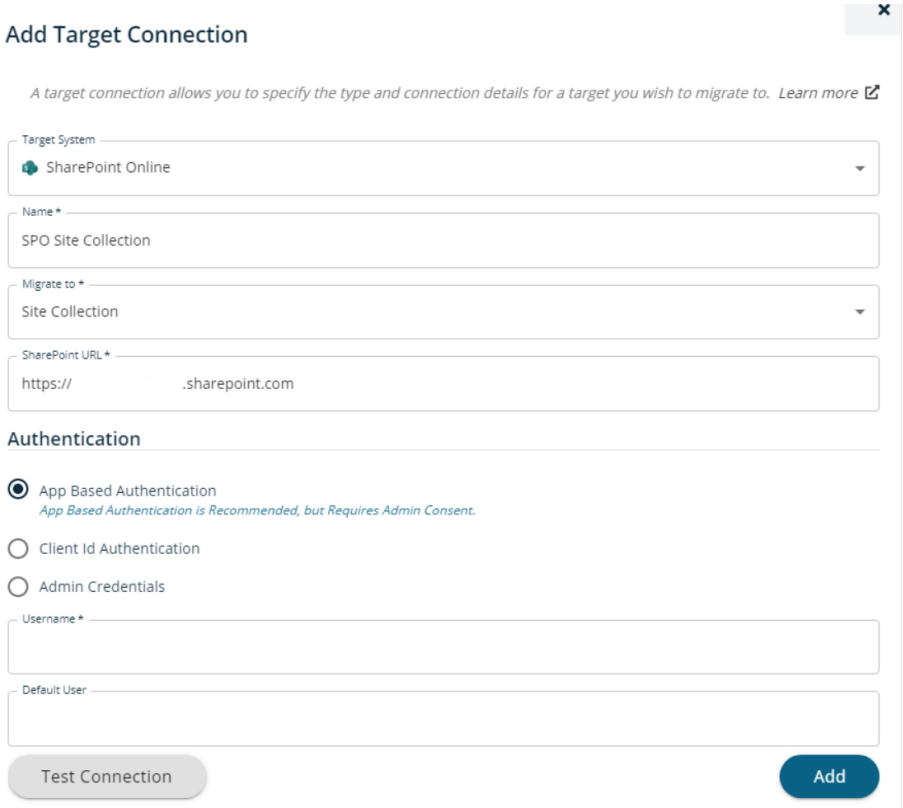


FIGURE 28 - USING APP BASED AUTHENTICATION

App Based Authentication Settings:

- **Username**: The Username for connecting to SharePoint. If connect is to **Tenant** option, then user should be SharePoint Online Admin or Global Admin of the Office365 tenant. If connection is to **Site Collection,** then the user should be a site collection administrator.
- **Password**: The Password for the above-mentioned user.
- **Default User**: The default user to be used for setting Author/Editor fields in the target Site Collection if the source user is not found. This field is optional. See Processing empty source user fields.

Click on the "Test Connection" button, a screen will pop-up to sign into the account of the whitelisted username by Microsoft.

Click on the “Add” button to add the SharePoint connection.

SharePoint Online - Client Id Authentication

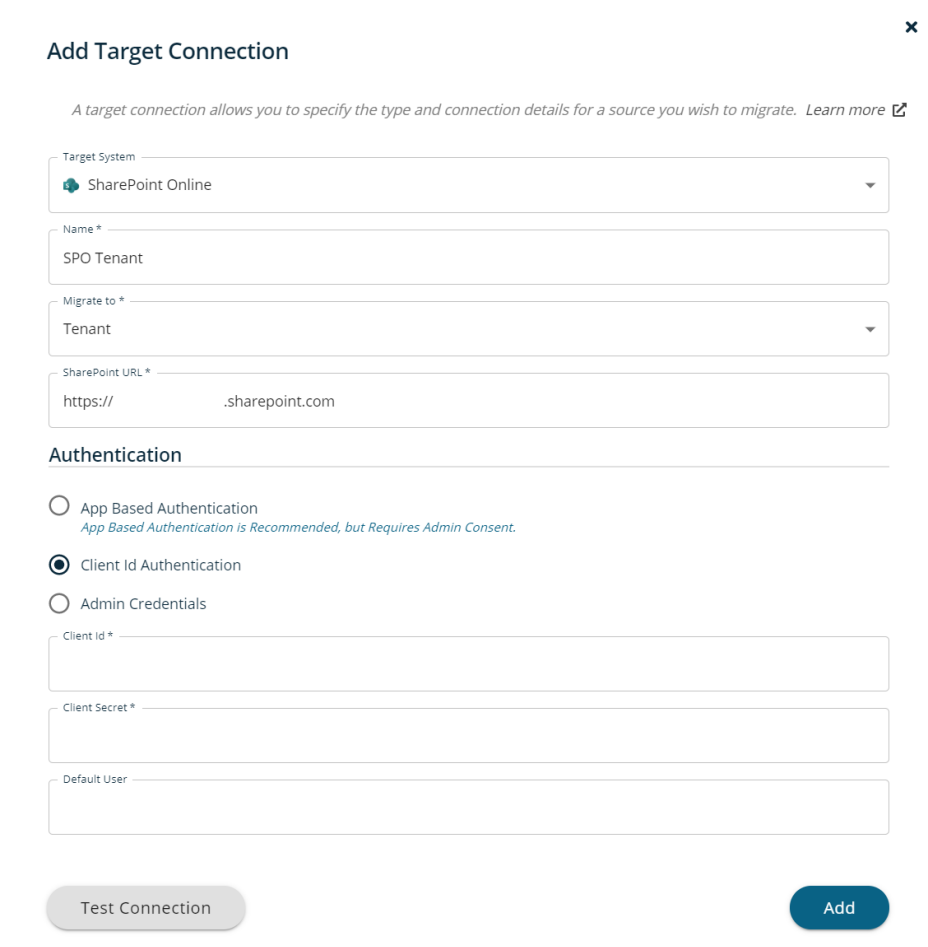


FIGURE 29 - USING CLIENT ID BASED AUTHENICATION

Client Id Authentication Settings:

- **Client ID**: The app-only principal Client ID.
- **Client Secret**: The app-only principal Client Secret.

- **Default User**: The default user to be used for setting Author/Editor fields in the target Site Collection if the source user is not found. This field is optional. See Processing empty source user fields.

Click on the "Test Connection" button to verify that the details entered are correct.

Click on the “Add” button to add the SharePoint connection.

Note: To set up app-only principal Client ID/Client Secret. Please refer the below link:

https://docs.microsoft.com/en-us/sharepoint/dev/solution-guidance/security-apponly-azureacs

SharePoint Online - Admin Credentials Authentication

NOTE: The method of connection below is NOT recommended for production use as it will decrease migration speed due to increased levels of throttling from SharePoint Online.

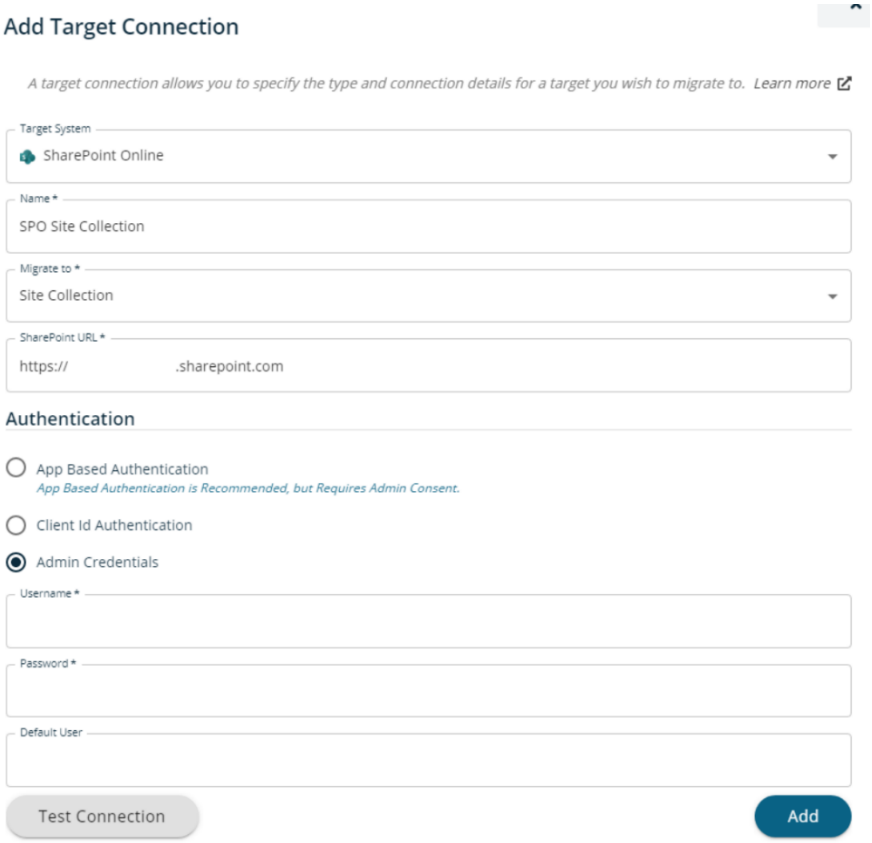


FIGURE 30 - USING ADMIN CREDENTIALS AUTHENICATION

Admin Credentials Authentication Settings:

- **Username**: The Username for connecting to SharePoint. The user should be a site collection administrator for the target Site Collection or SharePoint Online Admin for the tenant level connection.
- **Password**: The Password for the above-mentioned user.

- **Default User**: The default user to be used for setting Author/Editor fields in the target Site Collection if the source user is not found. This field is optional. See Processing empty source user fields

Choosing the Correct APP ID for Authentication

When connecting using this method, it's still necessary to use an Application ID to authenticate. By default this will require the PMA Application ID d1e5e128-9660-4380-aabb-3e0061c1047c to be authorised on the tenant. If using Sensitivity Labels then the Application ID should instead be 0a8da96c-bf61-4839-9ede-2d0b6647414c .

If this is not possible then you may use the Microsoft SPO Management client id instead which is pre-registered on all tenants. This can be done by changing the PmaAppId configuration key as follows:-

```
Add key="SharePoint.PmaAppId" value="9bc3ab49-b65d-410a-85ad-de819febfddc" />
```

Once configured, the connection will work as the Application ID is already approved.

WARNING: USING THIS APP ID MAY CAUSE THROTTLING TO OCCUR SO MIGRATIONS MAY BE SLOWER

Click on the "Test Connection" button to verify that the details entered are correct.

Click on the "Add" button to add the SharePoint connection.

Note: CPS supports migration to GCC (Government Community Cloud). So if the domain is other than ". sharepoint.com" then update the application configuration file typically located here ("C:\Program Files\Proventeq\Proventeq Content Suite\Proventeq.MigrationAccelerator.exe.config"). These keys can be found in the <AppSettings> section.

<add key="SharePointOnlineDomainPart" value=".sharepoint.us" />
<add key="SharepointOnlineAzureEnvironment" value="USGovernment"/>

Select Target Connection Path

Select the root path items should be migrated to. The migration target specified here is combined with information in structure mapping to determine the exact location where items will be migrated to.

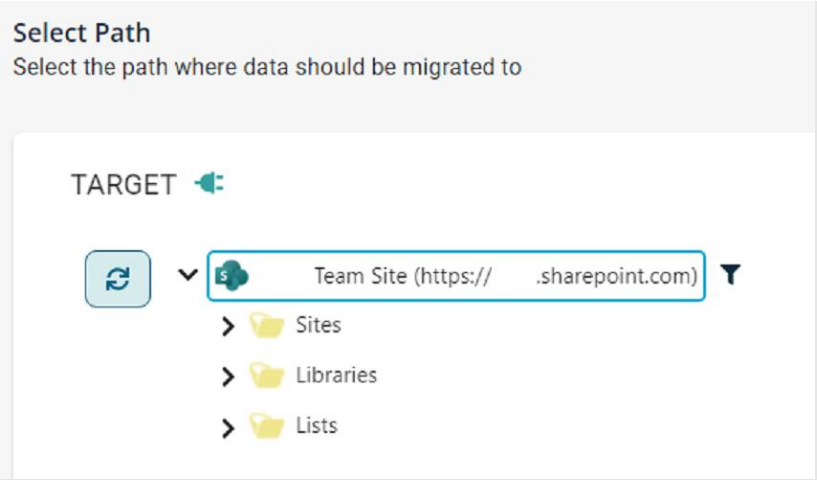


FIGURE 31 - SELECTING TARGET PATH

3.3.2. Configuration when connecting to Tenant

The following are steps to add a SharePoint Online Tenant as a Target. A Tenant level connection is useful if the migration will be to more than one Site Collection.

If the connect is to Tenant then all the Site Collections in that tenant will be listed below.

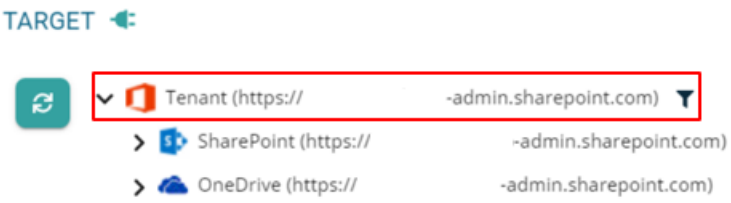


FIGURE 32 - CONNECTING TO TENANT

If the Migration is to OneDrive select OneDrive.

Expand to select a collection site

Click on ▼ to apply a filter to find the site of interest

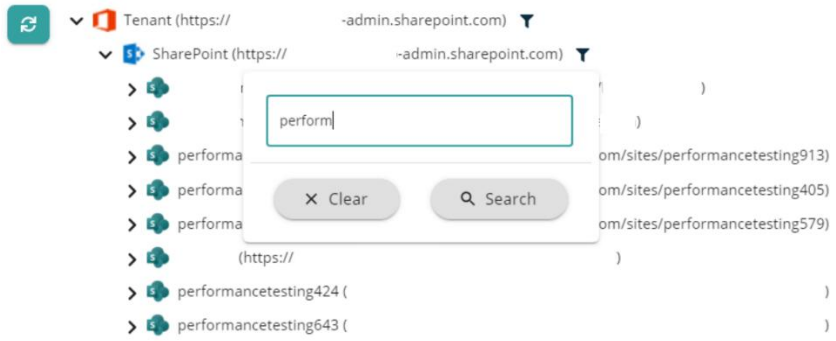


FIGURE 33 - FILTERING SITE COLLECTION LIST

NOTE: The selected item can be a specific folder, Document Library, Site Collection or Site. The higher level the node selected then the more information that will need to be specified later in Structure Mapping to specify exactly where data should be migrated to. For example, if no Library is selected here, then that information will need to be provided in Structure Mapping. See Phase 2 – Structure mapping for more information.

After select the appropriate node, Click **Next** to specify additional job settings.

3.3.3. Adding M-Files Target

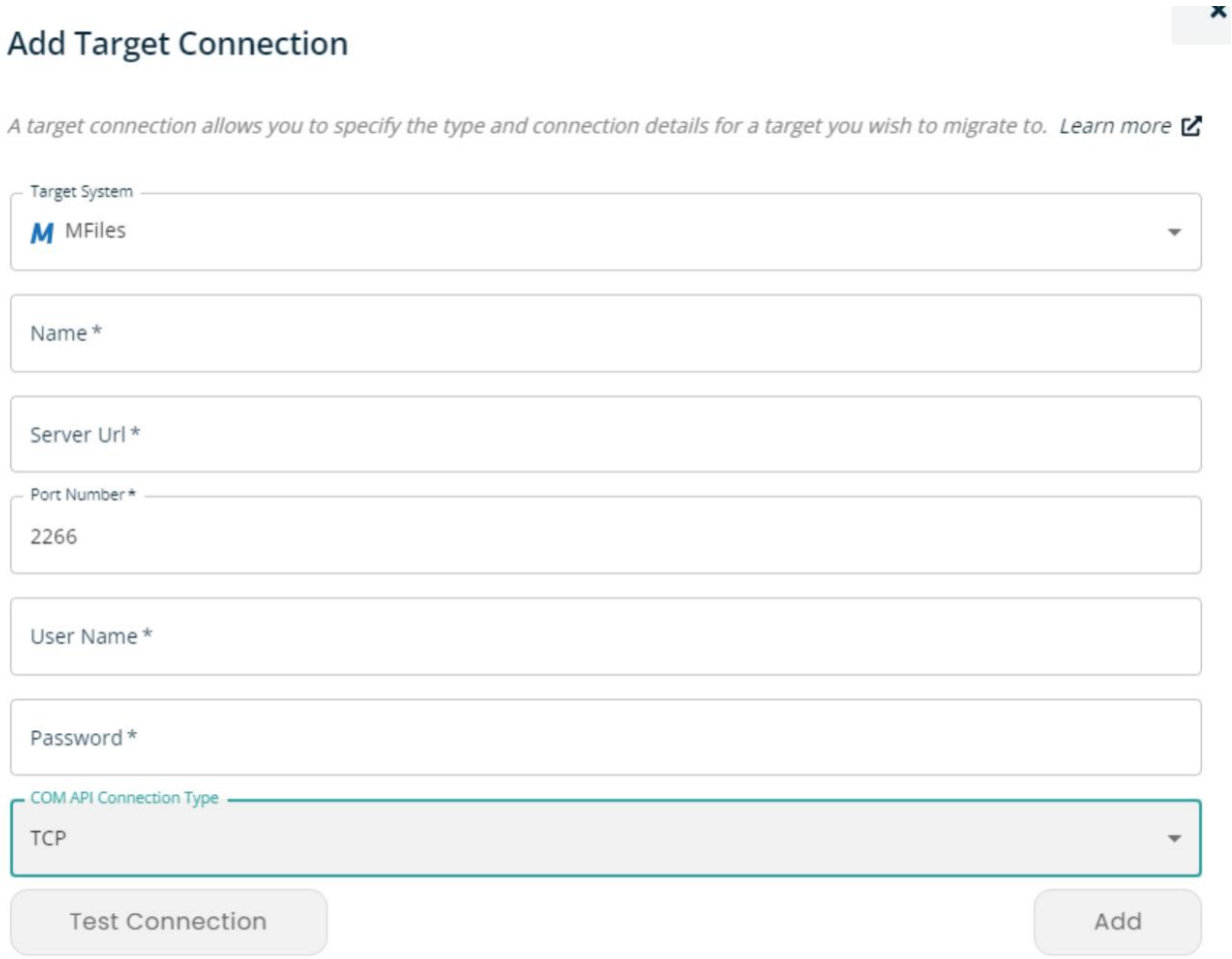
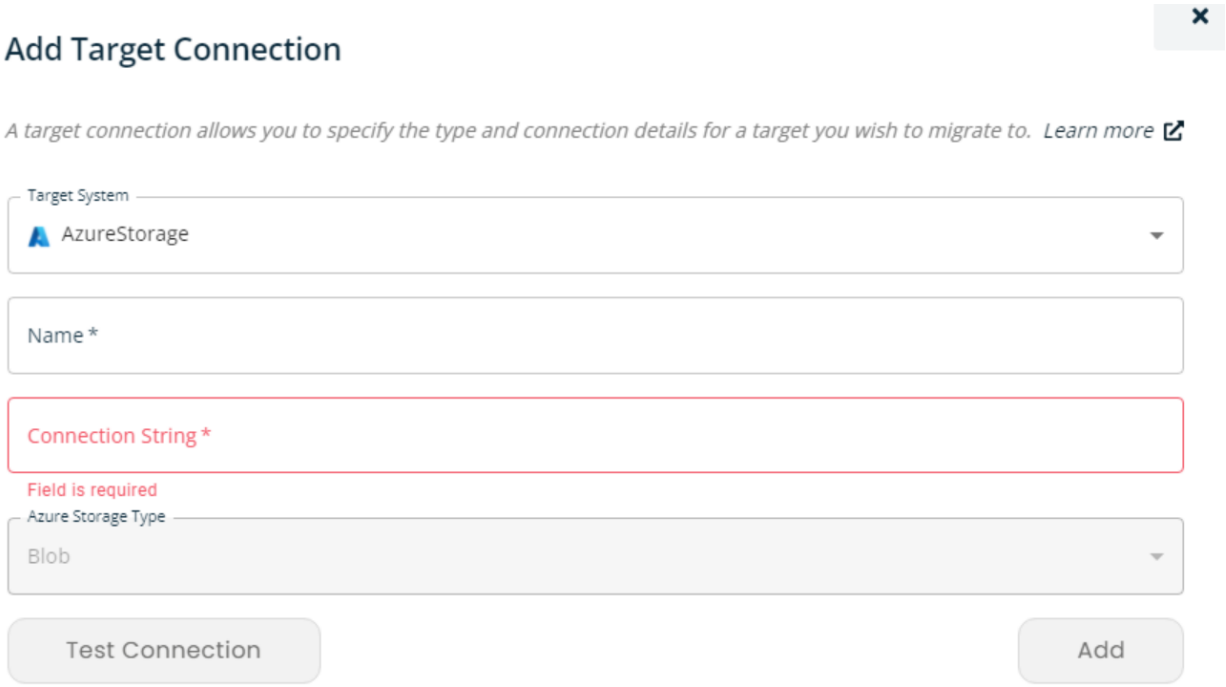


FIGURE 34 - M-FILES TARGET CONNECTION

Name: Name of connection to uniquely identify the connection within the workspace.
ServerUrl: URL of the target M-Files server.
Port Number: Port number for the COM API utilization
Username: Username for connecting to M-Files server. The user should be an administrator for the target M-Files system.
Password: Password for the above-mentioned user.
COM API Connection Type: Connection protocol - HTTPS, TCP or gRPC connection. Selecting gRPC and select port 443 if connecting to M-files cloud.

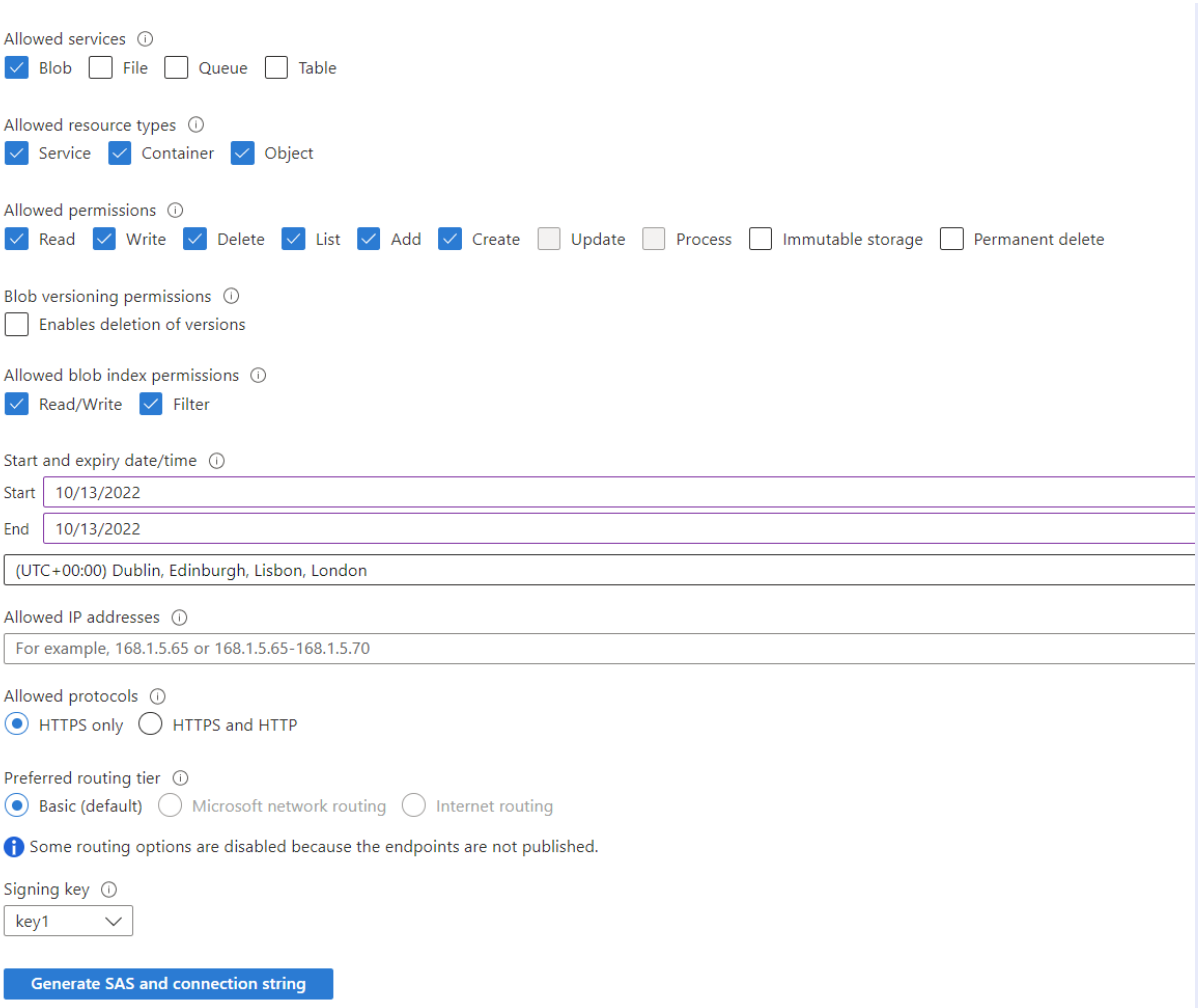
3.3.4. Adding Azure Storage Target



Name: Name of connection to uniquely identify the connection within the workspace.

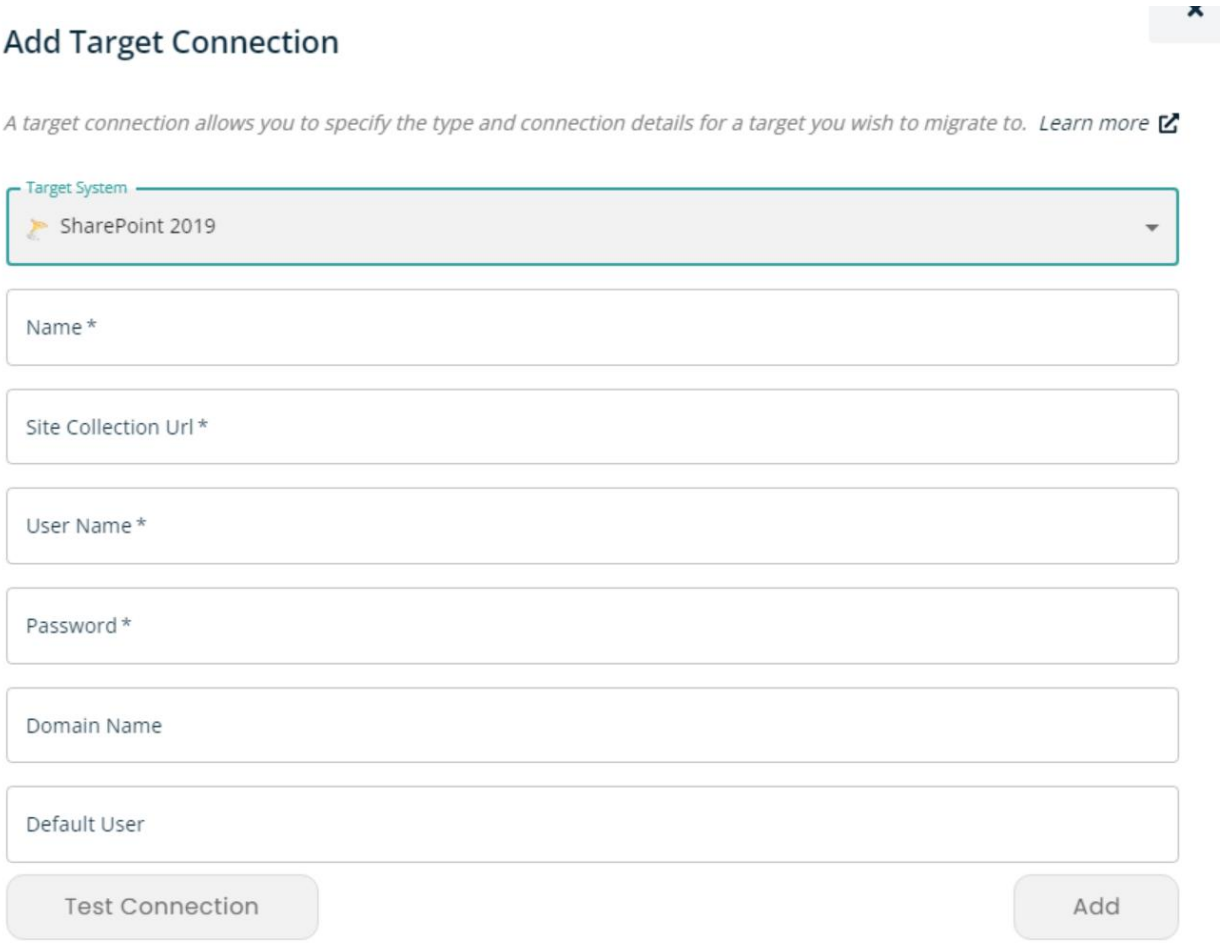
Authentication: There are 3 methods of authentication with AzureStorage. Connection String, Azure AD and Shared Access Token.

Connection String: To configure Azure target connection, you will need to enter a Connection String from a Shared Access Signature with these properties as a minimum:-



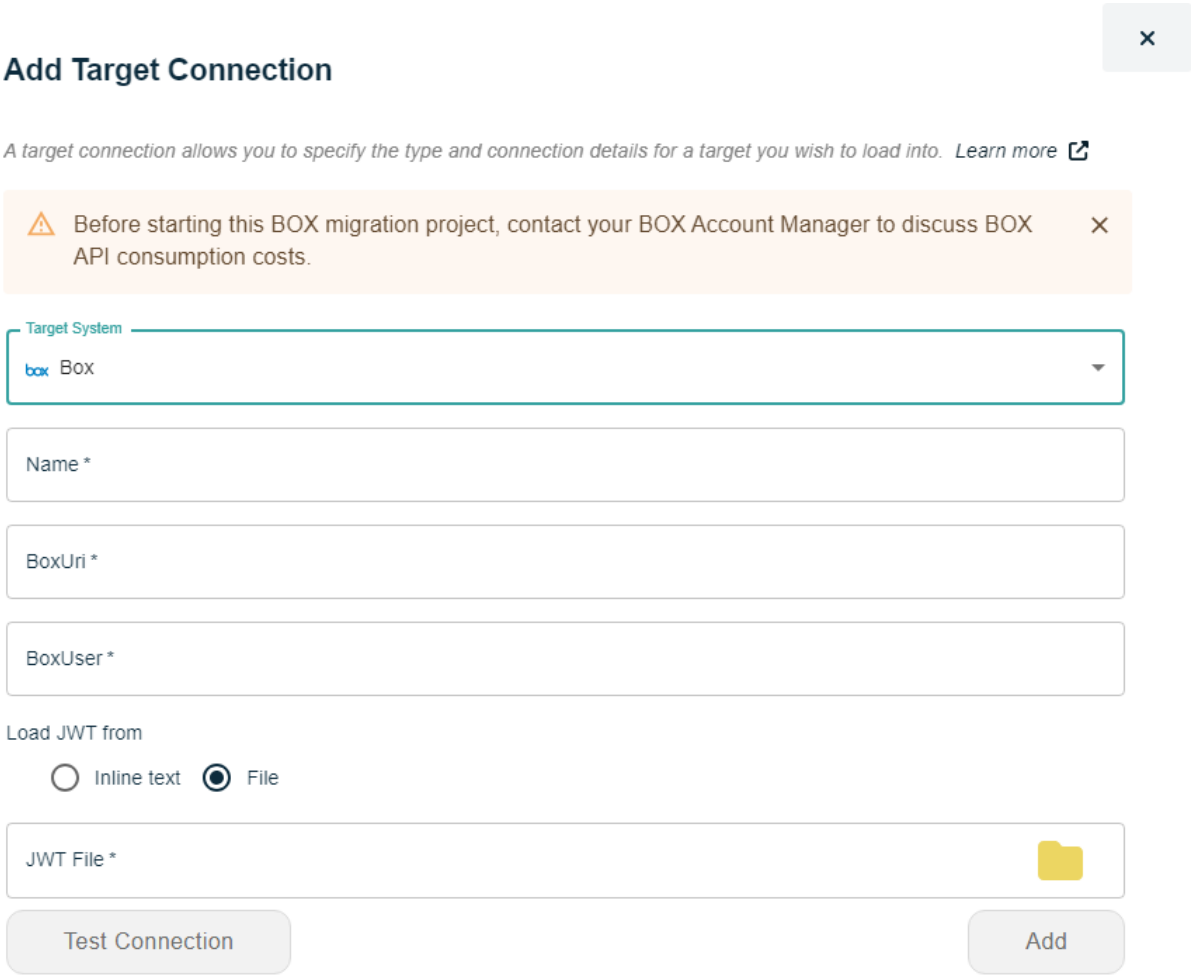
Note: Since blob storage does not support the concept of folders, empty folders will not appear in Azure blob storage.

3.3.5. Adding Sharepoint 2019 Target



Name: Name of connection to uniquely identify the connection within the workspace.
Site Collection URL: Full URL of target site.
Username and password: Account specified requires Admin rights on the site collection.
Domain Name: Windows domain that the SP2019 system is located in.
Default User: The default user to be used for setting Author/Editor fields in the target Site Collection if the source user is not found. This field is optional. See Processing empty source user fields.

3.3.6. Adding BOX Target



Name: Name of connection, to uniquely identify the connection within the workspace.
Box URL: URL of the Box system. Usually this will be https://app.box.com
Box User: Username of Box enterprise admin account
Load JWT from:
The CPS Box connector supports the JWT (JSON web token) method of authentication. The token can either be located in a .JSON file or select 'Inline text' and enter the token in the specified field.
File: Location of a file containing JWT token
Inline text: String containing JWT token

4. MIGRATION GUIDELINES

Before a migration begins it is critical to understand the data within the existing environment and how it should be represented in the target. This is part of the migration planning process and CPS provides features that allow you to view and then define how this data is migrated and should be presented in the target.

> **NOTE:** Throughout this document the term **Migration** is used. In-fact, Migration of data is always carried out by copying of the data. The process of migrating data to the target does not modify or delete the source data.
>
> **NOTE:** Information on software and hardware pre-requisites and product installation is contained in the Proventeq Content Productivity Suite Installation Guide. This guide presumes the software installation is complete.

Project Planning

4.1 Proventeq Migration Accelerator is a sophisticated migration product, so it is important to plan the migration to ensure that the migration runs smoothly.

Questions to ask when planning for migration include:

- What systems to migrate, their physical location and how much of the data is to be migrated.
- What requirements/expectations are there regarding project milestones, migration completion.
- How to split the migration to make it manageable.
- Differences between the Source and Target in terms of metadata, users, security, and folder structure.
- Should the existing folder structure remain or be changed.
- Do existing access permissions need to be carried over to the target.
- How will end-users be affected, communication plans for end user.
- Source system readiness – Ensuring any documents are 'check-in' prior to discovery, that source system is in a good state of health.
- Target system readiness.
- What to do with the Source once migration has been completed.

When starting a migration project, the end-users are a vital part of a successful project. Communication is important and end-users need to know.

- Why the content is being migrated
- What the benefits will be to them
- When the migration is going to happen
- What the expectations are upon them

Any upfront communication also allows users to do any clean up prior to the migration and as a result less data will need to be migrated. End-users will also need to know what the new environment will be, how it will be structured and how to access their data after the migration. Some training may be required for them to take full advantage of the new environment.

> NOTE: It's important to understand that by default the location data being migrated to is new and so users should not be working in that area, potentially manipulating folders or documents in locations that are actively being migrated to. If it's essential for this to be allowed, contact technical support to understand ramifications and potential workarounds.

4.2 Migration Terms and Concepts

This section describes important terms and concepts that are used throughout this guide. It's strongly recommended that you read and understand this section before reviewing the rest of the User Guide.

Staging Database

This is the SQL database that gets created during the product installation. It holds all the configuration information that will be specified as well as the metadata of every source item and their status.

Source

The Source is where items are to be migrated from.

Target

The Target is where items are going to be migrated to.

Containers

This term is used to describe a object which can contain items. So a Folder, a Document Library, cabinet, workspace are all differernt types of Containers.

Migration Jobs

Migration jobs are the basic framework for a migration.

A Migration Job consists of: -

- Details of Source and Target
- Settings to determine how source data should transferred in the target
- Settings to control the Migration.

All operations run within the context of a Migration job. Migration projects are generally split into multiple migration jobs. The order in which the migration jobs get processed depends upon criteria such as data volume and business priorities etc.

Discovery

Discovery runs in the context of a job and is the first operation that needs to be run after a job is created. During Discovery, information on items, folder/container structures, types of content and security information is collected from the source specified in the current Job and stored in the CPS staging database. Item contents are **not** captured during discovery.

The items discovered can be seen using the **Items List** feature.

Depending upon the size of the environment, Discovery could take several hours or more to complete.

Pre-Checks

The Pre-Checks feature is set of configurable, pre-defined checks that are run against the information gathered by the Discovery process. It can identify many potential issues that may prevent a succesful migration into the target, looking at things like target system restrictions, limitations, empty files and duplicate file names. The result of running Pre-check is a PowerBI report where issues can be reviewed.

Example of a Power BI report after running **Pre-Checks**.

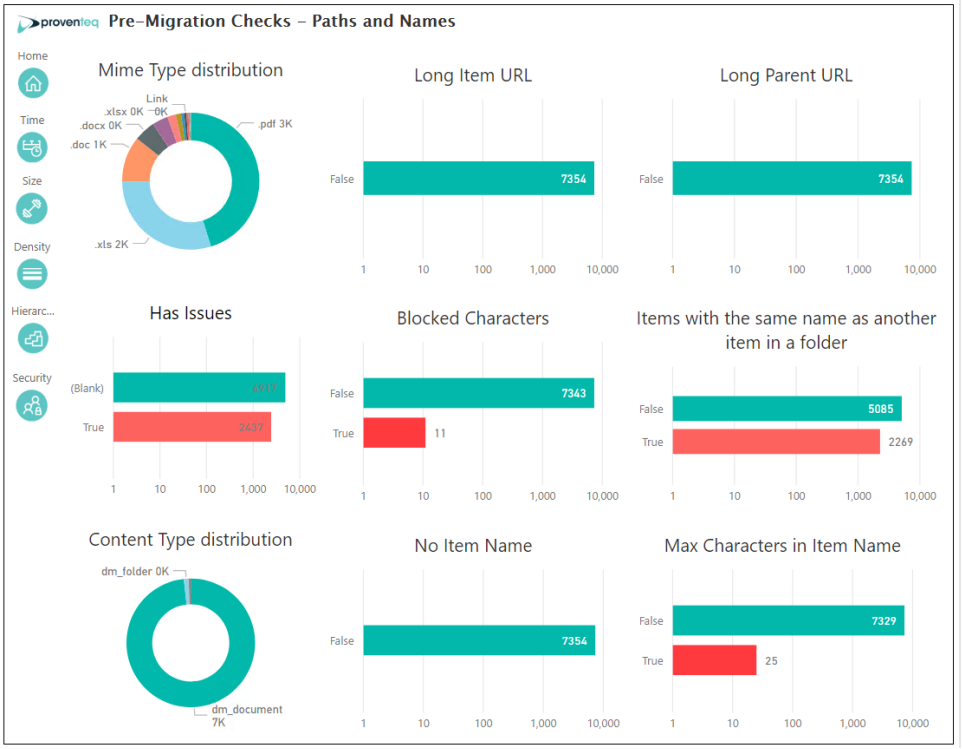


FIGURE 35 - SAMPLE POWER BI REPORT

You can also view the above data using the Pre-Check report.

Migration Settings

These settings control several features related to how data is migrated.

The **Structure Mapping** feature gives control over how the folders/containers in the source should be represented in the target. It provides the ability to simply use the existing hierarchy structure or restructure the folders as part of the migration.

Content Types Terms

Metadata is the term for properties of an item, for example Author or Created Date.

Content Type – A content management system will contain many types of items such as document, folders, invoices etc. These are known as 'Content Types'. In some systems these may also be knows as Document Types/Classes/Categories or Object Types. There are both Source Content Types (what the source system supports) and Target Content Types (what the target system supports). Each content type may have different metadata fields associated with it.

The **Content Type Mapping** feature controls how properties of a particular Source Content Type are mapped to a specified Target Content Type. For example, when an item of Source Content Type 'Invoice' is migrated, an item of type 'Legacy Invoice'(as defined by Content Type Mapping settings) should be created in the target and the metadata mapping information(see below) will be used to determine how metadata properties in the target item should be populated. CPS ships with very useful defaults mappings and these can be customised as required.

The **Metadata Mapping** feature controls how each target metadata property is populated for each target content type e.g. how the 'Description' or 'Amount' metadata properties is populated for the 'Legacy Invoice' Target Content Type.

CPS ships with very useful content type mapping defaults and applies these (this is termed **Automated Content Type Mapping)**. They can be customised as required.

Custom Classification is an advanced feature that allows items in source to be mapped and processed based upon their metadata properties, overriding the content type assigned in source. See Custom Classification for more information.

Security Mapping is used to define how Users, Groups and permissions types from the source should be represented on the item in the target. For example if the Created By of a source document is "Percy Shelley" then in the target the Created By should be Percy.Shelley@acme.com (because in SharePoint this field is required to be an smtp address).

Shortcuts Links

These are items in the source system which are pointers to an item in another location. Some platforms support these calling them Shortcuts or Links. In some occasions there may even be several of these links pointing to a single source document, this is known as multi-filing.

> Note: The use of shortcuts in SharePoint is generally not recommended.

Embedded Links

These are pointers contained within documents that link to another item. e.g. a Word document stored in Documentum may contain an Embedded Link to another Documentum item.

> Note: Both the shortcut links and embedded links features will impact migration speed if enabled.

Migration

Migration is a set of individual steps listed below to prepare and then load items into the target. These sequential steps can be executed individually so that in the early phase of the migration project the output of each can be verified before executing subsequent steps.

The **Item Manifest** is a data structure that contains all the information about an item to be migrated, excluding the item content itself. It will contain the location where an item should be migrated to, all the values for the target metadata fields and so on. Steps 1-2 provide the information for Step 3 which creates this structure for each item.

The processing steps are sequential, so an item will first be processed by Step 1, then Step 2 and so on.

Step 1 Extract – Captures metadata for each source item into the CPS Staging database.

Step 2 Classify – Determine what the target content type of the item should be based upon the extracted meta data from step 1,or custom classification rules and source->target content type mapping configuration.

Step 3 Transform – Uses the output from Step 2 as well as information from structure, content type and security settings to create a data structure for each item called the Item Manifest.

Step 4 Load – Retrieves the item content from the source and combines it with the Item Manifest from Step 3 and uploads this into the target.

Packaging

To improve migration speed, items are 'packaged' before uploading to SharePoint.

> Note: During the pilot phase of the migration, or after reconfiguration, it is strongly recommended to run each of the 4 steps above seperately and the results reviewed before executing the next step.

Summary of Migration Process

A migration is a project where several **Migration Jobs** are created, each specifying details of a **Source Connection** and a **Target Connection.**

4.3. In the **Discovery Phase,** the **Discovery process** is run for each job tobuild an inventory of items in the source and gather basic details of each. The details gathered are visible using the **Items List** feature and can be analyzed using the **Pre-Checks** feature to identify potential issues (as shown in the **Power BI** reports). The **Power BI** reports also displays lots of other useful information to help the customer understand the content to be migrated.

In the **Migration Preparation Phase**, the user can review existing **Folder Structures, Content Types** and **Security Information**. Using an understanding of business requirements and the environment, **Stucture Mapping** can be used to define what the target folder hierarchy should look like. **Content Mapping** can be used to control how the **metadata** fields in source items should be moved to the target and finally, **Security Mapping** used to specify how user and security information in the source should be transferred to the Target.

Once these are complete, the **Migration Phase** can begin. Items are processed based upon these settings above to create an **Item Manifest** for each item which in the **Load** step is combined with the retrieved item content and uploaded to the **Target.** During the **Load** step, if the option **Update Embedded Links** is select then these will be updated in Items as they are migrated.

If there is a requirement to migrate shortcuts then the **Shortcut Links Migration** features can be used to migrate these after all items have been successfully migrated.

4.4.

Verify the Migration

Like any data migration, it's critical to ensure that data has been migrated as expected. Special attention needs to be paid during the pilot phase, reviewing the processing of different content types, all of the source meta data are migrated as required, items are migrated to the correct location etc.

4.5. As with any migration, a comprehensive User Acceptance Testing (UAT) plan must be in place to ensure the business users are satisfied there are no issues and the migration is meeting the business requirements.

Communicate with End-Users

End- users need to be aware that their data is being migrated and once the data has been migrated they need to check that they can access the data. It is worth considering providing training for End-Users so that they can learn the new features of their environement. Take into account that End-Users may be away for an extended period of time, for example maternity leave or long term ill health. Eventually you will need to decide what to do with the original Source environment.

4.6. High-Level Migration Plan

- **Configuration**
  - Apply CPS License(s).
  - Create Migration Job(s)
    - Specify details of source.
    - Specify details of target.

  Both will require permissions to access the source/target.

- **Phase 1 - Discovery**
  - Run discovery
  - Run pre-checks, review output and take corrective action were necessary.
  - Understand existing source folder structures, content types and security.
  - Using understanding of source, have detailed discussions with the customer regarding metadata. What metadata is on each content type, if and how that should be represented in the target.
  - Discuss with customer, requirements for target folder hierarchy structure.
  - Understand data volumes in various locations to help develop plan for order of data migration.
  - Using understanding of existing data and business migration requirements go to phase 2 to apply these.
  - Work with the business to develop UAT program.

  Based upon the outcome of discovery, you may then use this to decide overall migration architecture i.e. The numbers of application servers etc. that are required.

- **Phase 2 - Migration Preparation**
  - Structure Mapping - Choose appropriate Structure mappings option and create structure mapping entries if necessary. Some configurations of the target system may also be required.
  - Content Type Mapping- Understanding existing content in detail and how it should be reflected in the target.
  - Metadata – For each content type, what metadata properties are used currently; what and how should these fields be moved into the target.
  - Security Mapping - Setup user, group, and permission mapping to assign new values to existing user and groups etc. associated with the existing data in the source.
  - Review migration options, in particular Embedded Link resolution.

To ensure the migration runs smoothly and will meet business needs, it's **strongly recommended** to review and get 'sign off' on these data migration settings with the appropriate parties.

- **Phase 3 - Migration**

Note: Executing a step will also execute the previous steps if not already run.

- Execute **Step 1- Extract**, review extracted meta data from source system and ensure it is correct.
- Execute **Step 2 – Classify**, review content types that have been assigned to items.
- Execute **Step 3 – Transform**, this will use the information in structure mapping, content types etc. and create a package of data for each item (terms an item manifest) which should be reviewed and will then be used by step 4.
- Execute **Step 4 – Load,** will use the item manifests created by step 3, retrieve the actual item from the content source and load it into the target.
- Migrating Shortcuts Links- After migration, execute the Shortcut Links Migration which will migrate any shortcuts in the source system into the target. Note: It's not recommended to continue to use shortcuts in SharePoint.

- **Verification -** After migration, review the migrated data to ensure
  - Migrated to correct location
  - For each source Content type, meta data and item content are correct
  - Security information is correct
  - Links between items are working (if option to convert was selected)
  - Shortcuts are recreated and working (if option to migrate was selected)

5. MIGRATION JOB CREATION

A migration project is typically broken down into one or more Migration Jobs. For each job, the source system and source path, together with the target system and target path is specified. By selecting a particular job within the client, all other operations work in the context of those settings. It's possible to switch between migration jobs in the client but only one operation can be executing at any one time on the server.

Often several jobs are created even if there is just one source and one target. In doing so it's possible to break down the overall migration into several manageable chunks by specifying specific source paths for migration or other filter criteria.

Depending upon the project, several CPS servers may be required. It is recommended that the Source and the CPS server(s) should be in the same location (i.e. to keep latency as low as possible).

> NOTE: If using several CPS servers together with embedded link resolution or shortcut link migration features, ensure such links are not across CPS systems.

Migration Job creation is simple using the job creating wizard which is available using the  Migration icon on the Navigation Panel.

5.1.1. Create Migration Job

To manage jobs, click on  in the Navigation Panel. New jobs can be created, and the properties of existing jobs can be viewed.

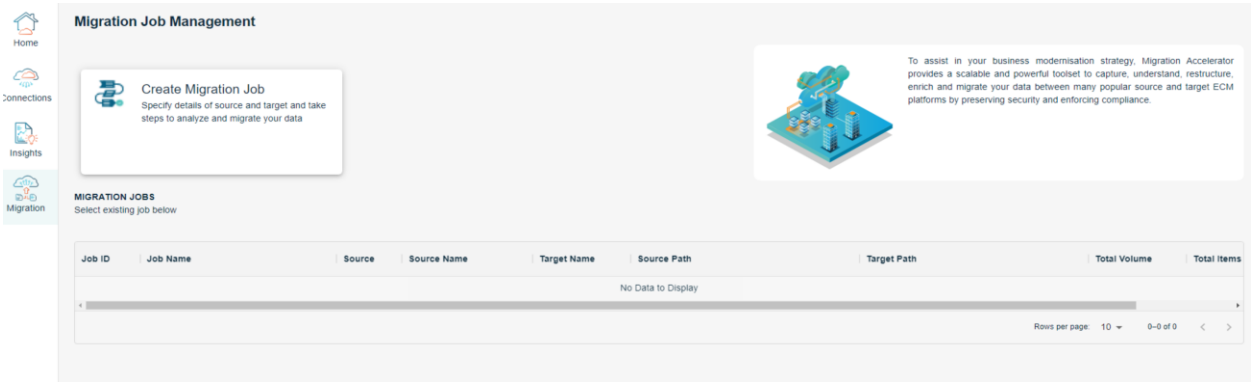


FIGURE 36 - MIGRATION JOB MANAGEMENT

Once click on create migration job, it will redirect to introduction page

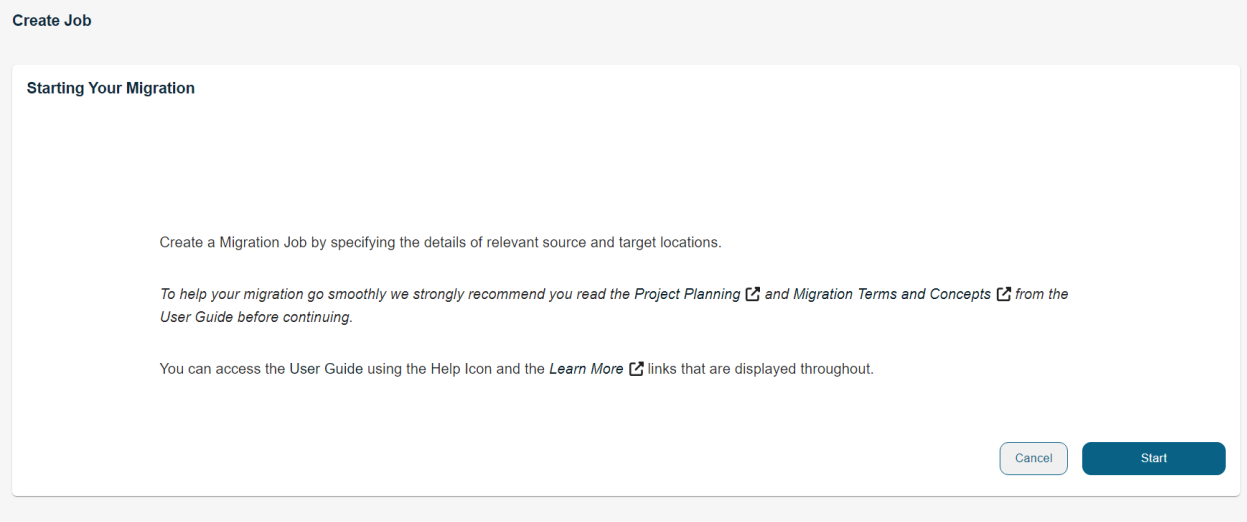


FIGURE 37 - CREATE MIGRATION JOB

This will be a information page about migration. Click on Start will be navigated to next screen of job creation.

5.1.2. Select Source Connection

This screen will show all existing source connections added using Connection modules. In case not added source connections from that module, there will be a Add New connection button to add relevant source connections.

From the list of connections, choose relevant source connections using radio button and click on Next button to move on next screen to select relevant location.

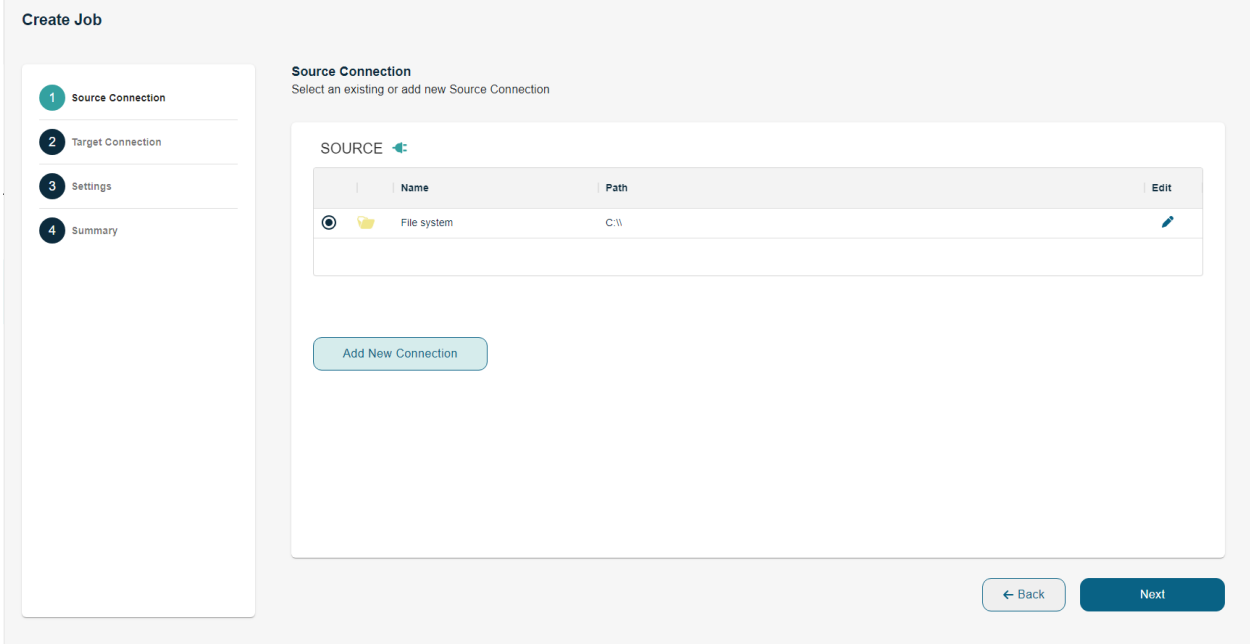


FIGURE 38 - SELECT SOURCE CONNECTION

5.1.3. Select Source Path

On this screen, select root node if want to create job from there or expand the root node to select any specific folder to create a job from it.

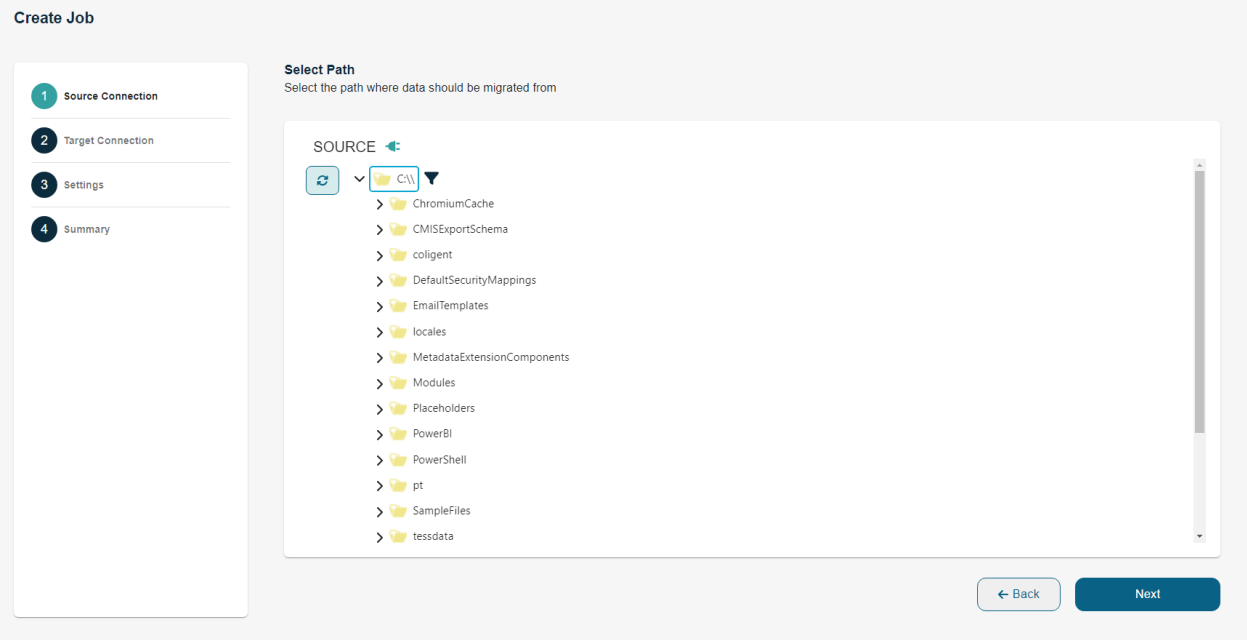


FIGURE 39 - SELECT SOURCE PATH

After selecting relevant location, click on Next button to move on screen to select target connection.

5.1.4. Select Target Connection

This screen will show all existing Target connections added using Connection modules. In case not added source connections from that module, there will be a Add New connection button to add relevant target connections.

From the list of connections, choose relevant target connections using radio button and click on Next button to move on next screen to select relevant location.

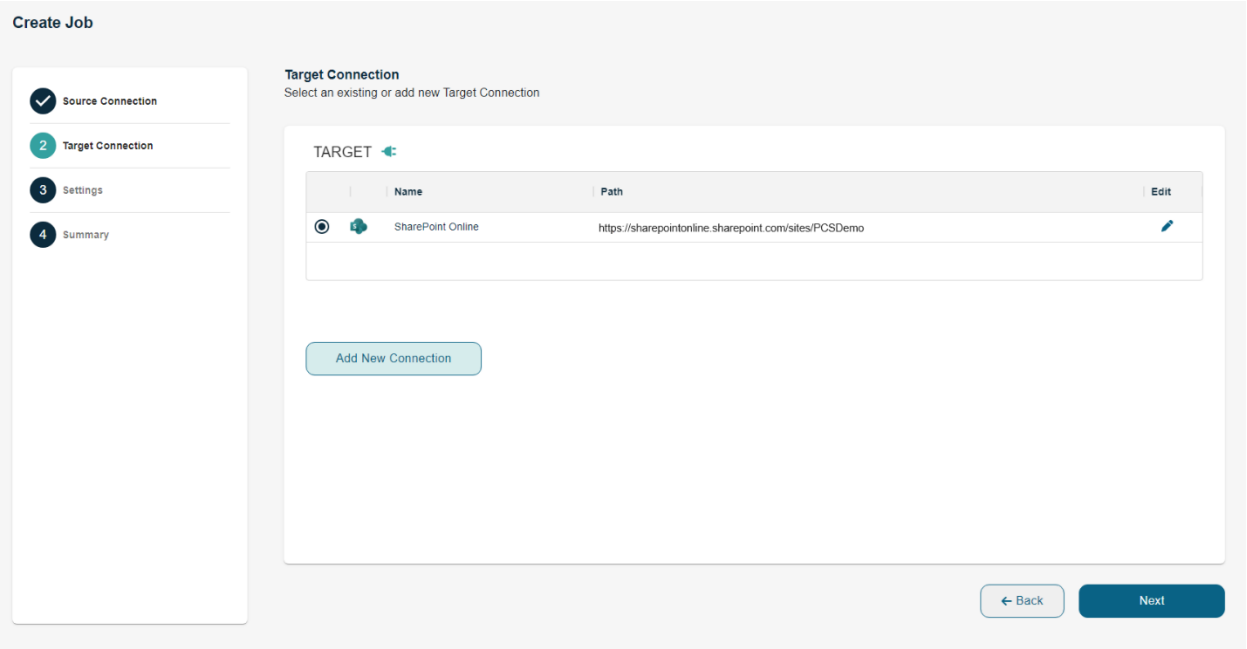


FIGURE 40 - SELECT TARGET CONNECTION

5.1.5. Select Target Path

On this screen, select root node if want to create job from there or expand the root node to select any specific location to create a job from it.

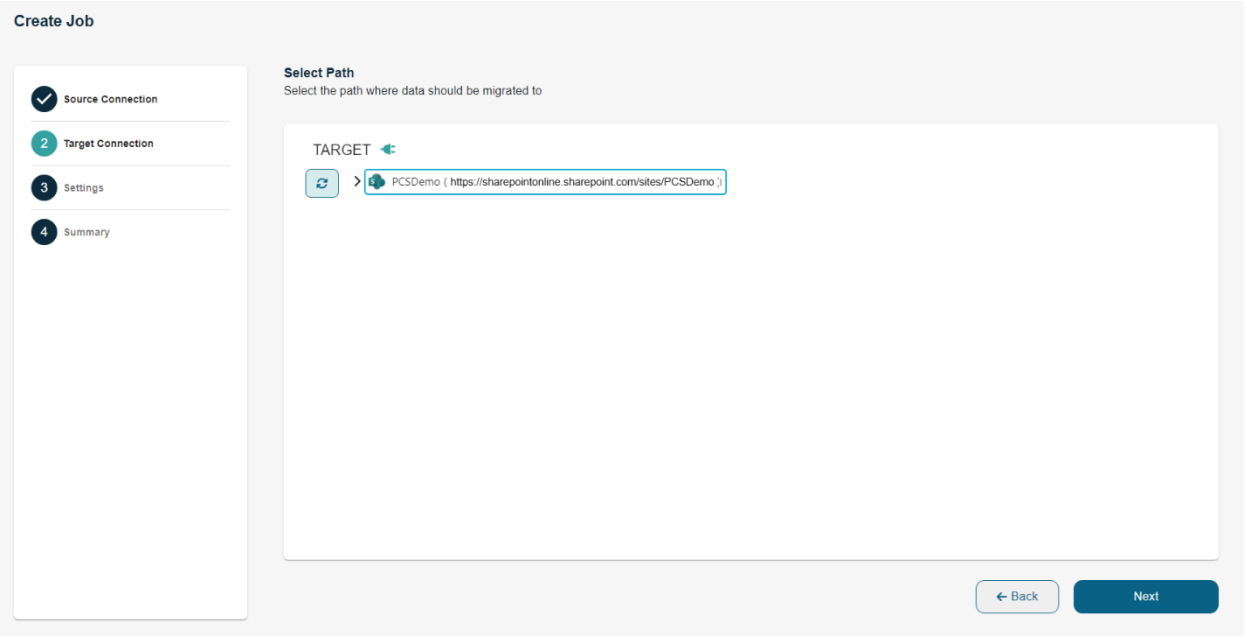


FIGURE 41 - SELECT TARGET PATH

After selecting relevant location, click on Next button to move on screen to choose settings for job creation.

5.1.6. Settings

On Settings screen, there are multiple options for choose settings.

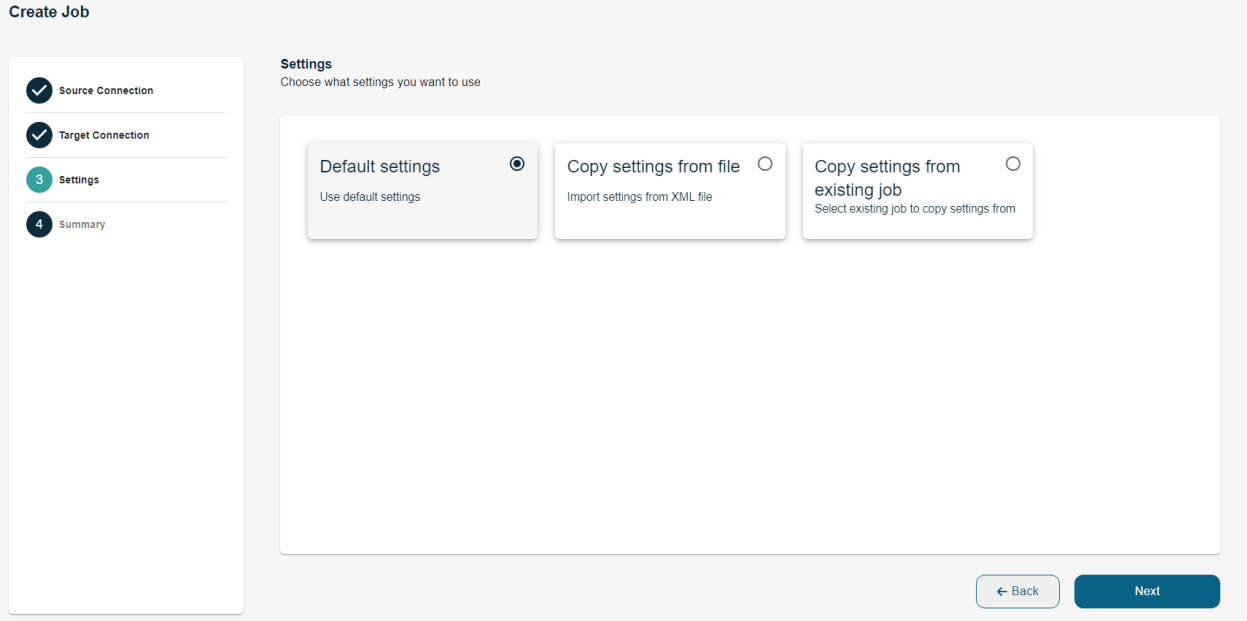


FIGURE 42 - SETTINGS

To simplify job creation, settings from other jobs can be copied and applied to this job. These settings are the Structure, Content Type and Security mapping. The following options are available: -

- **Default Settings**

  Job is created with default settings, and no information copied from existing jobs. No structure mapping settings or security mapping and only standard Content Type assignments will be configured.

- **Copy settings from file**

  In Create Migration Job, the Content Type Mappings and General settings from a job can be exported to files. Using this Copy settings option allows these to be imported. This is especially useful for copying these settings across CPS servers. Custom classifications and security/permission information is NOT copied.

NOTE: This option can only be used between jobs that have the same source and target. i.e. It's not possible to create a job migrating from File Share and use those settings to create a job migrating from Documentum.

- **Copy settings from existing job**

Allows the user to select an existing job to copy Content Type Mappings, Custom Classification and General settings from another job. It will NOT copy any custom content type settings as these are not transferrable between jobs. This option will only be displayed if other jobs have been created.

NOTE: This option will display only if at least one job present in application.

After selecting relevant setting, click on Next button to move on Jon summary screen

5.1.7. Job Summary

Before creating the job, review the details shown on the Job Summary screen and alter the job name if necessary.

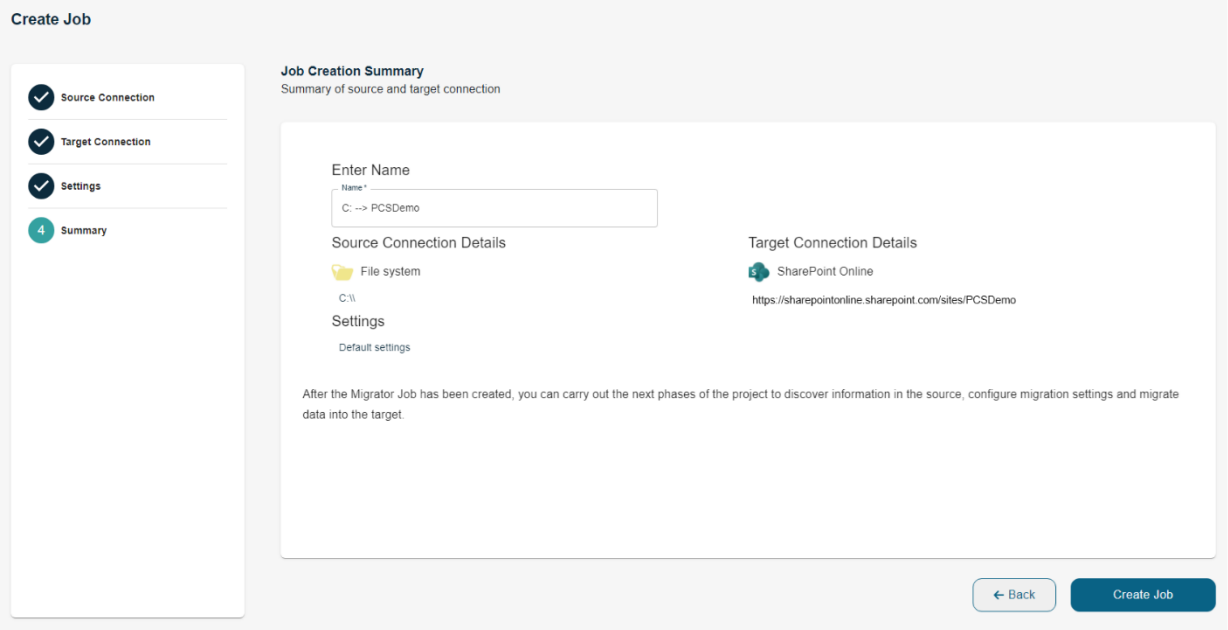
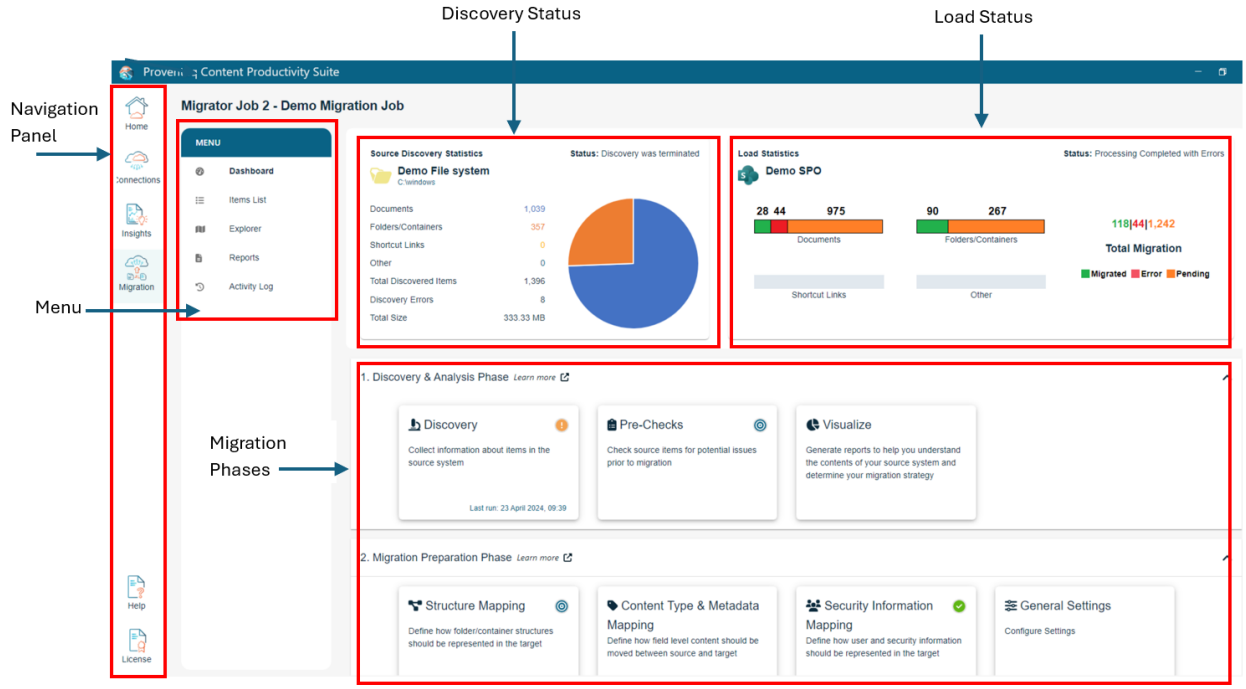


FIGURE 43 - JOB SUMMARY

Clicking **Create Job** will create the job and the Migration Dashboard will be displayed.

6. MIGRATION JOB DASHBOARD

The Migration Job Dashboard allows access to configuration and status information and execution of all migration operations. It is the main screen displayed when a job is selected from Create Migration Job and is also displayed after a job is created using the Migration Job Wizard.



6.1. FIGURE 44 - MIGRATION JOB DASHBOARD

Discovery Status

This section displays information gathered during the Discovery process.

Folders/Containers	Number of folders/containers found
Documents	Number of documents found
Links	Number of shortcuts links found
Other	Any item types not covered by the above types e.g. ShareFile Virtual Collections
Total Discovered Items	Number of items successfully captured during discovery (excludes discovery errors)
Discovery Errors	Number of errors during discovery
Total Size	Total size of items discovered (includes all versions of the items)
Status	Current status of Discovery

Load Statistics

This section is updated when the **Step 4 Load – Migrate to Target** operation is executed to migrate items into the target.

Documents – Broken down by Migrated, Errored and Pending
Folders – Broken down by Migrated, Errored and Pending

Migration Status	Status from the last migration run
Documents*	Broken down by Migrated, Error and Pending.
Folders*	Broken down by Migrated, Error and Pending
Links	Shortcut links broken down by Migrated, Error and Pending
Total Migration	
Migrated	Total Number of documents and folders that have been migrated
Error	Number of items that have failed migration
Pending	Items yet to be migrated

*When Include Total versions is not selected, if versions of the same item have different processing states it is represented as follows. Any version with Error will count as Error, otherwise any version with Pending will count as Pending, otherwise all versions have to be migrated to count as Migrated.

6.3. Migration Phases

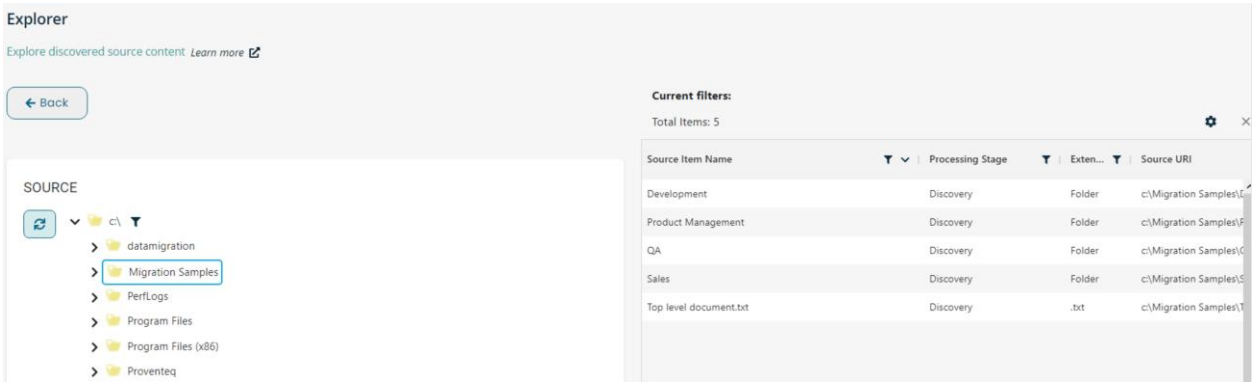
The dashboard displays options organised into the three project phases which control the migration. These will be described in later section.

- Discovery and Analysis
- Migration Preparation
- Migration

6.4 For information on other elements of the Migration Dashboard see below.

Explorer

The explorer feature allows the information gathered during discovery to beyasily understood by presenting it as a folder hierarchy which can be expanded to list folders and items.



6.5. Reporting

The reporting option provides access to the following reports.

- Discovery Report
- Embedded Links Report
- Pre-check Report
- Migrated Items Report
- Failed Items Report
- Migration Audit Report
- Performance Report

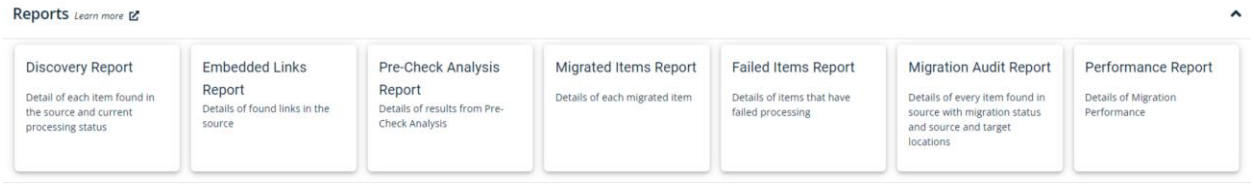


FIGURE 45 - CPS REPORTS LIST

When a report is displayed the list of reports minimises but the expand open can be used to display them again.

6.5.1. Discovery Report

Displays a list of all items that have been discovered in the source along with additional metadata information and overall status of the item. It's important to review this for correctness and completeness to ensure subsequent steps in the migration process proceed on a sound footing.

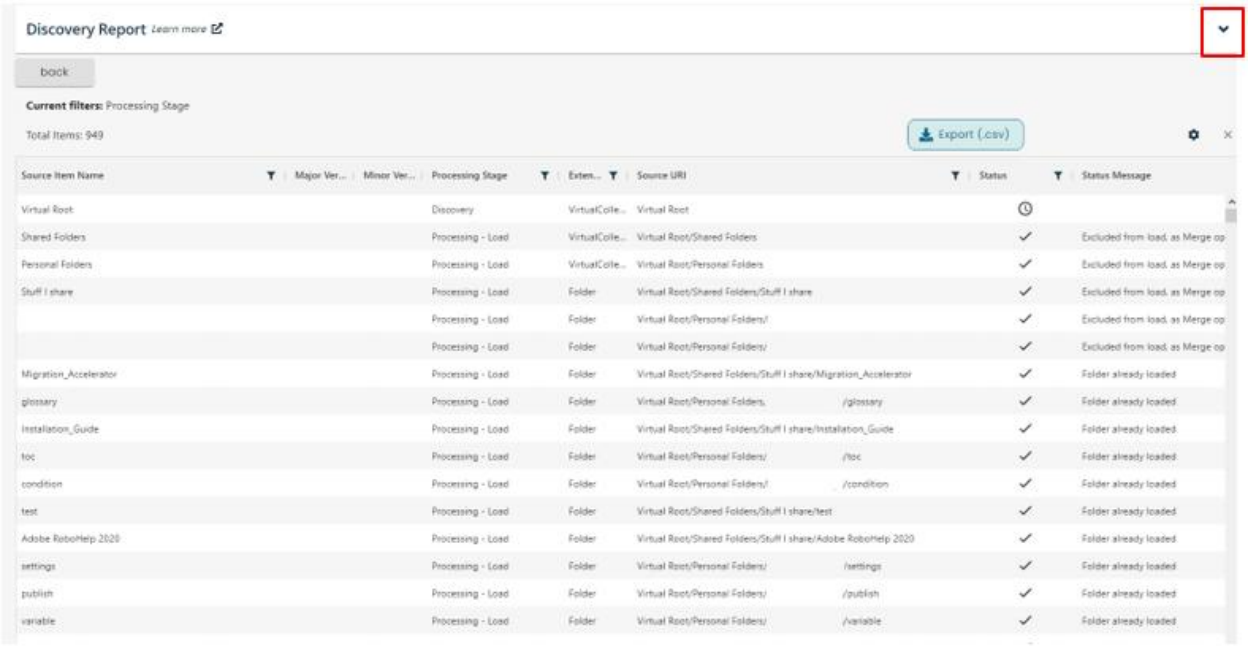
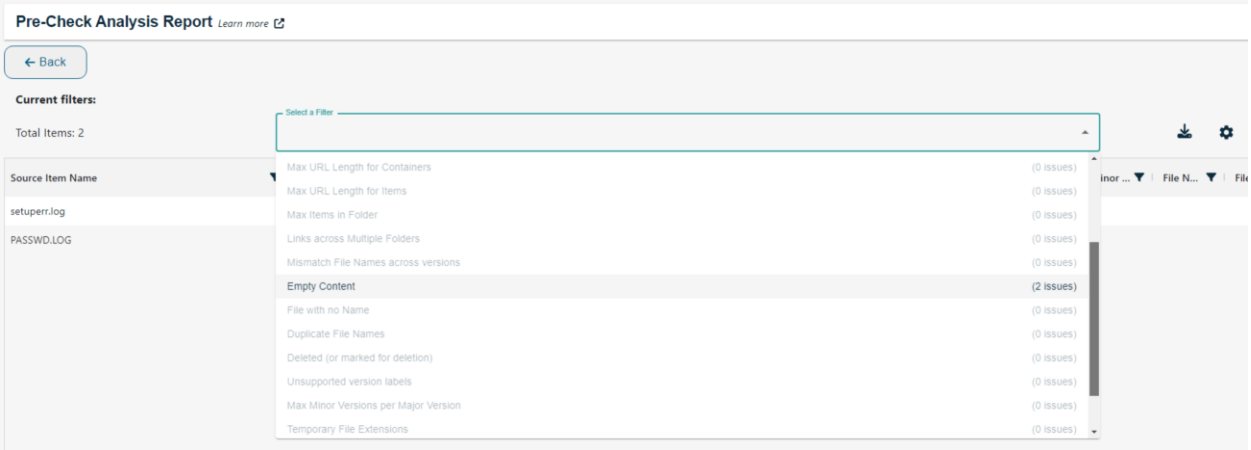


FIGURE 46 - EXAMPLE DISCOVERY REPORT

6.5.2. Pre-Check Report

Displays a list of items that have been identified as potential issues by the Pre-Check step. A breakdown of the issues by type appears in the drop-down list. Selecting the appropriate Filter will list the items affected by this pre-migration check.



An export option is available which will create a CSV file of the items that apply to the filter. It will include the result of whether the particular item is impacted by the pre-check against the source or the transformed data (if relevant).

6.5.3. Migrated Items Report

Displays a list of all items with a Status of Success

6.5.4. Failed Items Report

Displays a list of all items with a Status of Failed. This is useful to understand issues from a migration and what steps to take to resolve issues.

6.5.5. Error Summary Report

Over several releases, CPS errors have been enhanced to include an error code to simplify error resolution. These are a 3-digit number that will appear in the status message text and can be clicked to perform a search of the CPS support knowledgebase.

For fixed error codes that have occurred in the processing of a task the error summary report will list them and the number of occurrences.

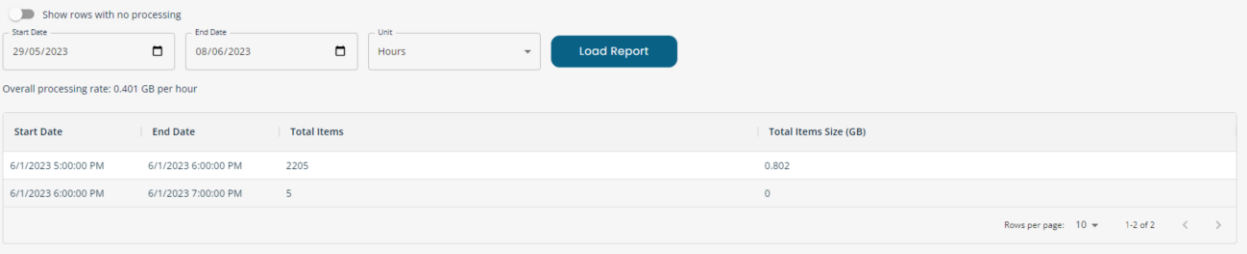
NOTE: Since not all errors currently have a fixed error code, the report also displays a total count of all other errors and more details of them can be seen in the failed items report.

6.5.6. Migration Audit Report

Displays a list of all items in the source, irrespective of processing status. This is a useful way of tracking over migration and often this data is exported and presented to the customer for their records.

6.5.7. Performance Report

Display migration statistics between the specified date period and summarised on hourly or daily basis.



By default periods of no processing will not appear in the table unless the toggle ‘Show rows with no processing’ is selected.

The overall processing rate (GB per hour) is also calculated and displayed. Time periods for which migration occurs are treated as whole units so for example if processing occurs over a 1 hour period spread across two hours, the hourly rate is based upon 2 hours. With longer processing runs this impacts the overall processing rate calculation less. Selecting the ‘show

rows with no processing' also affects the overall processing rate calculation because then hours during the time period with no processing are included in the calculation.

6.6. Activity Log

The activity log is accessible from the Menu option and displays the history of job processing. Each time Discovery, Pre-Check or Migration Step are executed the activity log is updated to reflect this.

Process Type	Process Status	Start Time	End Time	Status Message
Capture Metadata	Complete	9 November 2021, 16:09	9 November 2021, 16:09	Pre Migration Check found issues
Capture Metadata	Complete	9 November 2021, 16:09	9 November 2021, 16:09	Pre Migration Check found issues
Classify	Complete	9 November 2021, 15:33	9 November 2021, 15:35	Processing completed
Classify	Complete	9 November 2021, 15:29	9 November 2021, 15:32	Processing completed
Classify	Complete	9 November 2021, 15:10	9 November 2021, 15:10	Processing completed
Classify	Complete	9 November 2021, 15:03	9 November 2021, 15:05	Processing completed
Classify	Complete	9 November 2021, 13:23	9 November 2021, 13:26	Processing completed
Classify	Terminated	9 November 2021, 13:11	1 January 1970, 01:00	Left in inconsistent state
Discovery	Terminated	8 November 2021, 18:42	9 November 2021, 11:17	Cancelled
Capture Metadata	Complete	5 November 2021, 15:57	5 November 2021, 15:57	Pre Migration Check found issues

FIGURE 47 - ACTIVITY LOG

If a migration phase is executed the activity log records the last step as the Process Type, so in the above Classify is the Process Type, even though Capture Metadata step may have also been executed for some/all items.

7. PHASE 1 – DISCOVERY AND ANALYSIS

The Discovery process gathers information for each item contained in the specified source connection. Information gathered includes basic metadata, item version information, location etc. Discovery may take several hours or longer depending upon the amount of data to be captured and the performance of the source system. A filter can be created to limit the data to be discovered. This information is held in the CPS Staging Database and can be viewed from the Items List.

Anything that is to be migrated must be first successfully discovered in this phase. It's therefore vital to ensure Discovery is successful and the results contain all files to be migrated without errors.

This process will help confirm the source data volumes. This can then be used to help determine the number of CPS servers that may be required. It also has an impact on CPS licensing since licenses are volume based and assoicated with a specific server. For example, if there are several very large Sites Collections it may it recommended to process these on separate more powerful CPS servers so each of these servers would require it's own CPS license containing details of the volume that should be migrate.

> Note: Discovery only captures basic information on items, it does not capture custom metadata or item content. Only items that are supported by CPS will be captured during discovery.

7.1. Starting Discovery

NOTE: Before starting the Discovery process, it's strongly recommended that all relevant documents are 'checked-in' otherwise such items may not be captured by the discovery process.

Go to  on the Navigation panel to access Create Migration Joband then select the appropriate job.

FIGURE 48 - MIGRATION JOB LIST

After selecting a job the Migration Dashboard is displayed, Select **Discovery.**

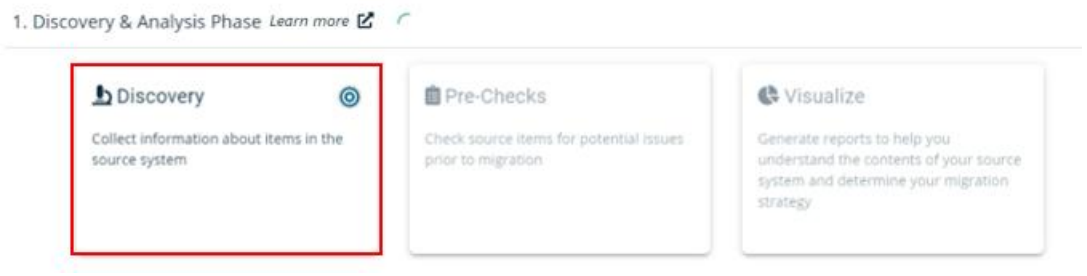


FIGURE 49 - PHASE 1 - DISCOVERY

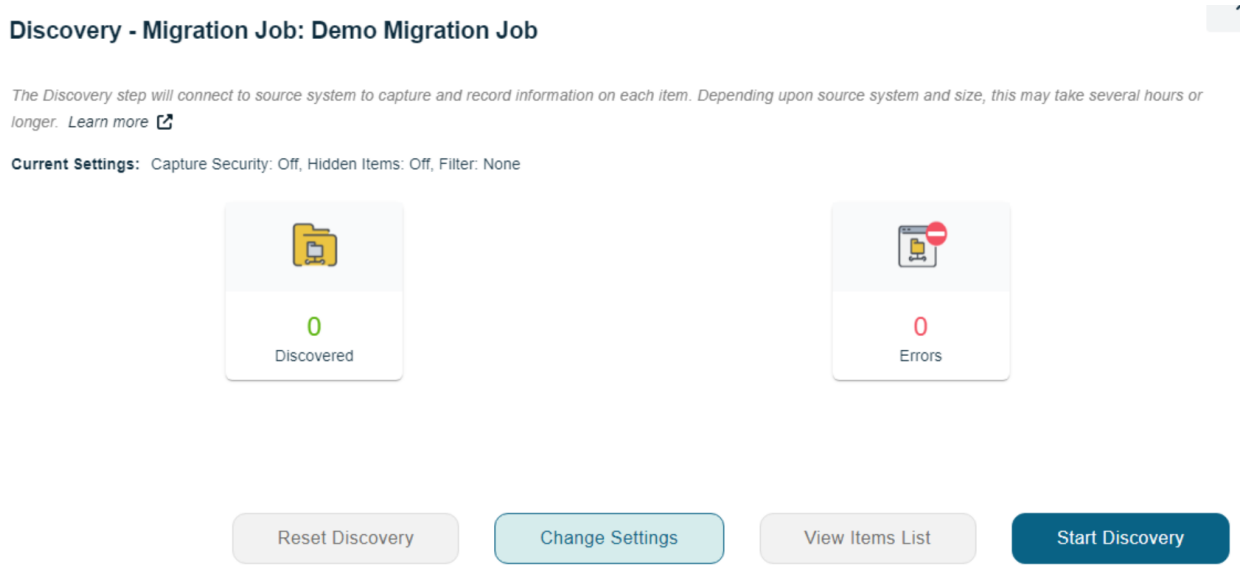


FIGURE 50 - STARTING DISCOVERY

Select **Start Discovery** option to begin the process of capturing information about the source environment into CPS. For many sources, the discovery progress will be displayed in the above screen. For IManage and EDocs, discovery statistics will only be shown at the end of discovery.

After Discovery is complete, the discovery statistics information on the Migration Dashboard will be updated to reflect the information captured during the discovery.

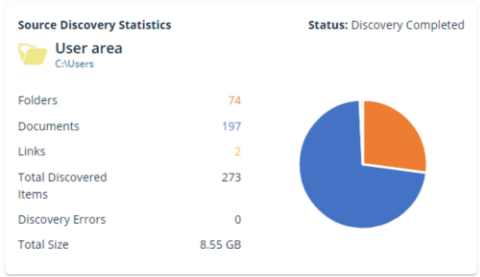


FIGURE 51 - DISCOVERY STATISTICS PANEL

7.1.1. Background Operations

Discovery, pre-check, and migration activities can run as background operations allow allowing the client still to be used during that time. Select the × button on the Discovery Progress/Migration whilst operation is running. The relevant statistics will be continually updated on the dashboard and the status will be displayed as **"In Progress".** Select the ⤢ icon to return the full progress screen.

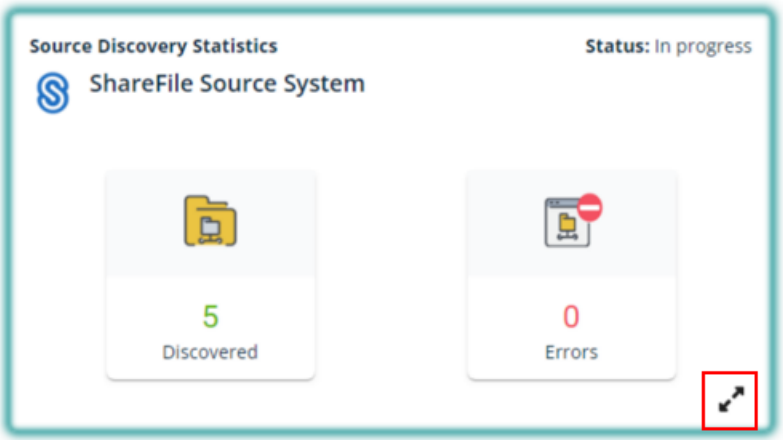


FIGURE 52 - DISCOVERY STATISTICS PANEL DURING DISCOVERY RUN

Only one background operation can execute on the CPS server at any one time. When a background process is running, it will disable options on the dashboard for the current job and all other jobs.

- Discovery – For current and all other jobs, all options available with the exception of Discovery, Pre-check, Migration and Shortcut Links
- Migration – All options disabled within the current job. For all other jobs, Discovery, Pre-check Migration and Shortcut Links are disabled.

The options become enabled as soon as the background operation completes.

7.1.2. Discover Errors

Any failures during discovery will count as discovery errors. Errors can often be at the folder level so even a single discovery error could mean that there are still many 1000's of items that need to be discovered so it's essential these errors are understood and resolved if possible.

7.1.3. Discovery Options

Use the **Change Settings** option to control the discovery of security information or apply a filter to what should be discovered. By default, this option is only available when Discovery has not previously been executed. See advanced discovery option in general settings for more information.

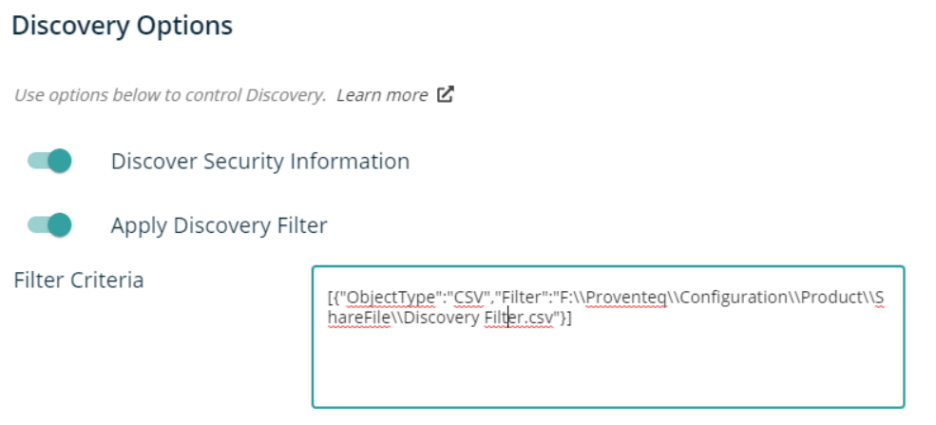


FIGURE 53 - DISCOVERY FILTER OPTIONS – EXAMPLE

For Documentum and File System, an additional discovery option to control the discovery of deleted items is also available.

Discover Security Information

In most cases, this option should be selected but in migrations where all the security information is determined by the target location then this option can be unselected which will improve discovery performance.

7.1.4. Version filtering

For some projects it can be useful to limit the number of versions of documents that are migrated. This can be done at the discovery stage by filtering the number of items discovered. Using this feature it's possible to limit the discovery to just the latest version of each document or the last 'N' versions, where N is a value specified by the user.

This feature is available for ShareFile source.

Delta behaviour when limiting versions.

If a version filter is applied and discovery occurs and the subsequent source item is edited then when discovery is run again, these new version(s) will be discovered, and this will not affect the already discovered items. So for example if version filter is last 3 versions and then a source document is edited twice and discovery re-run then 5 versions of the document will now be in CPS staging database to be migrated.

7.1.5. Discovery Filters

Discovery filters can be applied to limit the information captured during discovery. The filtering capabilities available vary depending upon the source system and whether CPS is configured to use the source system API or not. For example, connecting to Documentum using API provides different filter capabilities if connecting using DB. So, it's important to understand filtering requirements before fully configuring source connection.

For full details of filter capabilities and syntax see. 18.2.

Editing Discovery Filters

By default, after a discovery has begun it's not possible to edit the discovery filter. This is because it can create inconsistent results. For example having no filter and then adding a filter does not remove existing discovered items. Equally removing an entry in the filter does not remove the items from discovery.

However in some scenarios it can be useful to edit the filter, for example to correct a filter value that is specified incorrectly.

Based upon the scenarios above, if editing the filter will not create inconsistent results, then by selecting the Advanced discovery option in General Settings, it allows the Discovery filter to be edited even after discovery has started. Contact Technical Support for further information.

7.1.6. Reset Discovery

The **Reset Discovery** option is available after Discovery has been run and may be useful if the Discovery was incorrect and needs to be restarted anew or the size of the staging database needs to be reduced after a migration job has completed the migration. It will remove all existing discovered items, as well as structure, content, and security mappings. It does NOT delete data migrated into the target.

After the reset has completed, the **Start Discovery** option is displayed to allow a new discovery process to be started.

7.2. Delta Migrations

Any modifications to source items after discovery has completed will not be captured unless Discovery is run again. A Delta Discovery/Deta Migration is the term used for additional 'top up/incremental' migrations after the main migration has occurred and may be required when the source system is still active and there is a period of time between completion of the migration and switch over to the new target.

For example, a migration may take 1 week to run and during that time users are still actively altering the source and so these new documents and edits need to be detected and migrated before switching over.

A Delta discovery run will typically take longer than the original discovery due to the need to gather all information and compare discovery results.

To start a Delta run, select the Discovery option on the Migration Dashboard and select the **Start Incremental Discovery** from the Discovery screen.

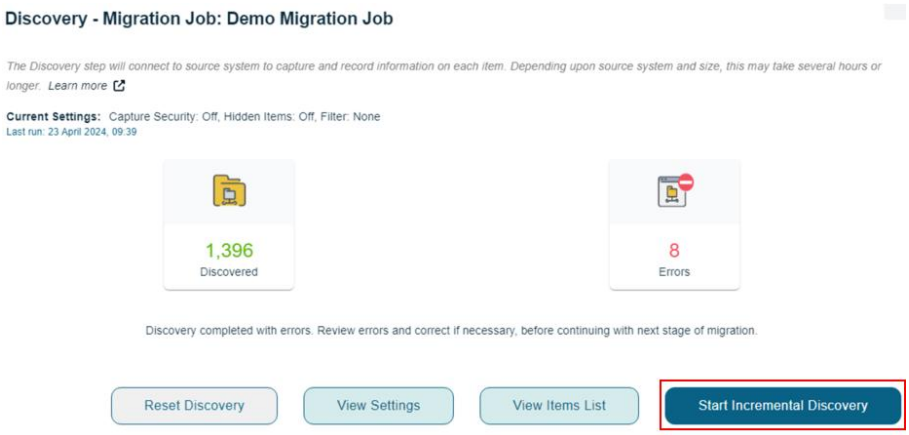


FIGURE 54 - STARTING INCREMENTAL DISCOVERY

The steps for a Delta migration are the same as for the initial migration.

Discovery will identify new/changed/deleted items and pre-check should be run in case any new data fails these checks. In most cases it may simply require a 'review' of the existing settings to review structure mapping, content types and security information.

Finally run the Data Migration and if relevant the Shortcut Links Migration

The table below explains the different delta scenarios and expected results.

Note: It is not recommended to move or rename or reparent folders and files while migration is underway, as it could lead to lots of duplicate content in target and it will consume additional license utilization.

Delta Scenario	Expected Result
New Folder is added	• New Folder would be created on target location.
Folder is modified/renamed	• New folder would be created in modified/renamed name and original folder will remain in the same target location which is migrated earlier.
Folder is moved/re-parented to different location	• New Folder would be created in moved/re-parent location and original folder will remain in the same target location which is migrated earlier.
Folder is deleted	• Folder and contents would not be deleted from target location.
New File is added	• New File would be created on target location.
File is modified	• File would be updated in delta run
File is renamed	• New file will be created with the new name and the original file will remain in the target.

File is moved/re-parented to different location	• New file would be created in moved/re-parent location and original file will remain in the same target location which is migrated earlier.
File is deleted	• File would not be deleted from target location.
New version is created	• New version is uploaded in delta run.
Version is deleted	• Version would remain in the target.

7.3. Items List

To launch the Items List, select the Items List menu option or launch from the Discovery dialog box. The Items List will display details of each item captured from the source during Discovery and will be updated by operations executed in the Migration Phase.

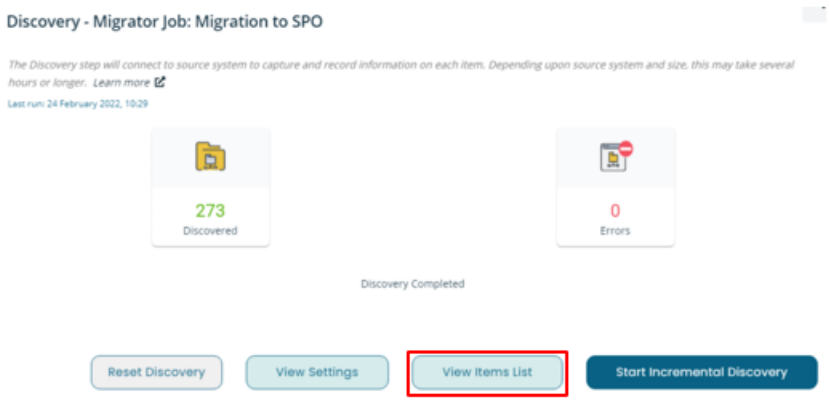


FIGURE 55 - LAUNCHING ITEMS LIST FROM DISCOVERY PROGRESS SCREEN

Columns displayed will vary depending upon the source being migrated. Filters can be applied to gain more insight into the data being migrated and cleared using the × icon. Additional columns can be added or removed using the settings icon ⚙.

When processing issues occur the details of the error can be found in the Status field.

Processing run: It is possible to filter on **Last Processing Run**.

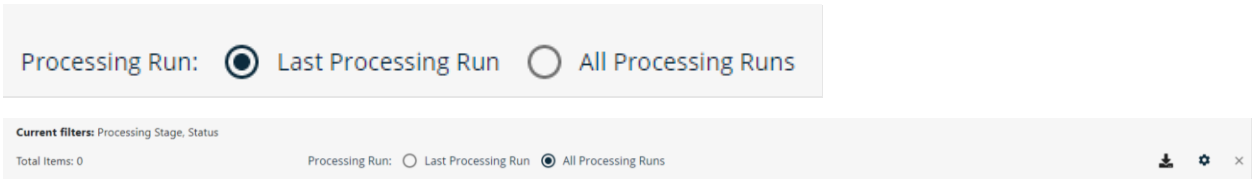


FIGURE 56 - PROCESSING RUN FILTER

Selecting Last Processing Run will filter the list items to only show items that were processed in some way in the last execution of a migration step.

Export to CSV

The Item List can be exported into a .CSV file using the option. If a filter is applied to the items list, it will apply to the exported list also. All columns will be exported.

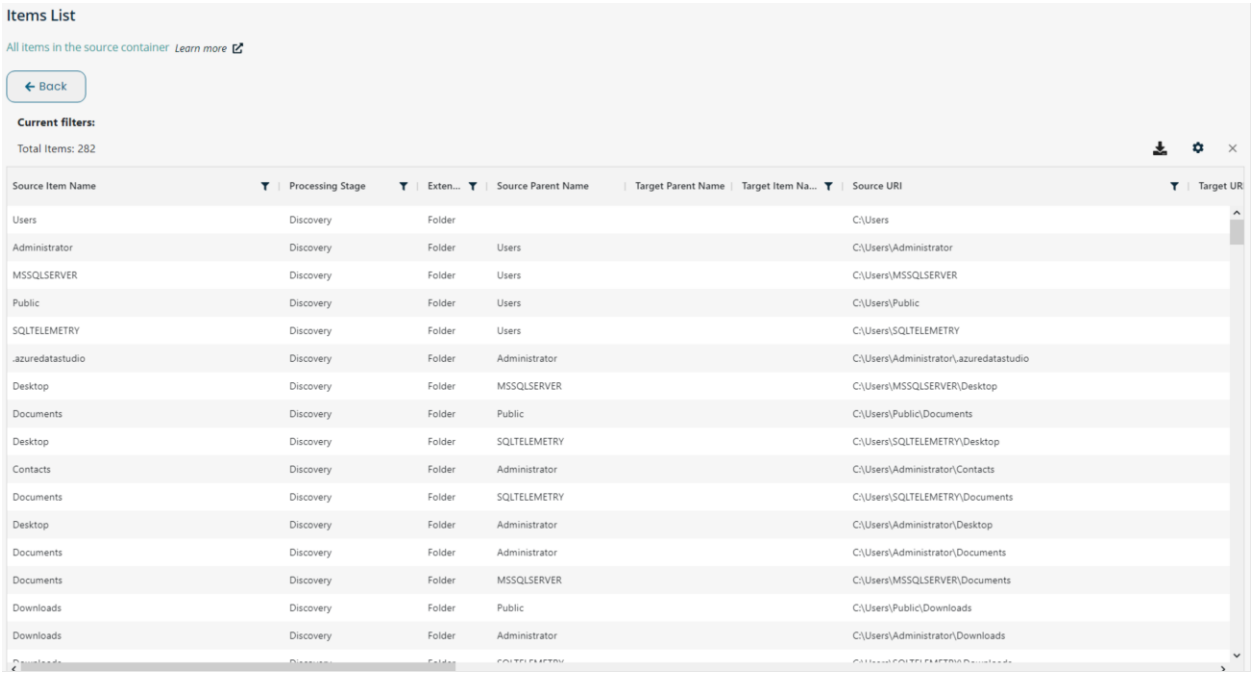


FIGURE 57 - ITEMS LIST

7.3.1. Item level information

Select the item to display the Item Details screen.

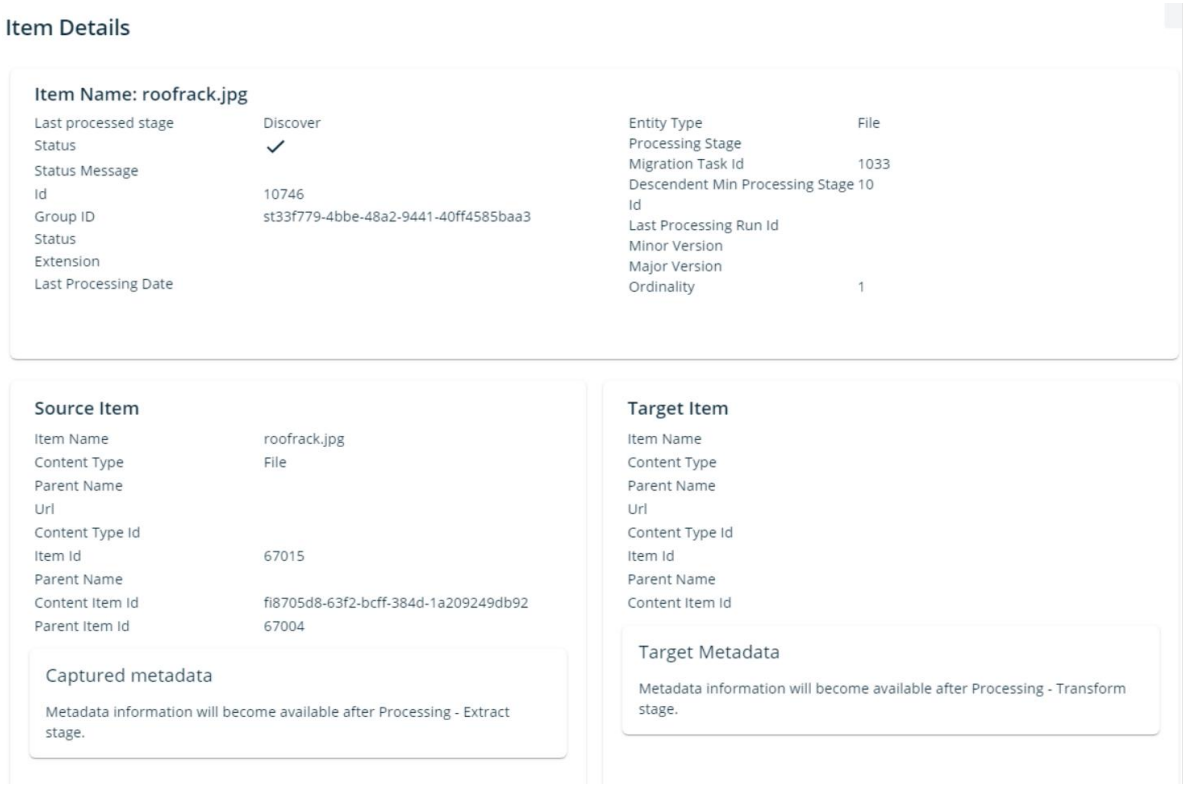


FIGURE 58 - EXAMPLE ITEM DETAILS

As an item moves through various process stages then additional information will appear as show in the table below.

7.3.2. Items Details

The following table lists all available columns, * indicates additional columns to **Summary** view.

Column	Description	Possible values	Stage Populated	Notes
Source Content Item ID	This identifies an individual document version in source.		Discovery	
Target Content Item ID	This identifies an individual document version in target.		Load	
Group ID	This identifies an individual document.		Discovery	The same document with several versions will have the same Group ID
Source Item Name			Discovery	
Major Version	e.g. If document version was 10.5 this would be 10		Discovery	

Minor Version	e.g. If document version was 10.5 this would be 5		Discovery	
Processing Stage		***Discover*** – Information gathered by Discovery ***Capture Metadata***- Step 1 of Migration process executed ***Classify*** – Step 2 of Migration process executed **Transform** – Step 3 of Migration process Load – Step 4(final step)	Discovery and Migration Phase	
Source Parent Name			Discovery	
Target Parent Name			Transform	
Target Item Name			Transform	
Extension			Discovery	
Source Url			Discovery	
Target Url			Transform	
Status			Migration Phase	
Source Content Type	This will contain either the content type assigned to the item by the source OR the a Customer Content Type assigned through custom classification		Classify	
Target Content Type*			Transform	
Ordinality	This represents the order in which versions items will be migrated		Discovery	This represents the order in which versions items will be migrated
Entity Type			Discovery	
Status Message				
SourceContentTypeId			Classify	
TargetItemId			Classify	

7.3.3. Source Virtual Collections

The items list may also contain items which don't physically exist in source but are captured in Discovery to represent the source hierarchy. For example, when discovering all the personal folders in ShareFile, a root item will be created in items list for "Personal Folders", which will act as the parent item for individual personal folders. Such items will have a source Entity Type of "VirtualCollection".

7.3.4. Target Structure Entries

As well as source items, some additional entries may appear in the items list termed 'Target Structure Entries'. They are added to represent structural elements that may need to be created by CPS during the migration. So for example if a SharePointLibrary needs to be created in target for splitting source folder hierarchy across libraries , an entry for this will also appear in the Items List so that the success or otherwise of that can be displayed. These items can be identified as they will have "StructureMapping" as the source ItemName.

7.3.5. Processing History

The details of an item being processed can be viewed from the Processing History tab on the Item Details screen.

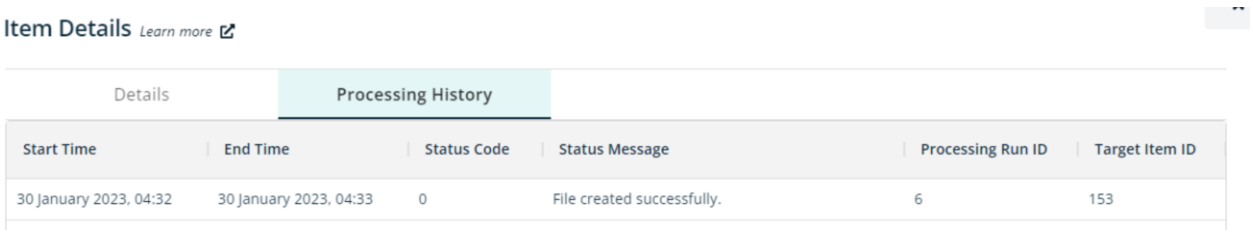


FIGURE 59 - ITEM PROCESSING HISTORY

7.4. Pre-Checks

Migrations between different sources and targets often have incompatability issues due different capabilites,limitations and behaviours in the target which could impact the migration.

The Pre-check option allows a set of checks against the source data to be executed, based upon the known limits of SharePoint. Each can be enabled/disabled and some have additional configuration parameters which can be customised as required.

If data has been actually been processed up to the Transform stage then a toggle will appear on the pre-check screen which controls whether pre-check runs against data generated by the Transform step.

NOTE: Migrations will still be attempted for items that are reported as potentials issues in the Pre-check. This is why it's important to review and correct as necessary to ensure the successful migration of items.

NOTE: The nature of pre-checks means that if Discovery is halted before completion then pre-check results may be inconsitent and cannot be relied upon.

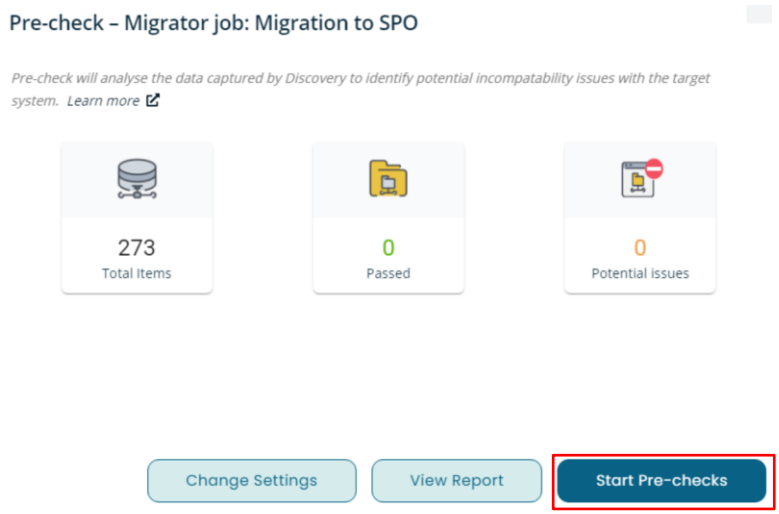


FIGURE 60 - STARTING PRE-CHECKS

7.4.1. Changing Pre-Check Settings

The 'Change Settings' option displays the options below and allows the checks to be enabled/disabled as well as specifying any relevant values.

Restrictions:

Restrictions are issues that may prevent the folder or file actually being migrated. So for example, because they contain characters in the filename which the target will not allow.

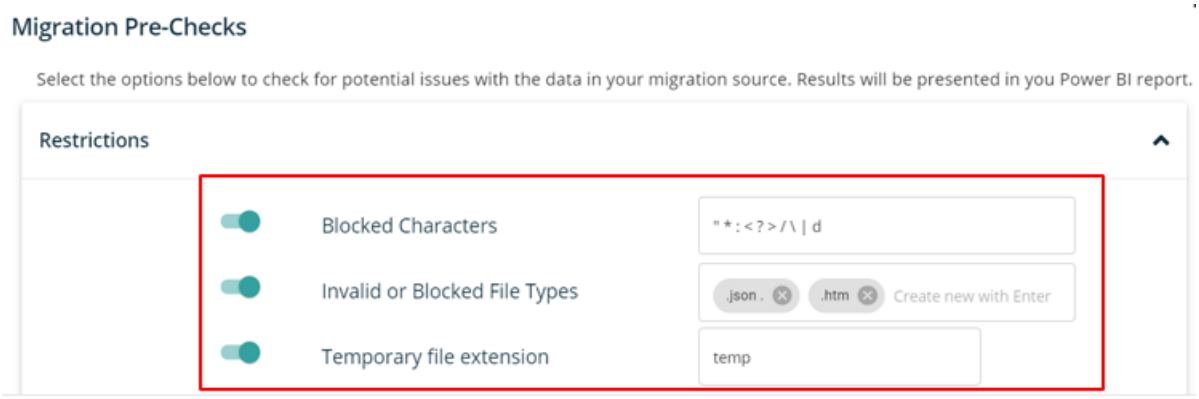


FIGURE 61 - PRE-CHECKS – RESTRICTIONS

Blocked Characters in Filenames:

Description: Checks for illegal characters in Source file/folder.

Impact: Items will be migrated but will contain replacement character as specified in General Settings.

Resolution steps: Use General Settings to assign a character to replace any invalid characters.

Invalid or Blocked File Types:

Description: Checks if file type of item is on the blocked file types list. Default list is empty as SharePoint currently supports all File types.

Impact: Items would fail migration but currently N/A.

Resolution steps: Exclude items from migration or convert to supported file type if applicable.

Temporary file Extensions:

Description: Checks for files with specified file type that is considered a temporary file extension and so is likely something that does not require migration. Use comma separated list to provide multiple file extensions.

Impact: Items would be migrated.

Resolution steps: Contact Technical Support.

Limitations

These are limitations in the Target that may prevent the folder/file being migrated. Unlike restrictions, limitations are those that are only an issue once a value exceeds the defined threshold based upon the limits of the target.

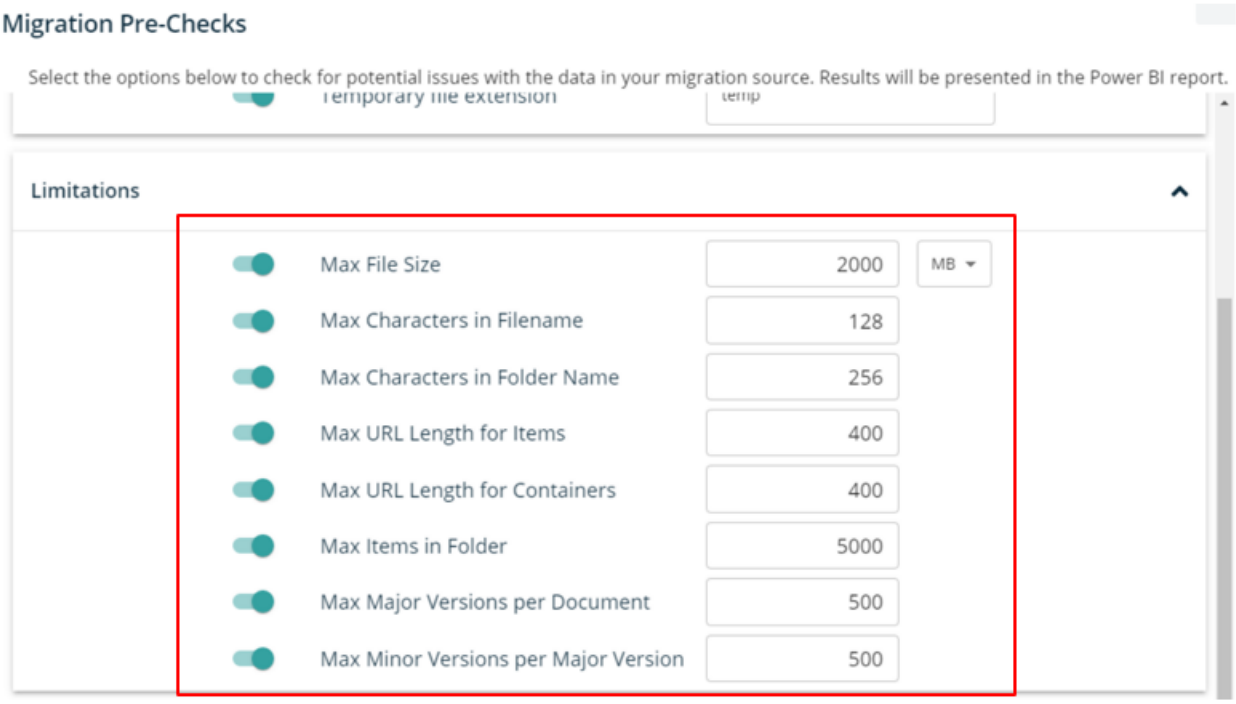


FIGURE 62 - PRE-CHECKS – LIMITATIONS

Max File Size:

Description: Checks whether the file size exceeds the maximum file size for SharePoint/SharePoint Online.

Impact: Items will fail migration.

Resolution steps: Migrate manually or exclude from migration.

Max characters in Filename:

Description: Checks whether the length of item name exceeds the SharePoint/SharePoint Online limits.

Impact: Items will fail migration.

Resolution steps: Rename source item or rename during migration by setting up a rule for item name mapping. See Populating Target using Functions where you can use functions to trim fields.

Max characters in Folder Name:

Description: Checks whether the length of folder name exceeds the SharePoint/SharePoint Online limits.

Impact: Items will fail migration

Resolution steps: Reorganise folders in source, use structure mapping or contact technical support for assistance.

Max URL Length for Items:

Description: Checks whether the length of a parent container exceeds SharePoint limits.

Impact: Items will fail migration.

Resolution steps: Contact Technical Support.

Max URL Length for Containers:

Description: Checks overall URL length

Impact: Items will fail migration.

Resolution steps: Rename item(s) in source, rename during migration by setting up a rule for item name mapping or shorten the path during migration by using structure mapping.

Maximum Items in Folder:

Description: Checks whether the folder contains more items than permitted within SharePoint Online Document Library folders.

Impact: Items will fail migration once target item limit reached in folder

Resolution steps: This can be resolved either by restructuring data in the source or through structure mapping during migration.

Max Major Versions per Document:

Description: Checks for number of major versions per item since SharePoint has a limit on the number of major versions supported.

Impact: Items will fail migration.

Resolution steps: Contact Technical Support

Max Minor Versions per Major Version:

Description: Checks for number of minor versions per major version since SharePoint has a limit on the number of minor versions supported.

Impact: Items will fail migration.

Resolution steps: Contact Technical Support

Duplication

Any duplication or empty files that may not get migrated into the Target

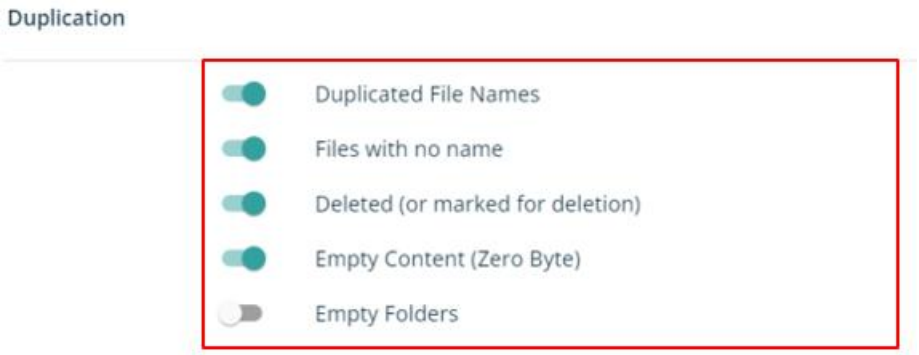


FIGURE 63 - PRE-CHECKS – DUPLCATION

Duplicate file names:

Description: Checks if there are items in the source dataset with the same name

NOTE: Unlike other pre-checks, this check can take several hours to run in larger environments. E.g. 8 hours in an environment of several million items.

Impact: Items will fail migration unless one of the resolution steps below is used.

Resolution steps: In General Settings, Duplication there is the option “Create a new file with a unique name” to control the migration of duplicate named items and provide a unique name as required (by default this is enabled). Alternatively, restructuring the source system and rerun discovery or resolve the conflict using structure mapping.

Files with no name:

Description: Checks whether the Item Name field is blank.

Impact: Items cannot be migrated

Resolution steps: Edit source item or create a function mapping to assign value. See Populating Target using Functions where you can use functions to apply a name in such scenarios.

Deleted (or marked for deletion):

Description: Checks whether the item is marked as deleted in the Source system.

Impact: Items that have been deleted/marked for deletion will still be migrated.

Resolution steps: Contact Technical Support

Empty content (Zero Byte):

Description: Checks whether the Item size is zero bytes as this could indicate issue with the source system

Impact: Items will fail migration.

Resolution steps: Items should be examined by a record officer/system administrator to determine whether the files truly are missing. if so, the issue can be worked around by loading placeholder content using the "LoadPlaceholdersForMissingOrEmptyFiles" advanced setting.

Empty Folders:

Description: Checks for folders with no content. By default, this check is OFF and so not executed.

Impact: Empty folders will be migrated.

Resolution steps: Review the list to see if such folders should be excluded from migration. By default, such items will be migrated, but the configuration key FilterEmptySourceFolders can be set to True, so empty folders are not migrated.

Other Checks

Other properties that may prevent **migration into the Target.**

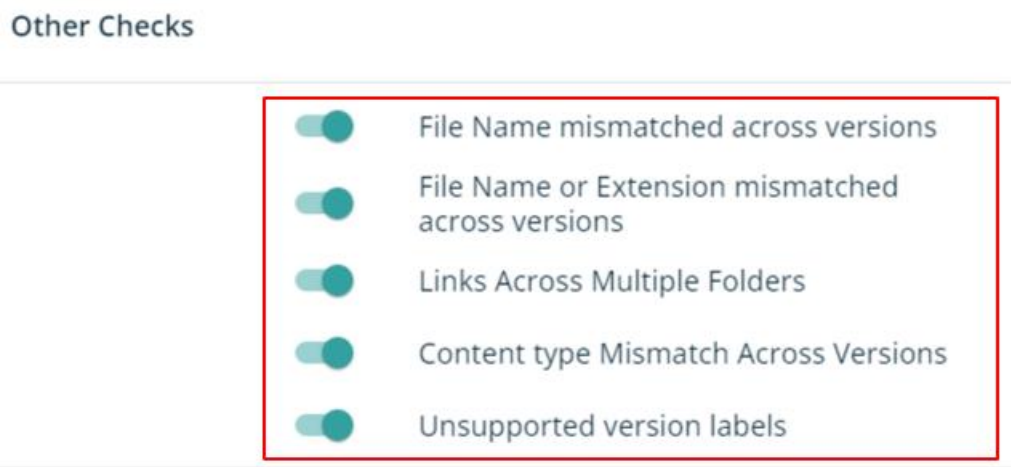


FIGURE 64 - PRE-CHECKS – OTHER CHECKS

Mismatch File Names across versions:

Description: Checks whether the versioned items having different file names. I.e. version 1 of a file is called Dept Training.docx and version 2 is called Dept Training(Old).docx. In SharePoint, all versions of the file must have the same name.

Impact: Items will be migrated and by default will have the same name as the first version

Resolution steps: For steps on how to use information from the latest version across all versions contact Technical Support

Mismatch File Names or Extensions across versions:

Description: Checks whether the filename or file extension of an items changes across versions of the item. SharePoint requires all names and extensions to be the same for all versions.

Impact: Items will be migrated and by default will have the same name/extension as the first version.

Resolution steps: For steps on how to use information from the latest version across all versions contact Technical Support

Links across Multiple Folders:

Description: Detects if source system has Shortcut links. In the Power BI report, these are labelled as content referenced by multiple folders.

Impact: Shortcut links need to be migrated if required.

Resolution steps: If Shortcut Links need to be migrated, ensure “Links” source content type is configured correctly in Content Type mapping and run Shortcut Links migration, after all items are successfully migrated.

Content type Mismatch Across Versions:

Description: Checks whether the Content Type/Category/Classification for all versions of a document are the same. Since SharePoint requires all version to be of the same Content Type.

Impact: Items will fail migration.

Resolution steps: Contact Technical support

Unsupported version labels:

Description: SharePoint online supports version labels like 1.1, 2.3 etc. Some source systems supported version labels such as 2.3.1 or even text label so such issues will be detected with this check.

Impact: Items will fail to migrate

Resolution steps: Contact Technical support

7.4.2. Running Pre-Checks

Select **Discovery & Analysis Phase->Pre-Checks** option and then select the Start Pre-checks. The checks normally take several minutes to execute.

7.4.3. Review and correcting pre-check issues

There are two ways to view pre-check results.

Power BI Report

Once the pre-check is complete, the number of potential issues will be displayed. Clicking on the 'View Report' will launch PowerBI to display the results. There are several pre-check related tabs which contain the results of each check. In general, any checks which return a result of 'True' are issues to be investigated.

Two warnings are displayed when opening Power BI

1. Potential security risk

Displays warning about 'Potential security risk' due to the file using multiple data source. Click **OK** to continue.

2. Native Database Query

Displays warning to approve running native query. Click **Run**

PowerBI will open up to display the Content Summary dashboard. Select the Pre-Check General tab to view the total number of issues.

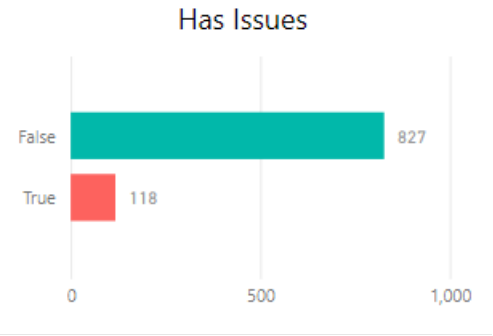


FIGURE 65 - TOTAL PRE-CHECK ISSUES

The result may be highlighted in amber to suggest a Informational/Warning as opposed to a failure. An example of this is the 'Empty Folders' warning.

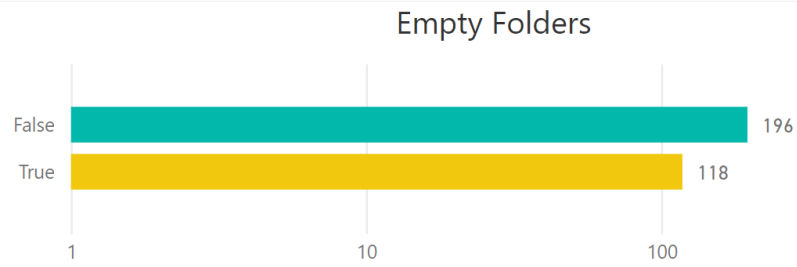


FIGURE 66 - EXAMPLE OF PRE-CHECK WARNING

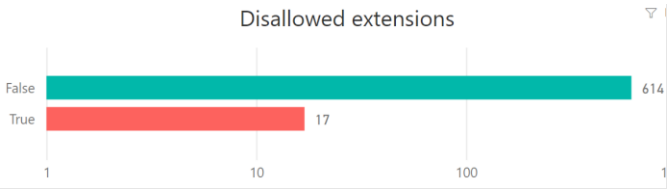


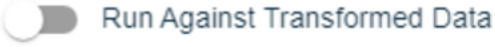
FIGURE 67 - EXAMPLE F PRE-CHECK ERRORS

In this example of an error, 17 items have file extensions not allowed in the target.

7.4.4. Pre-check against transformed data

By default, pre-check will process the source data that was captured during the discovery stage. During migration preparation steps later on in processing, you may be using structure mapping, functions or PowerShell scripts to transform the data for migration. Such operations may change the data in such a way that it may be acceptable to the target system (for example perhaps structure mapping is used to reduce folder path lengths).

If the Migration Phase, items have been processed up to the Transform stage then all such steps will be executed and so the new target paths/values will have been created. When this happens the pre-check will show a new option.



If this is selected, then the pre-check will run against the information created during the transform step above and relevant displayed statistics will also be displayed.

Pre-Check Report

There is a pre-check report available and a drop down list of issues and the number of items impacted will be displayed.

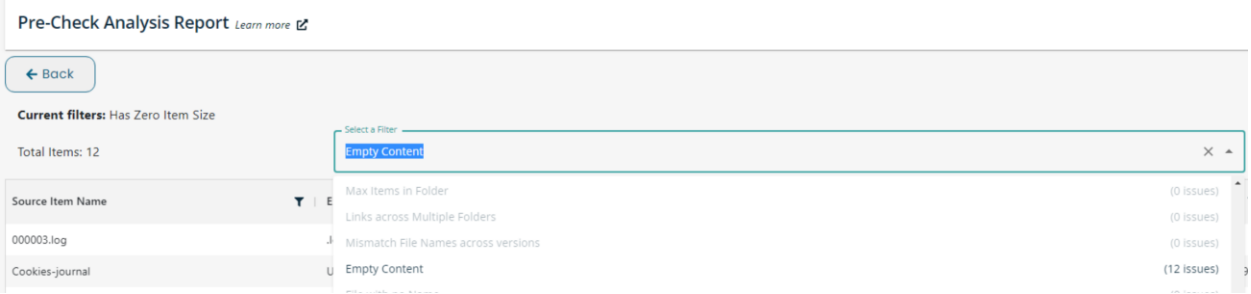


FIGURE 68 - PRE-CHECK FITER

8. Creating Detailed Migration Plans

Introduction

The completion of the Discovery Phase is an important milestone and provides the essential information necessary to help create a more detailed migration plan. Here are some of the key considerations when planning a migration.

8.1.

- Performing pilot/smaller scale migrations to validate migration configuration and understand performance to help identify any changes that may be required in the environment, including adding additional CPS servers.
- Understanding business priorities and data volumes to help determine the order of the migration.
- Initiate User Acceptance Testing (UAT) planning.

Delta Migration

If the source system is still active, then after an initial migration, customers may execute a 'top up' or otherwise known as a 'delta migration'. In this way, the bulk of the data is migrated, and the delta migration allows any subsequent updates that occurred on the source to also be migrated. Note that in such cases, the target is overwritten so it's important for users to understand this and only make and changes to documents in the source platform. For more information on Delta Migrations see Delta Migrations.

8.1.1. Content Type Planning

- Creating new Target Content Types

There may be source content types that have no current equivilent in the target and so decisions need to be made as to whether they should be mapped to an existing target content type or whether a new content type is created in the target.

Once these content types are created in the target then they can be mapped to using the Content Type Mapping feature.

- Mapping several source Content Types to single Target Content Type

In some migrations, customers may wish to simplify the existing content types and so map several different content types in a source to a single content type in the target.

- Separating single content type into many target content types

There can be scenarios were a single source content type needs to be separated out into more than one target content type. CPS allows the creation of 'Custom Content Type' and allows the user to define criteria using the metadata of the item to identify them as a particular content type which can then be mapped to a suitable target content type. See Custom Classification for more details.

- Sign off on metadata mapping

It's essential that all source fields are understood for all contents types so that metadata mapping is correctly configured and signed off to move the data to the target.

8.1.2. Structure Mapping

Each of the existing source locations needs to be reviewed and as applicable, the target location for these needs to be determined as part of this planning process. See Phase 2 – Structure mapping for more details.

When designing target layout be aware of the limitations in SharePoint that may create performance issues. See also:-

https://support.microsoft.com/en-us/topic/cf37b449-a2a7-4ef9-9f4d-aab826db5b30 - Understanding and managing large lists.

https://docs.microsoft.com/en-us/office365/servicedescriptions/sharepoint-online-service-description/sharepoint-online-limits for more information on SharePoint Online limits.

8.1.3. Security

The security mapping will export lists of users, groups and permissions that may need to be mapped. See Phase 2 – Security Mapping for more details.

Decision on what User in the target system should items be associated with if the source user is not in the target system e.g. Leavers

It is strongly recommended to review/simplify permissions as part of the migration. For example, putting permissions on the SharePoint Libraries and Folders rather than at an individual item level due to SharePoint limitations as SharePoint only allows 50,000 unique permissions in a single library.

8.1.4. Shortcut Links/Embedded Links processing

Decision on whether to migrate shortcut links from source to target. Note Shortcut Link migration occurs after all source items are migrated and is a much slower process.

Decision to convert links embedded within documents to resolve to new locations. This will also significantly impact migration performance. Note: Password protected/encrypted items cannot be processed.

9. PHASE 2 – STRUCTURE MAPPING

Structure Mapping is used to define where the source folders/containers are migrated to in the target. The location details in the Target Connection are combined with the information here to allow the full location to be determined. It is is one of the key elements within the Migration Preparation phase following on from successful Discovery and detailed planning.

It's important to bring this information and an understanding of the target and requirements of the business and end users to create a set of rules, to ensure data is migrated to the correct target location.

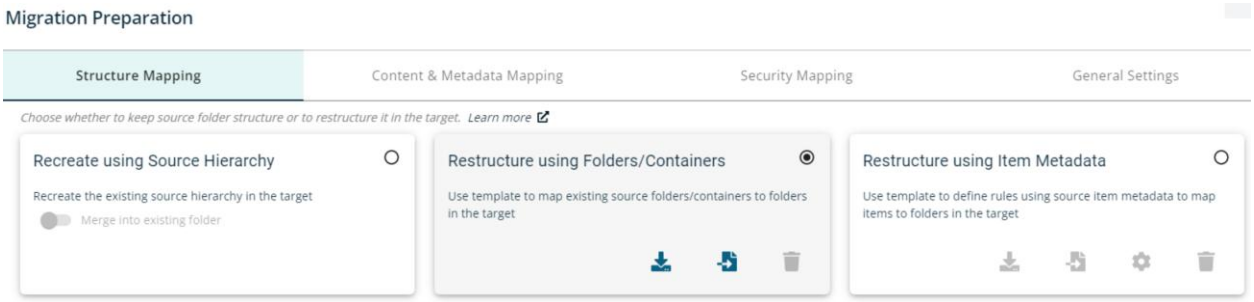


FIGURE 69 - PHASE TWO – STRUCTURE MAPPING OPTIONS

There are three structure mapping options that may be available depending upon the exact target location specified. See **Error! Reference source not found.**

- **Recreate using Source Hierarchy**

This will create the source folders below the specified target location in the migration job. This option is only available when the target location is a Document Library or Folder, and the source location is also a folder.

- **Restructure using Folders/Containers**

This feature allows folders to be reorganized during the migration by 'mapping' source containers to target containers. The assignment between source and target containers is done using a .CSV file which can be downloaded and then imported after editing.

- **Restructure using Item Metadata**

Metadata structure mapping is a feature whereby the target location for items is based upon customer written rules that use source item metadata properties. The assignment between source item metadata and target folders is done using a .CSV file which can be downloaded and then imported after editing.

Note: To update any structure mapping entries, it's necessary to first delete the existing structure mapping before then importing new mappings.

Once any items have been processed in the Migration phase then it's not possible to alter the structure mapping. To alter the mapping, it's necessary to use the Undo option which

will first delete any migrated items and provides the option to undo the current structure mapping.

9.1. Recreate using Source Hierarchy

Selecting this option will create a folder hierarchy in the target by creating the existing source folder/container hierarchy beneath the location specified in the Target.

The existing hierarchy in the source will be used in the target.

This option is only available if a Document Library or folder has been selected in the target connection since then no additional information is required to migrate the existing source hierarchy into these.

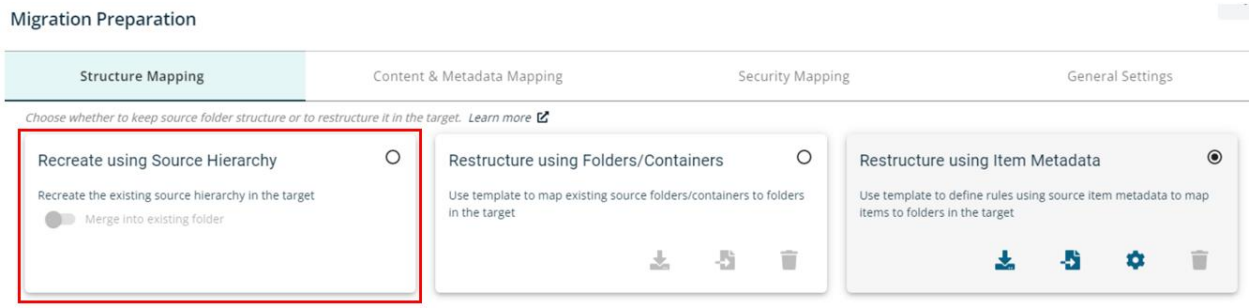


FIGURE 70 - RECREATE USING SOURCE HIERARCHY

- **Merge into Existing Folder Option**

Recreate structure mapping has the **Merge** into Existing folder option. By default, items will be migrated into the folder specified in the target. If unselected, then the final folder path specified in the source of the items will be used to first create a 'root/parent' folder beneath which the items will then be migrated. See TargetRootSetting

9.2. Restructure using Folders/Containers

Using this feature the customer can create a table called a **'Structure Mapping'** which lists source folder/containers and specifies where they should be placed in the target. When an entry is specified, all its children will also be moved beneath it, so it's not necessary to specify all source folders.

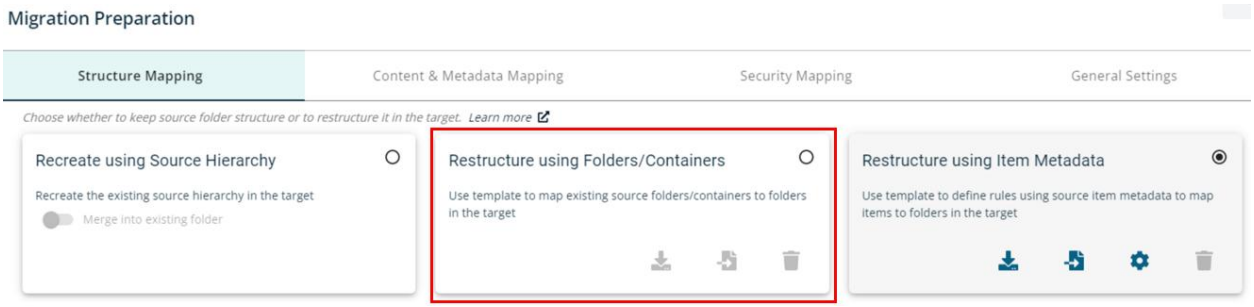
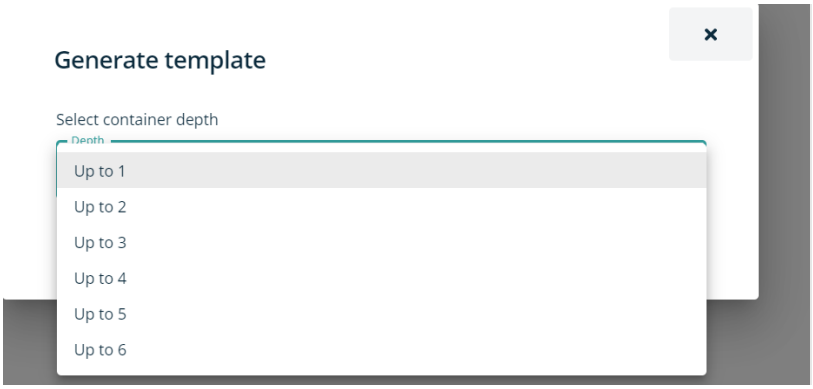


FIGURE 71 - RESTRUCTURE USING FOLDERS/CONTAINERS

In summary the process is to download source system structure information into a .CSV file. Edit this to add suitable target details and then upload the .CSV file.

Select the download icon to start Structure mapping. The .CSV file to be used to populate structure mapping will be populated list of folders from the source. When the download option is selected, the user is prompted for the folder path depth. This is the number of folders 'down' that it should export. Note the value is relative to the source path specified.



Example Source Folder Hierarchy:

```
Documents\Engineering\Project Diamond
Documents\Engineering\Project Diamond\Documentation
Documents\Engineering\Project Diamond\Beta Programme
Documents\Engineering\Project Diamond\Beta Programme\Internal
Documents\Engineering\Project Diamond\Beta Programme\SE Community
Documents\Engineering\Project Diamond\Beta Programme\External
```

Specify a folder depth of three would populate the CSV with unique folders that are 3 levels deep .

Folder Depth of 3

```
Documents\Engineering\Project Diamond
```

Folder Depth of 4

```
Documents\Engineering\Project Diamond
Documents\Engineering\Project Diamond\Documentation
Documents\Engineering\Project Diamond\Beta Programme
```

9.2.1. SharePoint Structure Basics

This section describes the basic principles of structures in SharePoint and may be helpful if the user is not very familiar with SharePoint.

The structure to locate any item in SharePoint is defined by 3 elements

SiteCollection
Document Library
Folder

Items can be stored at the Document Library level (i.e. the root) or within a Folder or Subfolder.

To migrate into SharePoint the SiteCollection, Document Library and Folder needs to be known. Depending upon the information specified in the Migration Job, all of this information may have been specified, if not, then the omitted fields need to be specified in Structure Mapping to allow these items to be migrated successfully.

9.2.2. Completing Structure Mapping Overview for Sharepoint

Source related fields in Structure Mapping (SourcePathID and SourcePath)

Depending upon the source. Structure mapping will either require the SourcePath or SourcePathID to be specified to identify the source location to be mapped depending upon the source. These columns will be populated automatically when the structure mapping template is downloaded.

Target fields in Structure Mapping

The information that's mandatory in structure mapping depends upon the full target path specified when the migration job was created. e.g. If a Site Collection was specified as the target location in the Migration Job, then TargetSubSite, TargetLibrary are required and TargetFolder is optional. (The folder names in the target will come from the source).

	SourcePathId	SourcePath	TargetSiteCollection	TargetSubSite	TargetLibrary	TargetFolder	TargetRootSetting
2		E:\FileSystem Test Data\Insurance		/	Insurance	/	Create
3		E:\FileSystem Test Data\Insurance\Health Insurance		/	Insurance	Health	Create
4		E:\FileSystem Test Data\Insurance\Car Insurance		/	Insurance	Car	Create
5		E:\FileSystem Test Data\Insurance\Travel Insurance		/	Insurance	Travel	Create
6		E:\FileSystem Test Data\Insurance\Health Insurance\Claims		SubSite	Claims	Health/Claims Docs	Create
7		E:\FileSystem Test Data\Insurance\Car Insurance\Claims		/	Claims	Car/Claims Docs	Create
8		E:\FileSystem Test Data\Insurance\Travel Insurance\Claims		/	Claims	Travel/Calims Docs	Create
9		E:\FileSystem Test Data\Insurance\Insurance Archive		Archive	Archive	Archive Docs	Merge
10		E:\FileSystem Test Data\Insurance\Others		Archive	Archive	/	Create

FIGURE 72 - EXAMPLE STRUCTURE MAPPING FROM FILE SHARE

Structure Mapping Field	Description
TargetSiteCollection	This is a URL that is required when:- A: Target path is to tenant level. B: To migrate to a Site Collection other than what's specified in target path.
TargetSubSite	Specify / to migrate to root, otherwise the name of the SubSite. If the SubSite does not exist, it will be created using the Team Site template.

TargetLibrary	Required if a Document Library is not specified in target path on the Migration job.
TargetFolder	Not mandatory. Specify if want to migrate to specific folder.
TargetRootSetting	Required if wanted to use MERGE setting.

FIGURE 73 - STRUCTURE MAPPING VALIDATION

When a structure mapping is imported, it will be validated using the logic from the above table to ensure that columns contain necessary values. With the exception of the TargetRootSettings field, it does NOT ensure that the values are correct/exist in the target.

Target SiteCollection

This is the **Full URL** of the Target Site collection.

Target SubSite

Specify / to migrate to root, otherwise the name of the SubSite. If the SubSite does not exist, it will be created using the Team Site template.

Target Library

If migrating to an existing library, the value needs to be the 'internal name' of the target library NOT the display name e.g. Shared Documents (Internal name) instead of Documents (Display Name). Note values are **case sensitive so specify the internal name exactly**. If the specified library does not exist, it will be created.

TargetFolder

Specify **/** to migrate to the root folder or else specify name of folder which will be created if it does not exist. Any value specified will overwrite the target folder.

TargetRootSetting

If TargetRootSetting field is blank or set to **CREATE**, the final folder path in the source path will be used to create a parent root folder in the target underneath which the data is migrated. A value of **MERGE** can be specified and in that scenario no root path is created, and the data is migrated directly into the target folder.

Adding comments to structure mappings

To make structure mappings easier to document and understand comment lines can be inserted into the structure mapping file. Any line that begins with a # or ! character will be treated as a comment and will be ignored during the structure mapping import and not appear in the client UI.

9.2.3. Completing Structure Mapping Overview for BOX

SourcePathId	SourcePath	TargetSiteCollection	TargetSubSite	TargetLibrary	TargetFolder	TargetRootSetting
P-258	AD1013.8 - Administration - Events	https://app.box.com/6144634745	/	/	TTTesting/Folder 1	Merge
P-266	AD1013.8 - Administration - Events/Documents	https://app.box.com/6144634745	/	/	TTTesting/Folder 2	Create
P-265	AD1013.8 - Administration - Events/Emails	https://app.box.com/22209797698	/	/	TTTesting/Folder 4/A/B	Merge
	AD1013.8 - Administration - Events/Documents/ 29 July 2008.pdf	https://app.box.com/22209797698	/	/	MergeAllContentsDocuments	Create
	AD1013.8 - Administration - Events/Documents/Other speakers.doc	https://app.box.com/22209797698	/	/	TTTesting/Folder 1	Create
	AD1013.8 - Administration - Events/Documents/ station.pdf	https://app.box.com/6144634745	/	/	TTTesting/Folder 4/A/B	Create
	AD1013.8 - Administration - Events/Emails/ To Do List .msg	https://app.box.com/22209797698	/	/	TTTesting/Folder 2	Create
	AD1013.8 - Administration - Events/Emails/FW: Christmas Cards.msg	https://app.box.com/6144634745	/	/	TTTesting/Folder 1	Create
	AD1013.8 - Administration - Events/Emails/FW: eCard from .msg	https://app.box.com/22209797698	/	/	TTTesting/Folder 4/A/B	Create

FIGURE 74 - EXAMPLE STRUCTURE MAPPING FOR BOX TARGET

Structure Mapping Field	Description
TargetSiteCollection	A combination of box domain and userId
TargetSubSite	Specify /
TargetLibrary	Specify /
TargetFolder	The hierarchy and the user
TargetRootSetting	Required if wanted to use MERGE setting.

For files don't specify any SourcePathId (GroupId).

9.2.4. Structure Mapping Examples

Structure Mapping example files are in C:\Program Files\Proventeq\Proventeq Content Suite\SampleFiles

Example File Share Structure Mapping

	SourcePathId	SourcePath	TargetSiteCollection	TargetSubSite	TargetLibrary	TargetFolder	TargetRootSetting
2		E:\FileSystem Test Data\Insurance		/	Insurance	/	Create
3		E:\FileSystem Test Data\Insurance\Health Insurance		/	Insurance	Health	Create
4		E:\FileSystem Test Data\Insurance\Car Insurance		/	Insurance	Car	Create
5		E:\FileSystem Test Data\Insurance\Travel Insurance		/	Insurance	Travel	Create
6		E:\FileSystem Test Data\Insurance\Health Insurance\Claims		SubSite	Claims	Health/Claims Docs	Create
7		E:\FileSystem Test Data\Insurance\Car Insurance\Claims		/	Claims	Car/Claims Docs	Create
8		E:\FileSystem Test Data\Insurance\Travel Insurance\Claims		/	Claims	Travel/Calims Docs	Create
9		E:\FileSystem Test Data\Insurance\Insurance Archive		Archive	Archive	Archive Docs	Merge
10		E:\FileSystem Test Data\Insurance\Others		Archive	Archive	/	Create

FIGURE 75 - EXAMPLE STRUCTURE MAPPING FROM FILE SHARE

In this example, assume that the SiteCollection has been specified as **AcmeSure** in the Target Connection in the Migration Job. This means it is not required in the structure mapping. This is a FileShare migration and so the SourcePath field is used to specify the source path. By using the download structure mapping template feature the SourcePath will be prepopulated.

Row 1

- TargetSubSite is set to **/** since not migrating to a Subsite
- TargetLibrary contains name of Document Library **Insurance**
- TargetFolder is set to **/** to migrate to the root of the Insurance Document Library

- TargetRootSetting is set to **Create**. This means the last folder Path in SourcePath will be created as a root.

Result: -
Items in “E:\File System Test Data\Insurance” are migrated to “AcmeSure\Insurance\Insurance\”

Row 2
As above, TargetFolder is specified so items from Health Insurance folder in the Source are migrated to **Health** folder with the Insurance DocumentLibrary.
Result: -
Items in “E:\File System Test Data\Insurance\Health Insurance” are migrated to “AcmeSure\Insurance\Health\Health Insurance”

See how the **Create** TargetRootSetting means that the “Health Insurance” folder is created below the Health Target Folder.

Row 9 – Example using MERGE
- TargetSubSite is set to **Archive** so migration is to SubSite called **Archive** which will be created if it does not exist.
- TargetLibrary contains name of Document Library **Archive**
- TargetFolder is set to **Archive Docs** to migrate to the Archive Docs folder (which will be created within the Archive Sub Site if it doesn’t exist)
- TargetRootSetting is set to **Merge**. This means items in the root of the source folder will be migrated into the root of the target.

Result: -
“E:\File System Test Data\Insurance\Insurance Archive” are migrated to “AcmeSure\Archive\Archive\Archive Docs”

Note how because TargetRootSetting is set to Merge, “Insurance Archive” is not in the target path. If it was set to Create, items would be migrated
E:\File System Test Data\Insurance\Insurance Archive migrated to AcmeSure\Archive\Archive\Archive Docs\Insurance Archive

Below is an example where the Target connection is at the Tenant level and no specific Site Collection is specified so the **TargetSiteCollection** is mandatory as it specifies the Site Collection the items should be migrated to.

SourcePat	SourcePath	TargetSiteCollection	TargetSubSite	TargetLibrary	TargetFolder	TargetRootSetting
	FileSystem\Test Data\Insurance	https://demo.sharepoint.com/sites/Insurance	/	Insurance	/	Merge
	FileSystem\Test Data\Insurance\Health Insurance	https://demo.sharepoint.com/sites/Health Insurance	/	Insurance	Health	Create
	FileSystem\Test Data\Insurance\Car Insurance	https://demo.sharepoint.com/sites/Car Insurance	/	Insurance	Car	Create
	FileSystem\Test Data\Insurance\Travel Insurance	https://demo.sharepoint.com/sites/Travel Insurance	/	Insurance	Travel	Create
	FileSystem\Test Data\Insurance\Health Insurance\Claims	https://demo.sharepoint.com/sites/Health Insurance	Claims	Claims	Health/Claims Docs	Create
	FileSystem\Test Data\Insurance\Car Insurance\Claims	https://demo.sharepoint.com/sites/Car Insurance	/	Claims	Car/Claims Docs	Create
	FileSystem\Test Data\Insurance\Travel Insurance\Claims	https://demo.sharepoint.com/sites/Insurance	Claims	Claims	Travel/Calims Docs	Create
	FileSystem\Test Data\Insurance\Insurance Archive	https://demo.sharepoint.com/sites/Insurance	Archive	Archive	Archive Docs	Merge
	FileSystem\Test Data\Insurance\Others	https://demo.sharepoint.com/sites/Insurance	Archive	Archive	/	Create

FIGURE 76 - EXAMPLE STRUCTURE MAPPING FOR FILESHARE WHEN CONNECTING TO TENANT

Example Non FileShare/ShareFile Structure Mapping

For migrations other than ShareFile and File Share, structure mapping entries require that the SourcePathID is used to identify the target path. By using the download structure mapping template feature the SourcePathID will be prepopulated.

SourcePathId	SourcePath	TargetSiteCollection	TargetSubSite	TargetLibrary	TargetFolder	TargetRootSetting
D3FCBE579BEC5E526455DB6E6DE2B6CD			/	Insurance	/	Create
05AF79670949F0E3A1224ABE7238EAC8			/	Insurance	Health	Create
E0BA90466FCC529D8CC6ABFF8D64F9B2			/	Insurance	Car	Create
60BA5B81972624E46C48BB6C567F1184			/	Insurance	Travel	Create
87FF02A6C7F26144B5763E476923269F			SubSite	Claims	Health/Claims Docs	Create
656458C6194DDC87C1F2A6F5A6DF8496			/	Claims	Car/Claims Docs	Create
CEC49D75D30181E34FE0056AA8FC3C58			/	Claims	Travel/Calims Docs	Create
783C72F45664A9A07E85FEDC628669E5			Archive	Archive	Archive Docs	Merge
4FC73A3D3FAA9A468F05F515D17325CD			Archive	Archive	/	Create

FIGURE 77 - EXAMPLE STRUCTURE MAPPING USING SOURCEPATHID

SourcePathId	SourcePathType	SourcePath	TargetSiteCollection	TargetSubSite	TargetLibrary	TargetFolder	TargetRootSetting
090007c280004646			https://demo.sharepoint.com/sites/Insurance	/	Insurance	/	Merge
090007c280004b17			https://demo.sharepoint.com/sites/Health Insurance	/	Insurance	Health	Create
090007c280004b6c			https://demo.sharepoint.com/sites/Car Insurance	/	Insurance	Car	Create
090007c280005135			https://demo.sharepoint.com/sites/Travel Insurance	/	Insurance	Travel	Create
090007c280004604			https://demo.sharepoint.com/sites/Health Insurance	Claims	Claims	Health/Claims Docs	Create
090007c280004604			https://demo.sharepoint.com/sites/Car Insurance	/	Claims	Car/Claims Docs	Create
090007c280004604			https://demo.sharepoint.com/sites/Insurance	Claims	Claims	Travel/Calims Docs	Create
090007c280004b1c			https://demo.sharepoint.com/sites/Insurance	Archive	Archive	Archive Docs	Merge
090007c280004b19			https://demo.sharepoint.com/sites/Insurance	Archive	Archive	/	Create

FIGURE 78 - EXAMPLE STRUCTURE MAPPING USING SOURCEPATHID WHEN CONNECTING TO TENANT

Example Structure Mapping to OneDrive

When migrating to OneDrive, there needs to be a structure mapping entry for each user. The TargetSiteCollection URL is usually of the form **https://<tenant name>-my.sharepoint.com/personal/<user principal name>**

SourcePathId	SourcePath	TargetSiteCollection	TargetSubSite	TargetLibrary	TargetFolder	TargetRootSetting
-1	Personal Folders	https://acmecorp-my.sharepoint.com/	/	Documents	/	MERGE
foh9eb5c-514c-4946-a18d-a212f2ada3d9	Personal Folders/Percy Shelley	https://acmecorp-my.sharepoint.com/personal/percy_Shelley_acmecorp_com	/	Documents	/	MERGE

FIGURE 79 - EXAMPLE STRUCTURE MAPPING TO ONEDRIVE

9.3. Restructure using Item Metadata

This feature allows the target location for items to be based upon rules which use source item metadata properties. For example, documents should go to a folder called 'Financial Records' if the value of a Department property = 'Financial'. In the examples discussed below, only the Target Folder is being assigned for the item, but just like normal structure mapping, the mapping can define Site Collection, Library and Folder within the target.

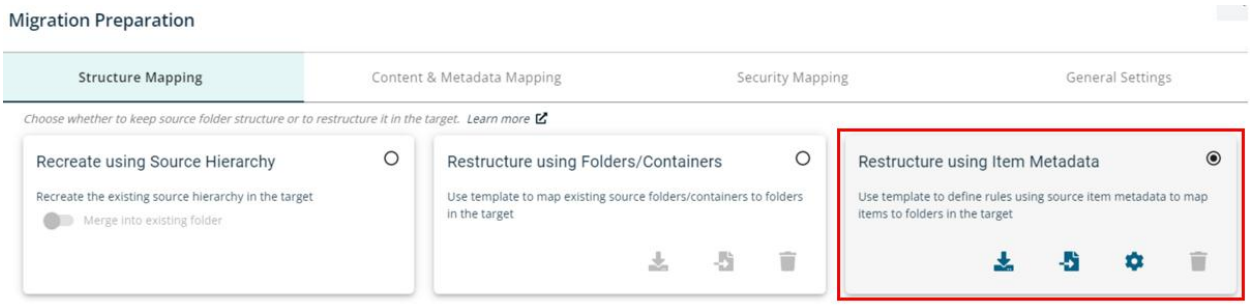


FIGURE 80 - RESTRUCUTURE USING ITEM METADATA

NOTE: The Pseudo code examples below are not how the rules are structured but serve to explain the capabilities that this feature can provide.

Example 1:

```
 IF Department=" Finance" AND CustomerType="Top50"
THEN
        Target Folder="Large Accounts Finance Data"
```

Example 2:

```
Target Folder="All Non Financials" ;  (You can define a 'catch all' target folder)
IF Department="Finance" AND CustomerType="Top50"
THEN
      Target Folder="Finance Large Accounts"
ELSE
     IF Department="Finance"
     THEN
           Target Folder="Finance Small Accounts"
     ENDIF
ENDIF
```

NOTE: When using this feature, it's essential to take a two pass approach to the migration. When in the migration phase, first execute **up to Step 3-Transform ONLY.** This will generate the necessary information about the location items will be migrated to. After this has been executed for **all items**, then Step 4 – Load can be executed.

9.3.1. Configuration Steps for Restructuring using Item Metadata

Metadata structure mapping works by defining a 'Target Assignment Expression' whose resulting output is examined for a match in the structure mapping table to determine the target to assign the item to.

Configuration Steps:-

1. Specify Target Assignment Expression
2. Populate the structure mapping table with values to compare with the output from step 1 and the target location.

The process to configure it will be explained by the example below.

Step1: Defining Target assignment expression

Scenario: Assign items to Target based upon the value of source metadata field 'ItemDepartment'. The name of source metadata fields that can be used are visible in Populating Target metadata field form.

Example Target Assignment Expression

```
Department='{ItemDepartment}'
```

Example Source Item Metadata

Item No	ItemDepartment	Result from Expression Evaluation
1	**HR**	Department ='HR'
2	**Finance**	Department ='Finance'
3	**Security**	Department ='Security'

Example Structure Mappings

SourcePath	TargetPath
Department ='HR'	HR Items
Department ='Finance'	Finance Items

When Item No. 1 is processed, during the Transform step, the Target Assignment Expression is read, the value from the metadata field *ItemDepartment* is retrieved and replaces the field name in the expression to generate the result, Department='HR' This is then used to scan for a matching entry in the structure mappings table and a row with the matching value is found so the TargetPath is read and that item is assigned with a TargetPath of "HR Items".

Similarly for Item No. 2,the result is Department='Finance' and the item is assigned to "Finance Items".

Item 3 would generate a result of Department='Security' but wouldn't find a match in the structure mapping table and so the item would fail to be migrated because no target would be assigned.

In Target Assignment Expressions, the variable name shown in this example 'Department' can a be any string, what's essential is that the output from the expression is in the structure mapping table.

NOTE: You cannot use comparators like "<>" nor Boolean operations like "OR" etc. in these Expressions. In these examples comparators = and Booleans like AND are just text to make the expression human readable. It's not the case that they're used during expression evaluation. Target Folders Assignment expressions are simply processed by replacing field labels with actual metadata values and the output value used to find exact matching entry in structure mapping.

9.3.2. Advanced Target Assignment Expressions

Target Assignment expressions support have additional capabilities.

In this second example Finance Items related to Top50 customers goes to Finance Large Accounts, otherwise Finance Items go to Finance Small Accounts. All other items will go to the "All Non Financials" folder.

In the example expression below, the first criteria is based upon two fields and then a second separate criteria is specified after the semi-colon.

```
Department='{ItemDepartment}' AND CustomerType='{ItemCustomer}';
Department='{ItemDepartment}'
```

Example Source Item Metadata

Item No	ItemDepartment	CustomerType	Output from Expression
1	**HR**		Department ='HR'
2	**Finance**	Top 50	Department ='Finance' AND CustomerType='Top 50';Department='Finance'
3	**Security**		Department ='Security'
4	**Finance**		Department ='Finance' AND CustomerType='';Department='Finance'

Example Structure Mappings

SourcePath	TargetPath
Department ='HR'	HR Items
Department ='Finance' AND CustomerType='Top 50'	Finance Large Accounts
Department ='Finance'	Finance Small Accounts

When Item 2 is processed the expression is evaluated and the first element of the output "Department ='Finance' AND CustomerType='Top 50'" is used to lookup on Structure Mapping and so assigns item to TargetPath 'Finance Large Accounts'.

For item 4, the Customer Type is empty so the output from first expression is Department ='Finance' AND CustomerType='' which has no matching entry in structure mapping so the next element of the output Department='Finance' is used for structure mapping lookup and so items are assigned to the TargetPath of Finance Small Account.

Target Assignment Expression order

It's essential to understand that expressions are processed from left to right, so in the above example the more 'specific' criteria is evaluated first. Only if an item doesn't meet the first criteria for 'Department AND CustomerType', is it then evaluated based on Department only.

Example of two expressions in the **WRONG** order.

```
Department='{ItemDepartment}' ; Department='{ItemDepartment}' AND CustomerType='{ItemCustomer}'
```

This is evaluated as: -

Department='Finance' and then

Department='Finance' AND CustomerType='Top50'

In this example every 'Finance' record would always meet the first criteria since it's broader, even though the more specific second rule is what is intended for some items.

9.3.3. Default Target Assignment

When using criteria to assign folders it's possible that some folders may not match any of the criteria specified. To cover this scenario, a 'Default' folder can be assigned to such items. If no 'Default' folder is configured, then these items will not be migrated.

Scenario: Items which don't match any specific values in the structure mapping table should be assigned a specific target location

Effectively this is just a case of creating criteria that is evaluated last of all and will always have a matching entry in the structure mapping table.

Department='{ItemDepartment}' AND Customer='{ItemCustomer}' **;**
Department='{ItemDepartment}'**;Default**

Example Source item metadata

Item No	ItemDepartment	Result of Evaluation

1	**HR**	Department =’HR’
2	**Finance**	Department =’Finance’
3	**Security**	Department =’Security’

In the structure mapping specify in the field SourcePath an entry for Default and a target value.

Example Structure Mappings

SourcePath	Target Folder
Department =’HR’	HR Items
Department =’Finance’	Finance Items
Default	All Non Financials

In this example, when an item that’s not in ‘HR’ or ‘Finance’ is processed the last entry is processed which is simple the string value of ‘Default’. This ‘result’ is used to scan the structure mapping table and finds a matching entry and assigns the item to that target folder specified. It will always find a match on ‘Default’ so items from the ‘Security’ department will be assigned to a target folder of “All non Financials”

Example Metadata_Expressions_Structure_Mapping.csv

SourcePathId	SourcePath	TargetSiteCollection	TargetSubSite	TargetLibrary	TargetFolder	TargetRootSetting
	DEFAULT		/	Insurance	/	Create
	Matter='Insurance' AND SubMatter='Car insurance'		/	Insurance	Car Insurance	Create
	Matter='Insurance' AND SubMatter='Car insurance Claims'		Claims	Insurance	Car Insurance/Claims	Create
	Matter='Insurance' AND SubMatter='Home Insurance'		/	Insurance	Home Insurance	Create
	Matter='Insurance' AND SubMatter='Property Insurance'		/	Insurance	Property Insurance	Create
	Matter='Insurance' AND SubMatter='Marine Insurance'		/	Insurance	Marine Insurance	Create
	Matter='Insurance' AND SubMatter='Fire Insurance'		/	Insurance	Fire Insurance	Create
	Matter='Insurance' AND SubMatter='Claims'		Claims	Insurance Claims	/	Merge

FIGURE 81 - EXAMPLE ITEM METADATA STRUCTURE MAPPING

9.3.4. Metadata Structure Mapping Configuration Steps

These are the steps to follow to configure Metadata Structure Mapping.

1-Specify Target Assignment Expression

Select the ⚙ option on the Restructure using Item Metadata

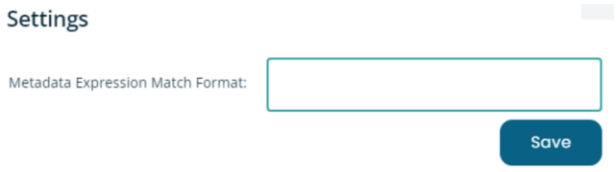


FIGURE 82 - ENTER ITEM METADATA STRUCTURE MAPPING

Enter the expression as required. e.g.

```
Department=’{ItemDepartment}’ AND CustomerType=’{ItemCustomer}’ ;
Department=’{ItemDepartment}’;Default
```

Once the above file is changed it’s essential for the CPS application is restarted.

2-Download Template

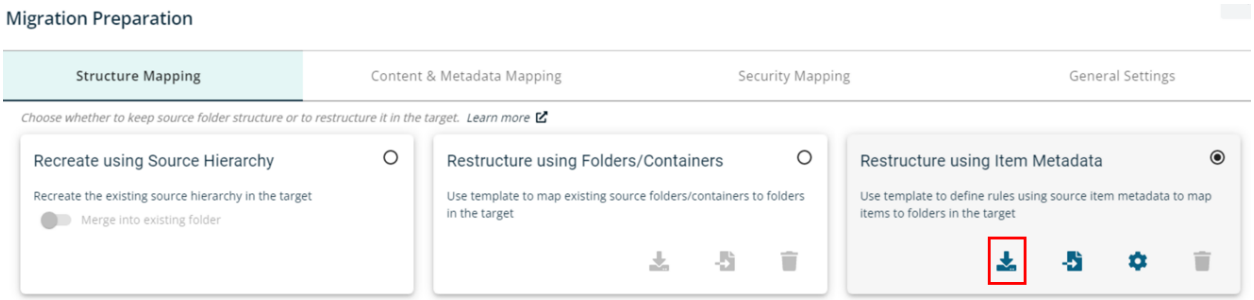


FIGURE 83 - DOWNLOADING TEMPLATE FOR ITEM META RESTRUCTURING

Select the **Restructure using Item metadata** and select the icon to download the template to use.

In the template the evaluated expression result should be placed in the SourcePath column. All other target related columns should be completed as required. See Restructure using Folders/Containers for more information.

Once the structure mapping file is complete, select the **Import** icon to import the file. The contents of the structure mapping will be displayed.

10. PHASE2 – CONTENT TYPES

Mapping Content Types

For details on what are content types, what is content type mapping see 4.2. Migration Terms and Concepts.

10.1. Until a Source Content Type is mapped, items of that type cannot be migrated. Automatic content type mapping will occur for 'well known' content types.

In this example there is an unmapped Content Type in the Source named 'File'. So, we are going to map 'File' to a Content Type of 'Document' in the Target.

Steps:

Click on Content & Metadata Mapping. The left pane shows **Unmapped Content Types** for the source, the right pane shows **Mapped Content Types**.

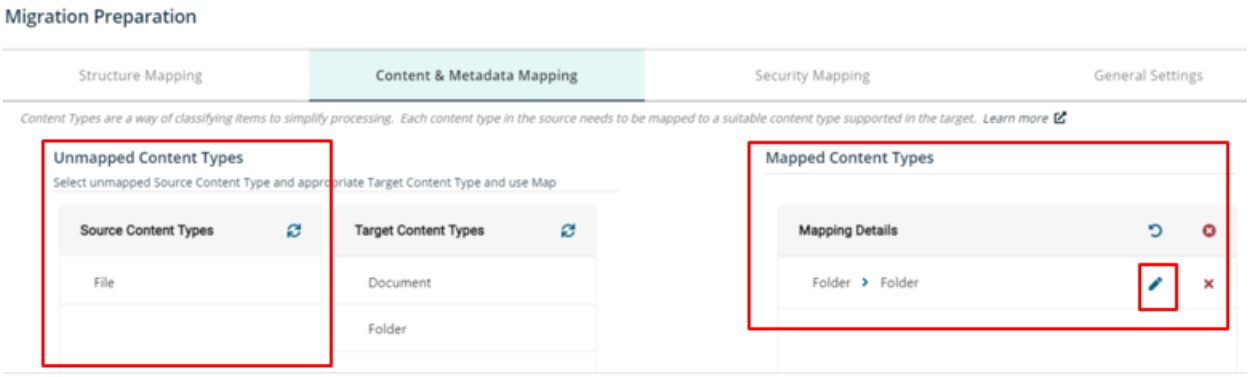


FIGURE 84 - MAPPING SOURCE TO TARGET CONTENT TYPE

Click on 'File' and click on 'Document' in the Target and select the **Map** button. The entry File>Document will appear in the Mapped Content Types table. The exact details of how metadata fields are moved between these mapped content types can be reviewed and edited by selecting the icon to display the Target Metadata Mapping.

10.1.1. Reviewing default content type mappings

CPS has an automatic content type mapping feature, but these should still be reviewed carefully because, even if a source content type is mapped, that doesn't mean every attribute on that type is assigned to a target field.

10.2. Target Metadata Mapping

Target metadata mapping is the process of defining for each target content type, how its metadata fields are populated. In many cases a field may be populated with the relevant metadata field from the source item, but this feature provides the opportunity to alter the data during the migration if required.

Some customers take the opportunity to simplify their data and reduce the number of content types and/or content of fields during migrations.

A target content type may differ from the source content type in several ways:

- Number of metadata fields
- Valid field values
- Mandatory v Optional

CPS provides a rich set of capabilities to smoothly migrate metadata in such cases.

Select the appropriate Content Type mapping from the Mapped Content Types to configure mapping.

FIGURE 85 - MAPPED CONTENT TYPES LIST

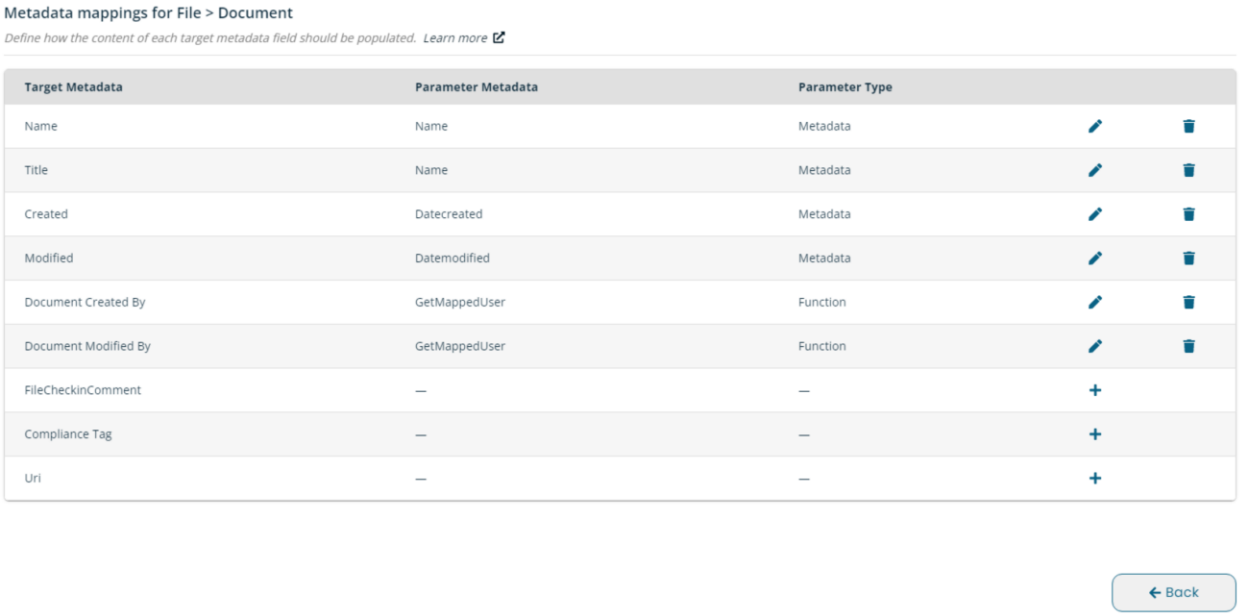


FIGURE 86 - CONFIGURING POPULATION OF TARGET METADATA

Metadata mapping configuration displays a table of target fields. For each field, the Mapping type displays the method used to populate the field (i.e. Metadata, Constant or Function) and the Source Metadata contain information on where the value comes from (either source field name, a constant/fixed value or name of function as appliable). Use + to add new mapping to a target metadata property or ✎ to edit existing mapping.

For each Target Metadata field, there are 3 different methods which can be used to populate it.

- **Source Metadata** - The target field is populated using the value from a specified source metadata field.
- **Constant** - The target field is populated with a specified fixed value.
- **Function** - The target field is populated using the output of a function or PowerShell script. Both of which may use source properties or other values to help generate the value to populate the target field. For example, populate the target by prefixing a string to the content in source field 'RoomID'. For the full list of functions available See 19.1. Built-in Functions for details.

Term stores

The setting of term store fields is supported, if multiple values need to be assigned they should be separates using ;# e.g. Term store 1;# Term store 2.

Hidden SPO fields

When migrating to SPO by default, all fields are shown, irrespective of their hidden status.

To alter this so that hidden fields are not displayed in metadata mapping, alter the CPS configuration file, setting the configuration key to True and restarting the CPS client.

```
<add key="SharePoint.IsDiscoverHiddenMetaFields" value="True" />
```

Sensitivity Labels

If setting nested sensitivity labels on items, they should be of the format:-
Test Auto Labelling Parent \ Copy Restriction

More details can be found here https://learn.microsoft.com/en-us/information-protection/develop/concept-supported-filetypes#file-types-supported-for-protection

10.2.1. Populating Target using Source Metadata

Selecting the Source metadata field will display list of source metadata field available. In this example below, the target metadata field Created is being populated by the DateCreated source metadata property.

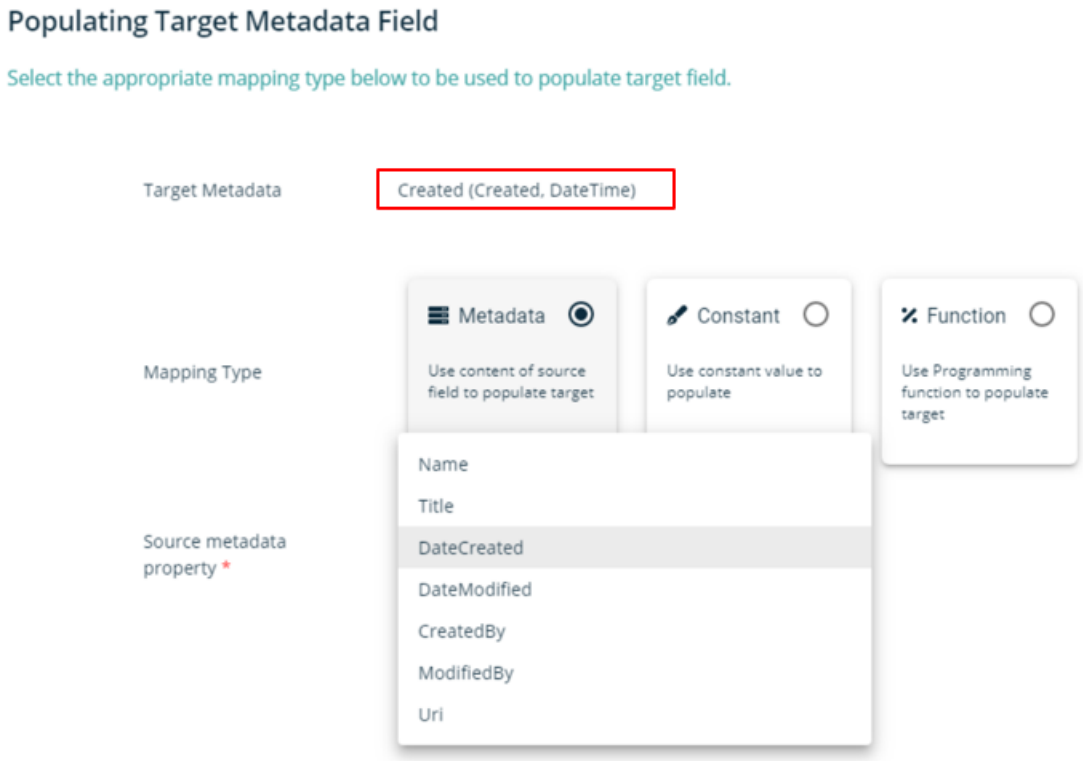


FIGURE 87 - POPULATING TARGET USING SOURCE METADATA

Selecting the Source metadata field will display list of source metadata field available. In this example above, the target metadata field Created is being populated by the DateCreated source metadata property.

10.2.2. Populating Target using Constant value

Use this option to specify a Constant(fixed)value for a target metadata field.

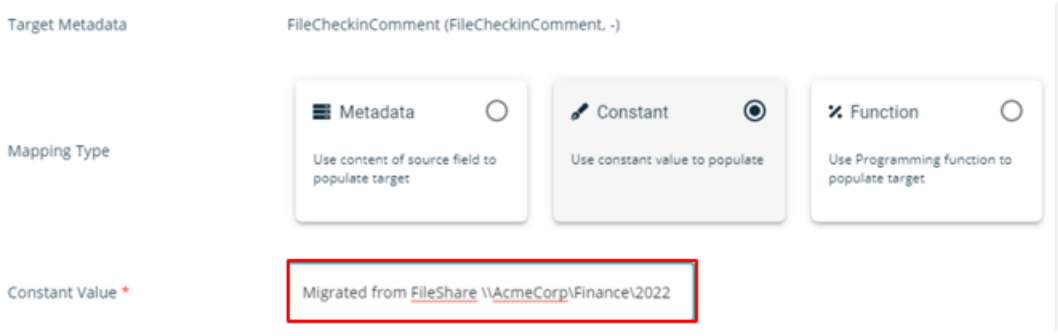


FIGURE 88 - POPULATING TARGET USING FIXED/CONSTANT VALUE

10.2.3. Populating Target using Functions

A function can take in one or more parameters and use these to generate a resulting output which is used to populate a target metadata field. The function can either be a built-in function or a custom created function using PowerShell.

> NOTE: Migration performance may be significantly impacted if using PowerShell functions, especially if processing .CSV files due to the time take to execute logic for each item. So script performance should be optimized where possible, especially for larger migrations.

Target metadata population using built-in functions

Choose the **Function** option and then the **Select** option

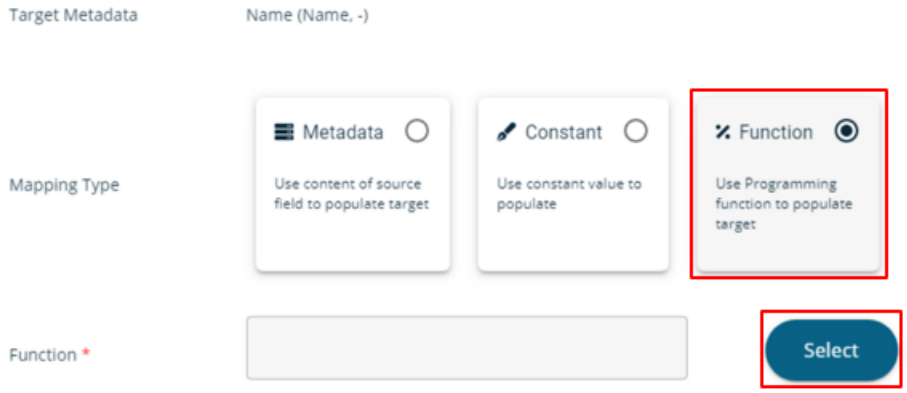


FIGURE 89 - POPULATING TARGET USING OUTPUT OF A BUILT-IN FUNCTION

A list of function areas is displayed broken down into the following categories.

Misc, StringFunctions, Math, Date & Time, Drupal8, OCR and CustomScripts

A full list of all the functions can be found here 19.1. Built-in Functions

Expand the relevant category and select the function to use and then press the **Select Function** button.

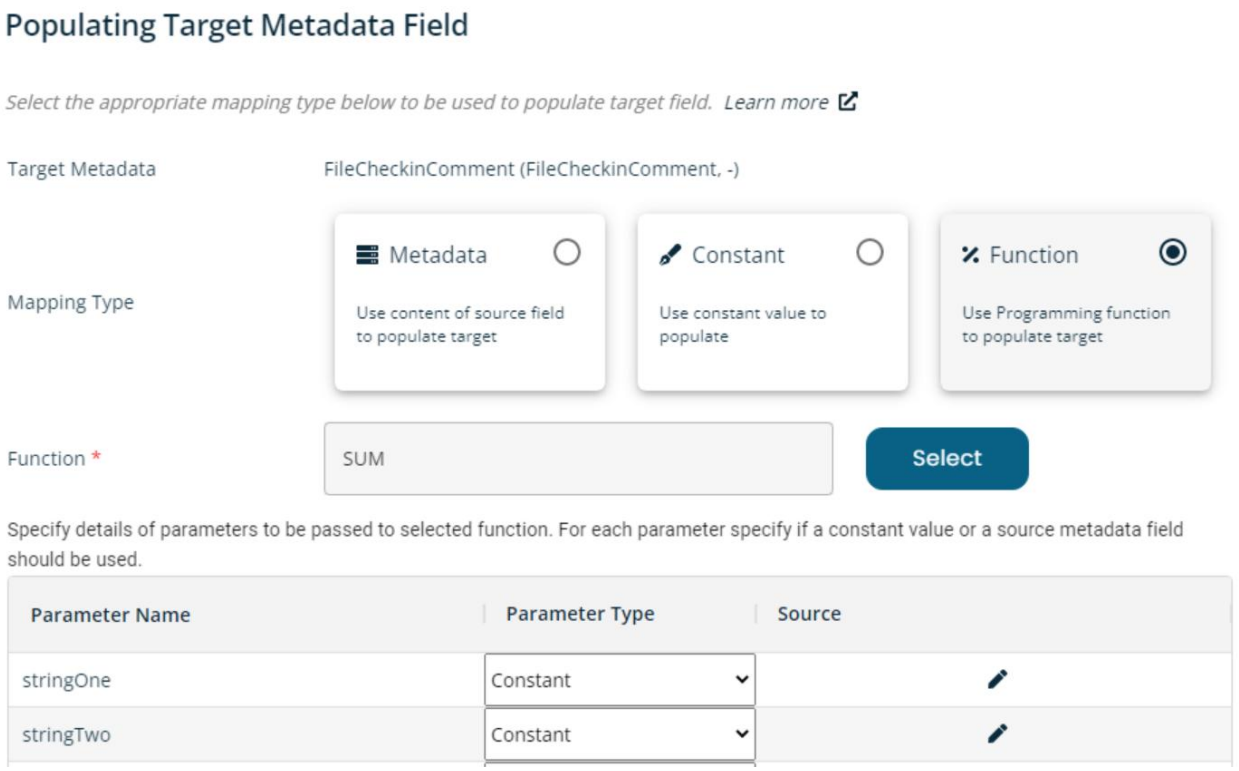


FIGURE 90 - CONFIGURING PARAMETERS TO PASS TO FUNCTION

Once the function is selected, the next step is to configure how values are passed into that function. A table is displayed showing each input parameter that the function requires.

The **Parameter Type** defines how the parameter is populated. This can either be **Constant** or **Metadata.** The source column should either contain the value for the Constant or the source metadata field name.

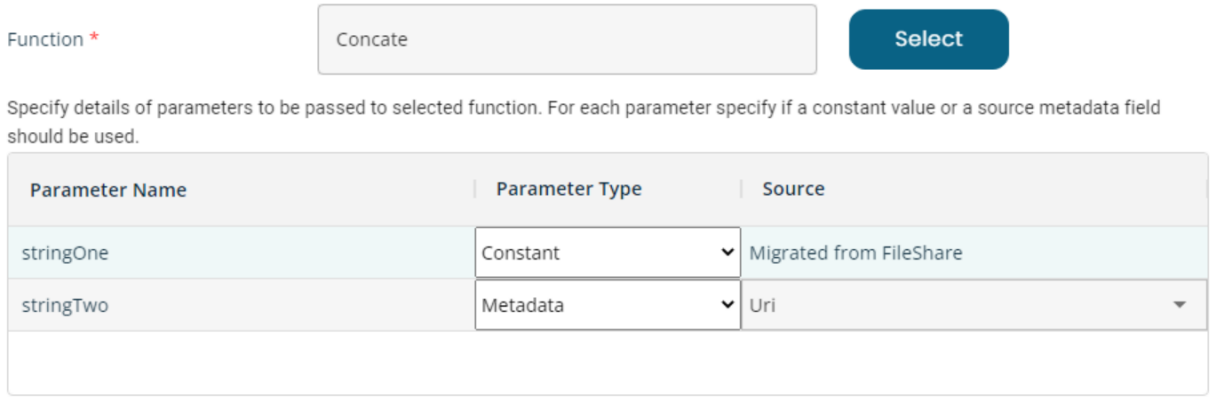


FIGURE 91 - EXAMPLE OF PARAMETERS PASSED TO FUNCTION

In this example a target field is being populated using the Concate function to join a fixed string value and the URI source metadata field.

GetMappedUser Function

Target fields of type User or Group should be mapped using the function GetMappedUser(as they are by default). This ensures that these field types are populated correctly using the information with the security mapping table.

Editing Built-in functions

All the built-in functions can be edited using the icon with the exception of those in the Misc Category.

NOTE: There is only one instance of these functions per server, so editing a function used by one job may have unintended consequences on other migration jobs.

FIGURE 92 - SELECTING BUILT-IN FUNCTION TO EDIT

This will display details of the function including the parameters and the PowerShell that implements the function. Each parameter can be referred to in the script as it's parameter name. So for example, a parameter called JobTitle will be accessible in the code as $JobTitle.

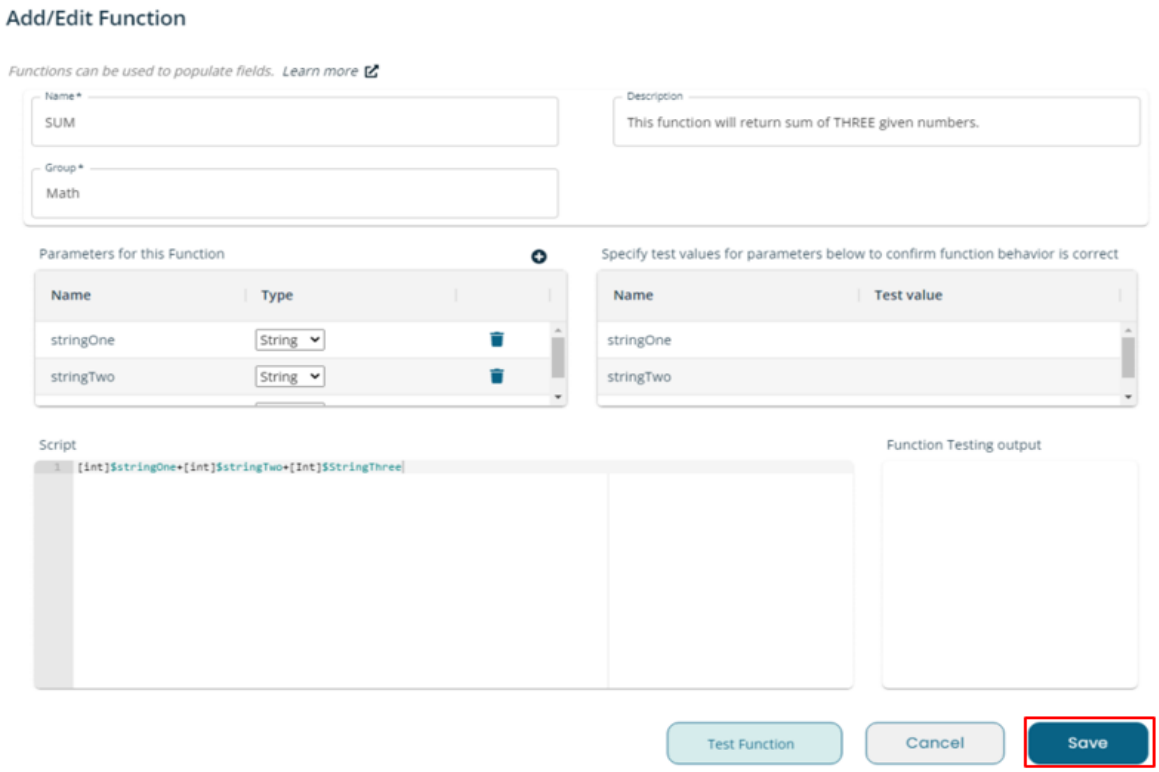


FIGURE 93 - ENTERING TEST PARAMETERS TO TEST FUNCTIONS

The built-in SUM function normally adds two values. In the example above, we've altered it to add 3 numbers. The new parameter has been named 'StringThree' and is available in the script as a variable $StringThree.

All the parameters will be listed in the Test Values section where test values can be manually specified. Use the **Test Function** option to execute the script using the value parameters specified to ensure it works as required.

10.2.4. Populating Target using PowerShell

As well as using built in functions it's possible to create your own functions. This feature is a powerful way to populate the contents of target metadata fields as it allows the output of a customer written PowerShell function to populate a specified metadata field. As with the built-in functions, a customer defined list of parameters can be specified.

Creating New Function

First, select the field to be populated with the function. On Content Type and Meta data mapping, select the for the Mapped Content Type to be altered.

In the table of target metadata fields, select for the metadata field to be populated with PowerShell function.

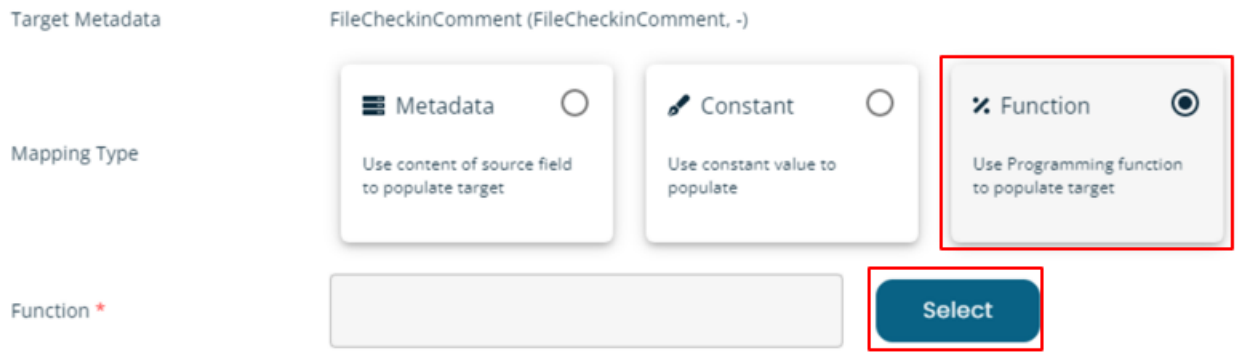


FIGURE 94 - POPULATING TARGET USING A NEW FUNCTION

On Mapping Type screen, Choose **Function** and then the **Select** button

On the Programming Functions screen select **Create New Function**

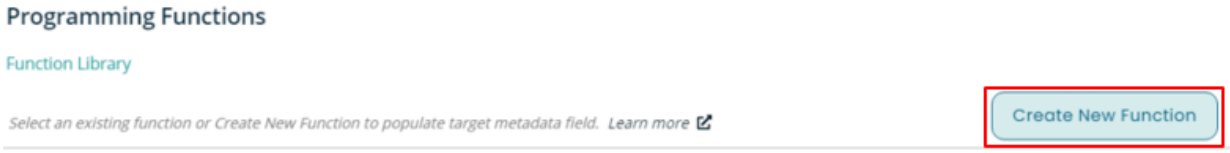


FIGURE 95 - CREATE NEW FUNCTION OPTION

New Function Definition

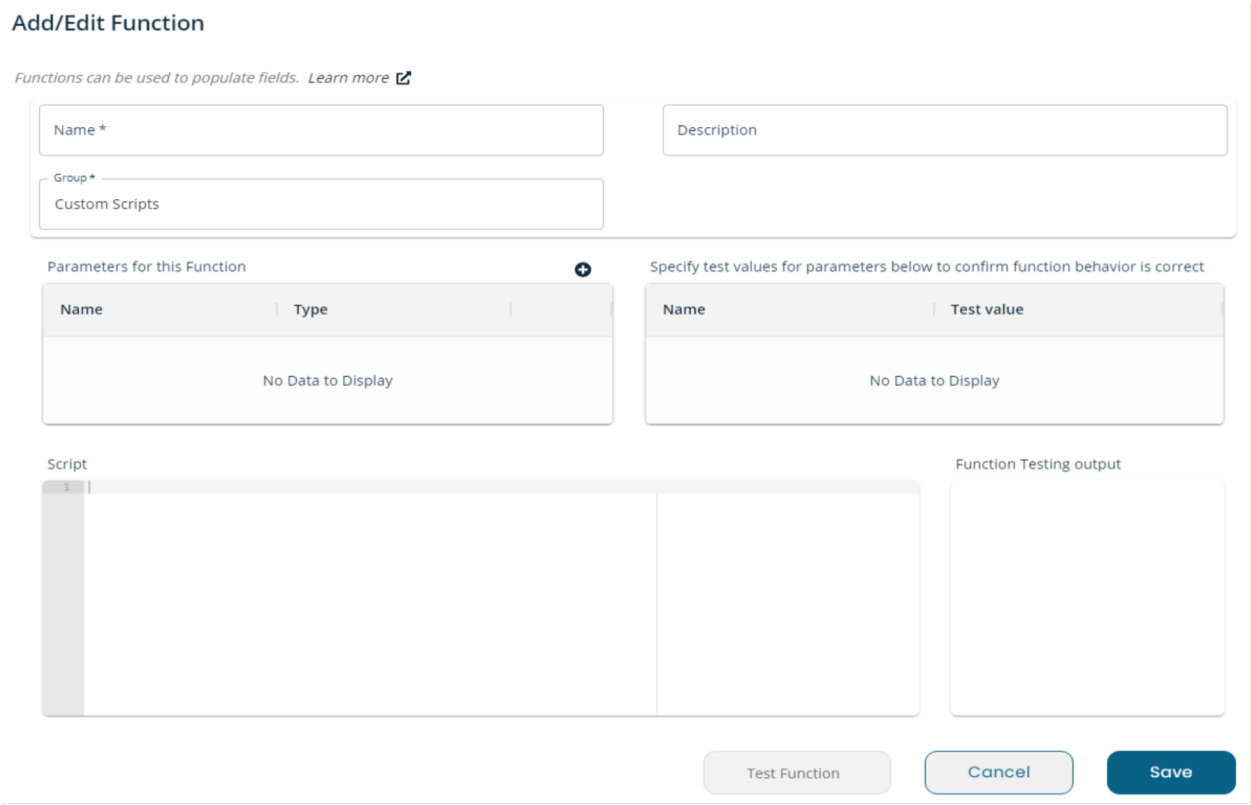


FIGURE 96 - DETAILS REQUIRED FOR NEW CUSTOM FUNCTION

Name: Name assigned for new function
Description: Description of function
Group: From the drop down, select **Custom Scripts**
Parameters for this function

Use the option to specify parameters that the function requires.

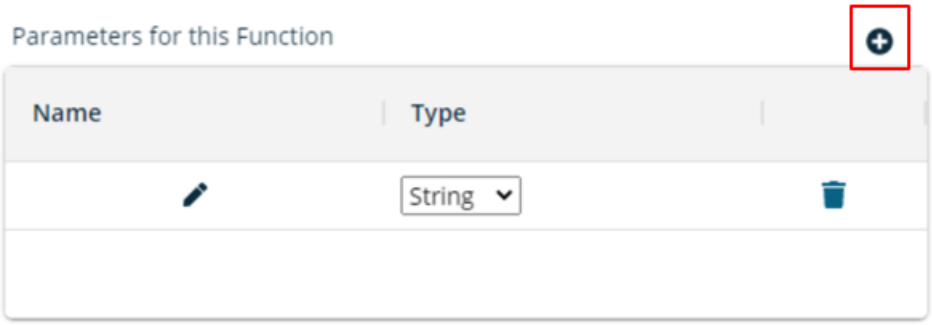


FIGURE 97 - DEFINING PARAMETERS FOR NEW FUNCTION

Use to specify the name and datatype for each function parameter.

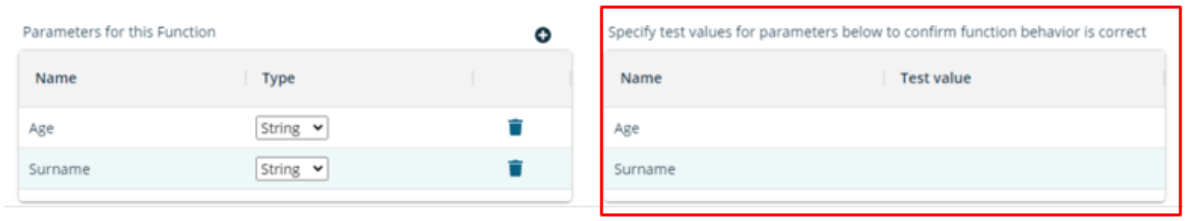


FIGURE 98 - ENTERING TEST VALUES FOR NEW FUNCTION

As parameters are specified, they will appear on the test parameters area. Use the option to enter test values for these parameters to pass to the PowerShell script. Use the **Test Function** option to execute the script using the test parameters. The output of which will appear in the Output window.

Example PowerShell Function

```
$age_bracket=Switch($Age)
{
   {$_ -le 10}                    {"Very Young"}
   {$_ -gt 10 -and $_ -le 17}     {"Young"}
   {$_ -ge 18 -and $_ -le 100}  {"Adult"}
   {$_ -gt 100}                     {"Very old!"}
}
```

```
return $age_bracket + "-" +$surname
```

The PowerShell script will have access to the passed in parameters So $Age and $Surname are variables which are available for example to use. It will use the $Age parameter to determine an 'Age Bracket' and concatenate that with the Surname to return this as the return value of the function which is then used to populate the Target metadata field.

Working Example

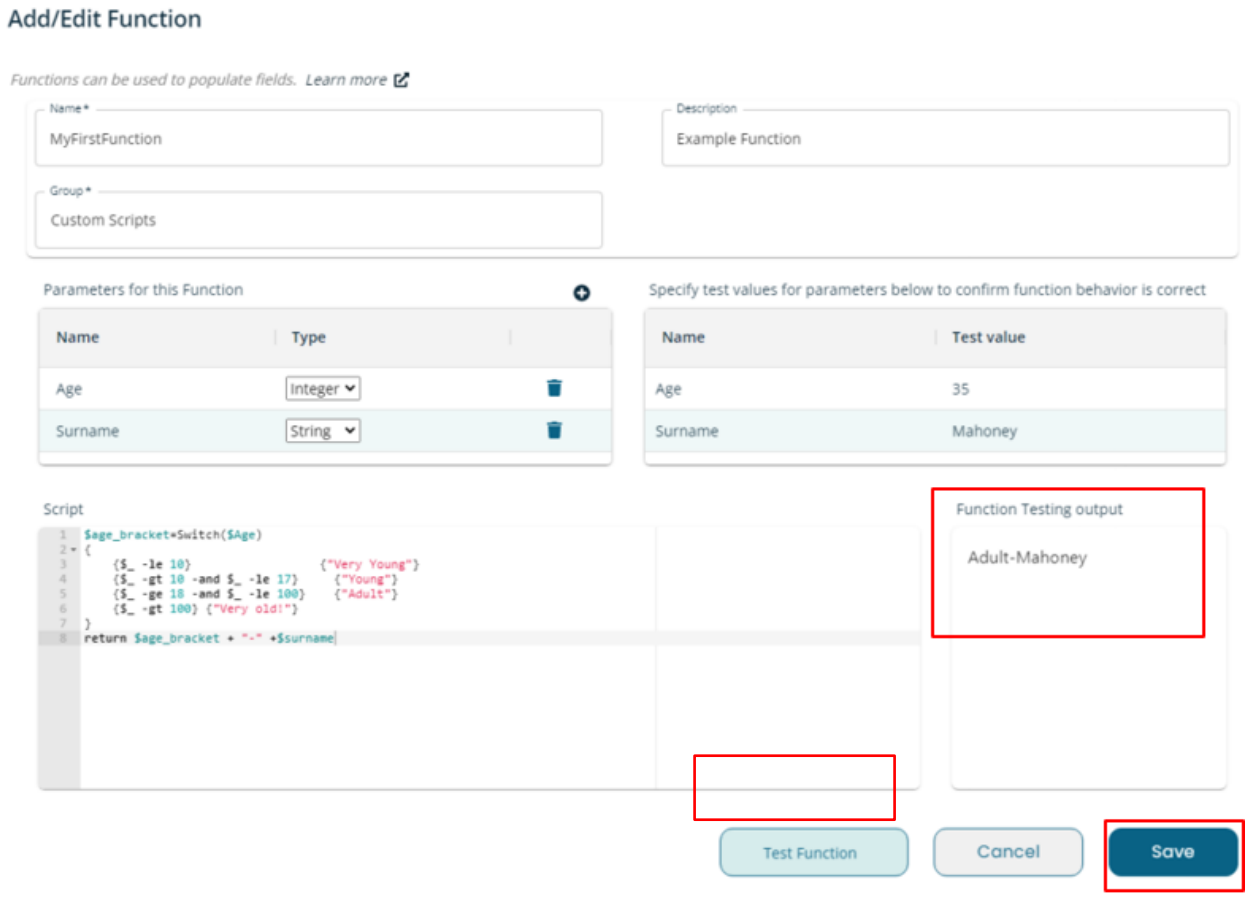


FIGURE 99 - EXAMPLE OF FULLY DEFINED CUSTOM FUNCTION

Use **Save** once the function is completed and then the next step is to configure how values are passed to the function. The table below will be displayed/

Use the Parameter Type field to specify whether the parameter should be populated with a constant value or a source metadata field. Use the ✎ to specify details.

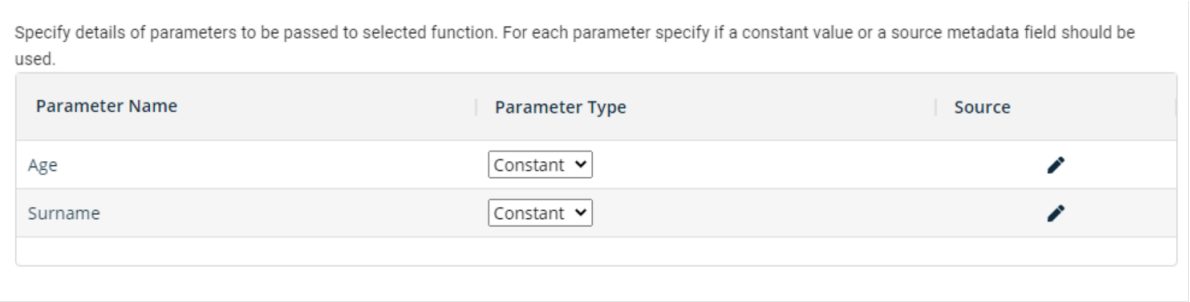


FIGURE 100 - CONFIGURING HOW PARAMETERS ARE PASSED TO CUSTOM FUNCTION

When using metadata type mapping, selecting the Source field will display a dropdown list of all the metadata fields for the relevant source content type.

10.2.5. Using Document ID Service

If the target SharePoint site has Document ID service feature enabled, then CPS will also allow you to map Document Id and Document ID URL fields during migration.

To add metadata mapping for Document ID and Document ID URL fields follow the steps given below:

- Open the content type mapping Settings.
- Select any document content type mapping.
- Select the Document ID Target metadata field and map it to equivalent unique source metadata field or if none exists, a custom PowerShell function could be written to auto generate a unique Document ID number.
- Document ID URL (Optional). By default, when a Shortcut Link is migrated, the URL of the item it points to will be the exact target path the referred to item was/will be migrated to. That means if the target item is later moved, the Shortcut will no longer work. If using the SharePoint Document ID service and, specifying this Document ID URL then this Document ID URL is populated into the Shortcut Link instead of the exact path. This then means that the Shortcut URL would still function even if the target item the Shortcut points to is moved.

A Custom Function can be used to generate unique Document Id URL for each document by adhering to below SharePoint Document ID URL format.

http://xxx.sharepoint.com/_layouts/15/DocIdRedir.aspx?ID=<DocID>,<DocID>

Where <DocID> is the name of the target metadata field storing the Document ID configured in steps above.

10.3. Custom Classification

Custom classification is a powerful feature to allow specific items to be assigned to a Target Content Type independent of what their source content type is. Customer defined criteria can be specified to identify such items. For example, a 'HR Items' custom classification can be

defined as any source items that have 5 digit number in the Name metadata field. Just like normal content types, these custom classifications then need to be mapped to a suitable target content type.

Example #1:

Scenario: In a File system migration, most files should be migrated 'as is' but for those items located in a specific folder the filename will contain an employee ID and there is a requirement for this employee ID to be used to populate a target metadata field for these items.

Solution: In this case, Custom Classification 'HR Items' can be created and used to detect these specific items and mapped to the target content type 'HR Items'. Afterwards, using the target metadata mapping feature for the 'HR Items' a target metadata field 'EmployeeID' can have a Function associated with it which would execute a customer written PowerShell script to process the filename source field parameter passed into it and parse it to return the employee ID which then populates the EmployeeID field.

In this way then, Custom Classification is being used to select the items which need to have their metadata properties processed in a different way to other items of the same source content type.

Example #2:

Documents and Images in ShareFile are saved as **Documents,** whereas **SharePoint Online** has a specific Content Type **Image.** Simply mapping items of Source content type **Documents** to **Image** target content type would mean images AND documents from ShareFile being migrated as Documents. Instead, we need a way to identify these images that are currently classified as Documents in ShareFile and process their metadata properties differently.

As many Custom Classifications can be created as required.

10.3.1. Creating a Custom Classification

In the steps below, a new Custom Classification will be created called 'Custom Images' and is defined by items in the source with file extension of .jpg or .png. This will be named 'Custom Images' and can then be mapped to the SharePoint target content type 'Image'.

To Add/Edit a Custom Classification select **Migration Preparation Phase** > **Content Type & Metadata Mapping.**

Click  next to Custom Classifications to open **Add/Edit Custom Classification** page

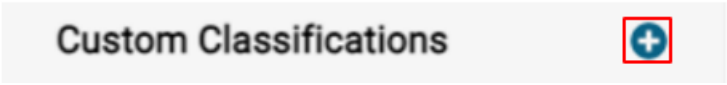


FIGURE 101 - CREATING NEW CUSTOM CLASSIFICATION

Custom Classifications can be edited or deleted by clicking on respectively.

The **Add/Edit Custom Classification** screen is displayed to define a new Custom Classification.

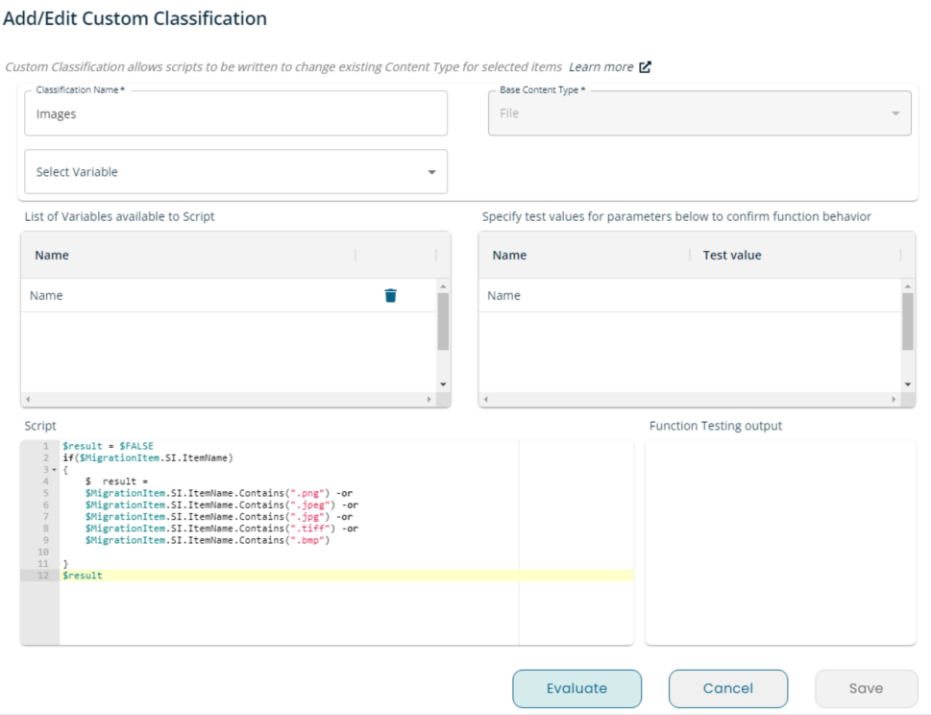


FIGURE 102 - EXAMPLE CUSTOM CLASSIFICATION

Classification Name: This is the name for this custom classification.
Base Content Type: When creating a classification, select an existing content type to use as a template for source metadata fields that will be available for mapping to target.
List of Variable available for Scripting: The variables listed will contain relevant source metadata properties based upon the selected base content type and can be used by the script. Each will then be available as their name in the script, so Title attribute is stored in $Title variable for example.
Specify test values: When developing the script, use this to specify test parameter values to confirm the PowerShell is working as expected.
Script: This is where the PowerShell script should be written. The script should evaluate the parameters passed to it and should return $true if the item should be assigned the associated custom content type, otherwise return $false.

Click **Evaluate** to run the PowerShell using the test values provided.
The output of the script will appear in the Function Testing output window.
Once testing is complete, use Save option to create the new custom classification.

The **Add/Edit Custom Classification** script should evaluate to **True** for custom classification to work properly. Anything that evaluates to **False** will skip the added Source classification settings.

Following the steps above, the Custom classification 'Custom Image' is then available to map to an appropriate Target Content Type (e.g. Image)

Note: It's essential to review the metadata mapping once the custom classification is mapped to a target content type.

Note: Custom classifications will only get executed if they have been mapped to a target content type.

Custom Classification Execution

The actual execution of these classification scripts will occur when running the Step 2 – Classify in the Migration Phrase. For each item, each of the script(s) will be executed till one returns a value of 'True'. So, for example if there are 3 Custom Classifications, then Step 2 Classify - Assign Target Content Type will run each script, passing to it the relevant item metadata values till one of the scripts returns 'True' or all scripts have been executed for the item. If none of the scripts return 'True' then the original Source Content Type is assigned to the item.

Note: The order in which the classification scripts are executed against each item is random. That means it's essential that these scripts do not 'overlap' one another i.e. two scripts could potentially return 'True' for the same item. When developing a script to identify an item if there are other custom classifications, it may be necessary to update those scripts to ensure that they will exclude such items.

10.3.2. Custom Classification Scripting Objects

The following objects can be used in the PowerShell script to classify items.

- **MigrationItem**
  - **SI** - The source representation of the item
    - **ItemName -** The name of the item as shown in the user interface.
    - **MimeType -** The file extension of the item.
    - **Uri -** The logical path of the item.
    - **AlternateUri -** The path to the binary if the item is a file.
    - **GroupId -** The document identifier that is common across all document versions.
    - **VersionLabel -** The major version.
    - **SubVersionLabel -** The minor version.
    - **EntityCategory -** 0 for containers, 1 for files and links.
    - **EntityType -** File, Link or Folder.

- **ContentType -** The content type classification of the item.
- **ItemSize -** The size of the content (in bytes).
- **Level -** The depth of the item in the folder tree.
- **HasUniqueSecurity -** Whether the item has security, information assigned.

  - **SIC -** The source metadata exposed as an XML string.

- **Log** - A logging object of type log4Net.ILog - this can be used to write messages to the application log with statement such as $Log.DebugFormat("Doing x for item with path {0}", $MigrationItem.SI.Uri);

Example Classification Scripts

Description	Example Script

Classifies Media files based upon Video or Audio content type	$result = $FALSE
	If ($migrationItem.SI.ItemName)
	{
	$result =
	$MigrationItem.SI.ContentType -eq 'Video' -or
	$MigrationItem.SI.ContentType -eq 'Audio'
	}
	• $result

Classifies Image items based upon file extensions (png,jpeg,jpg,tiff or bmp)	$result = $FALSE
	If ($MigrationItem.SI.ItemName)
	{
	$result =
	$MigrationItem.SI.ItemName.Contains(".png") -or
	$MigrationItem.SI.ItemName.Contains(".jpeg") -or
	$MigrationItem.SI.ItemName.Contains(".jpg") -or
	$MigrationItem.SI.ItemName.Contains(".tiff") -or
	$MigrationItem.SI.ItemName.Contains(".bmp")
	}
	$result

Classifies items based upon size and example of using Logging.	$result = $FALSE
	if ($MigrationItem.SI.ItemName)
	{
	if ($MigrationItem.SI.ContentType -eq 'File')
	{
	if ($MigrationItem.SI.ItemSize -gt 1048576)
	{
	$Log.DebugFormat("Large Item sized {1} for item {0}",
	$MigrationItem.SI.Uri,$MigrationItem.SI.ItemSize)
	$result = $true
	}
	Else

	{
	$Log.DebugFormat("Small Item sized {1} for item {0}", $MigrationItem.SI.Uri,$MigrationItem.SI.ItemSize)
	}
	}
	}
	$result

Classifies all files to a single content type	If ($MigrationItem.SI.EntityCategory -eq 1)
	{
	return $true;
	}
	else
	{
	return $false;
	}

11. Phase 2 – Security Mapping

Items in a source will have some kind of user or permission information associated with them and the security mapping feature can be used to help transfer this information into the target if necessary. For example, the Created By of a document in the source may be a name of a user but in the target, this needs to be in an email address format.

Security Mapping uses a table of source values which is edited to associate these entries with a new target value. This is typically achieved by editing a .CSV file which can be exported and imported into the client.

Security mapping is required because such information stored on the source may not be directly transferred 'as is' because it may not be compatible with the requirements of the target platform. For example a document author may be a simple text string in the source but must be an email address in the target.

Security Entities: User, Group, Permissions

11.1

User and Group mapping is mandatory step, it allows the conversion of User and Group data to something suitable in the target system so that if User "ShelleyP" or Group 'Senior Leadership' appears on the source item then replace this with a suitable user or group as specified in the mapping e.g. Percy.Shelley@acme.com. The result of the mapping will affect the CreatedBy/ModifiedBy fields in the target. Any user or groups not mapped will cause the relevant items to fail migration unless the 'Default' feature is used, see below for more details.

Permission mapping is an optional step. It's the process of mapping a type of access right on the source to a suitable access right on the target. For example, a user may have read access on a folder in the source and this needs to be mapped to the equivalent Read access in the target. This information is used with the user/group mapping so that User "ShelleyP" who has "Read" access on the source is represented as Percy.Shelley@acme.com with "Read" access for the migrated item.

NOTE: If no permission mapping is specified then no specific permissions are applied to migrated items, and so they would inherit their access permissions from the target.

For each of the entities, there is the option to download the relevant data to .CSV file were it should then be modified (preferably in Excel) to add mapping information and then the updated file can be imported back into CPS.

NOTE: Permission mapping capabilities are limited when migrating to M-Files or File System. Contact technical support for more information.

Configuring User Security Mapping

Example: Download a User template, modify the CSV file to add user mappings and import the modified file.

11.2.

Select **Migration Preparation Phase> Security Information Mapping** and then on the **User** card, select the Download Template icon**.** The .CSV file will be populated with user information that may need to be mapped to a suitable target.

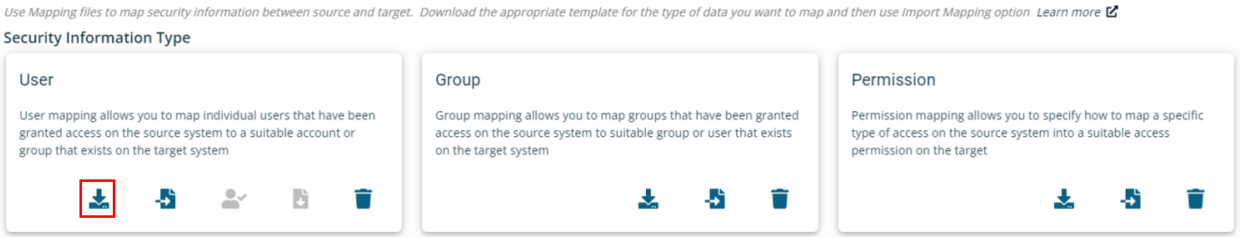


FIGURE 103 - DOWNLOADING USER SECURITY INFORMATION

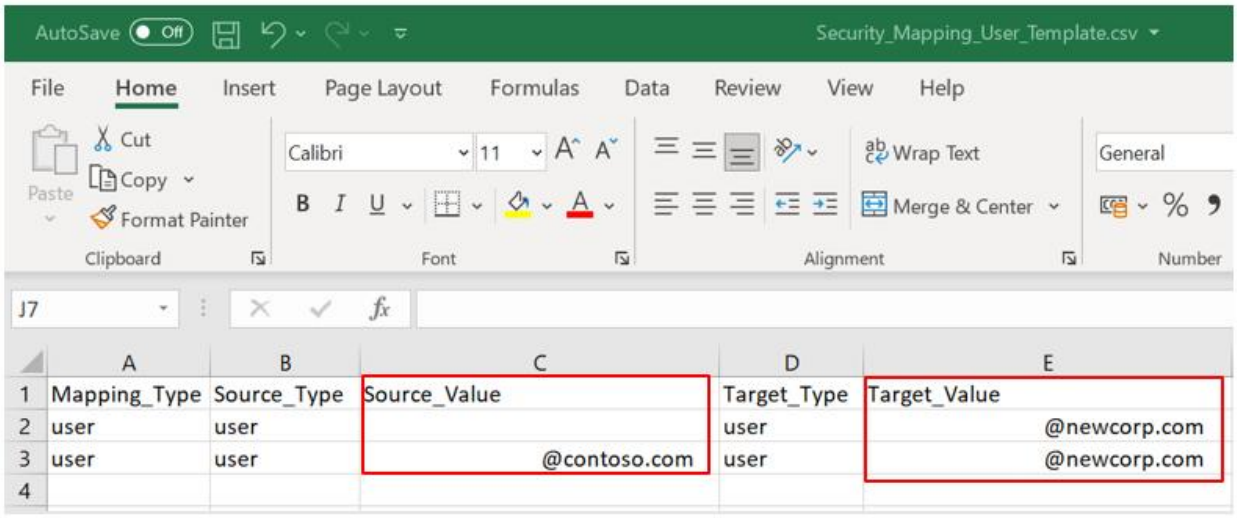


FIGURE 104 - COMPLETED USER MAPPING EXAMPLE

Edit the .CSV file to add a Target type and Target value that exists in the Target. Only the Target_Type and Target_Value columns need to be edited.

Target_Type - Defines whether the entry should be mapped to a **User** or **Group** in the target. If a user UPN is external then external sharing needs to be enabled on the tenancy.

NOTE: Target_Type should only be Group if using SharePoint Groups.

Target_Value - UPN for the respective user or group.

NOTE: A user may appear in two rows, as both a “Display Name” and an email address as shown in the example above and both of these need to be assigned target values.

NOTE: The .CSV file is populated using values from the CreatedBy and ModifiedBy fields in the source. So additional entries may be required.

Once completed, the modified file should be imported by selecting Import icon on the User card and select the updated security mapping file.

The process for modifying Groups and Permissions is the same as for Users.

Target Content Types may contain User or Group type fields which may need to be populated during migration. The information in these mapping tables will be used to map source values to new target values for these fields using the GetMappedUser function.

Note: If an entry has no target information or the target value specified does not existing then the relevant items will fail migration. Security Mapping 'Catch All'

After import, use the Validate User details option to check the target details are correct and review them using the Download Verified Security Mapping option.

In some scenarios, e.g. if a user has left the company, it makes no sense to map security entries for these users and often these users are not in the target system at all. However, the target platform may still require some user information to allow the items to be created and thus migrated.

Within the security mapping file a 'Default' value can be specified which is used for any user that does not appear at all in the User Mapping table.

Example *Default User* Mappings Entry

Mapping_Type	Source_Type	Source_Value	Target_Type	Target_Value
User	User	Default	User	orphanusers@acme.com

In the above case, any users with no entries in the mapping file with be assigned the specified target value. If a user entry is added but target information is blank or does not exist, the items will fail migration. So, it's important users without a value target entry are removed from the mapping table.

If no default entry is specified, then any items without an unmapped user or group will fail with an error similar to: -

Error: Failed: Get Mapped User: Did not find Mapping for User : [acme\shelleyp]

If a user or group is mapped to a target that does not exist the migration will fail with an error similar to: -

Error: SharePoint Error: The specified user percey.shelley@acme.com could not be found

Default entries can be applied both to user and to groups.

11.2.1. Processing empty source user fields

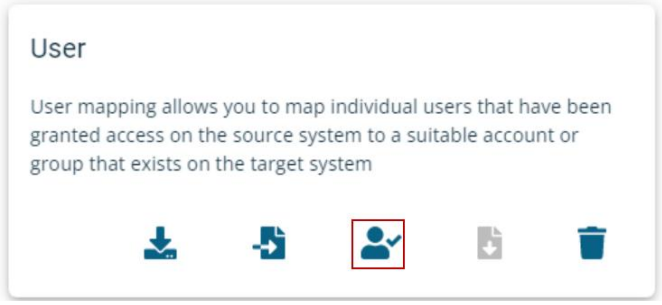
In some cases, the User field in the source system may be blank, in that case a User entry for PVQ_BLANK will be listed in the downloaded user mapping file. The target value for this should be the UPN for an account to be used when the source user field is empty.

If an entry with a target value of PVQ_SKIP is created then the value specified in the **Default User** field on the relevant source connection will be used for empty User fields.

Mapping_Type	Source_Type	Source_Value	Target_Type	Target_Value
User	User	Use Default User Value	User	PVQ_SKIP

11.2.2. Verifying User Mappings

The target address specified for users in security can be validated before actual migration using the Verify option.



Selecting this option with check each target address to ensure it is valid and report any errors.

After the check is complete, select the “Download verified security mappings” option. This will generates a report which is in the same format as the user mapping report that was imported but includes a new ‘VerificationResult’ column with the status of the verification of the target, (either “User found”, “User not found” or “Target not specified”) . Review the report and if any entries have the “User not found” or “Target not specified” status then correct these in the mapping file or target system as appropriate, delete the existing mapping and import and verify again.

11.2.3. Sample Security Mappings

Example User Mappings

Mapping_Type	Source_Type	Source_Value	Target_Type	Target_Value
User	User	EVAdmin	User	evadmin@acme.com
User	User	Percy Shelley	User	percy.shelley@acme.com
User	User	Nevil Hanson	User	n.hanson@acme.com

FIGURE 105 - EXAMPLE OF USER SECURITY MAPPING

Example Group Mappings

Mapping_Type	Source_Type	Source_Value	Target_Type	Target_Value
Group	Group	Senior Leadership	Group	slt@acme.com
Group	Group	Security Team	Group	Security@acme.com
Group	Group	Data Protection Team	Group	dp@acme.com

FIGURE 106 - EXAMPLE OF GROUP SECURITY MAPPING

Permission Mapping

NOTE: If there is no requirement to migrate container or item level security but to migrate only CreatedBy and ModifiedBy fields, then permission mapping is **not** required.

Example Permission Mappings

Mapping_Type	Source_Type	Source_Value	Target_Type	Target_Value
Permission	AccessRights	FullControl:AccessPermit	PermissionLevel	Full Control
Permission	AccessRights	Delete:AccessPermit	PermissionLevel	View Only
Permission	AccessRights	Read:AccessPermit	PermissionLevel	Read
Permission	AccessRights	Write:AccessPermit	PermissionLevel	Edit

For Mapping type permission, Mapping_Type, Source_Type and Target_Type are always the same as above, it's only the mappings in Source_Value and Target_Value that will need to be changed.

The target values should contain the permission role. E.g. Contribute, Full Control or whatever additional roles the site administrator may have created.

Samples files can be found installation folder which is by default, C:\Program Files\Proventeq\Proventeq Content Suite\SampleFiles.

12. PHASE 2 – GENERAL SETTINGS

General Settings contains settings controlling behaviour of migration jobs. Such as how to manage Embedded Links, handline characters that are valid in the Source but invalid in the Target etc.

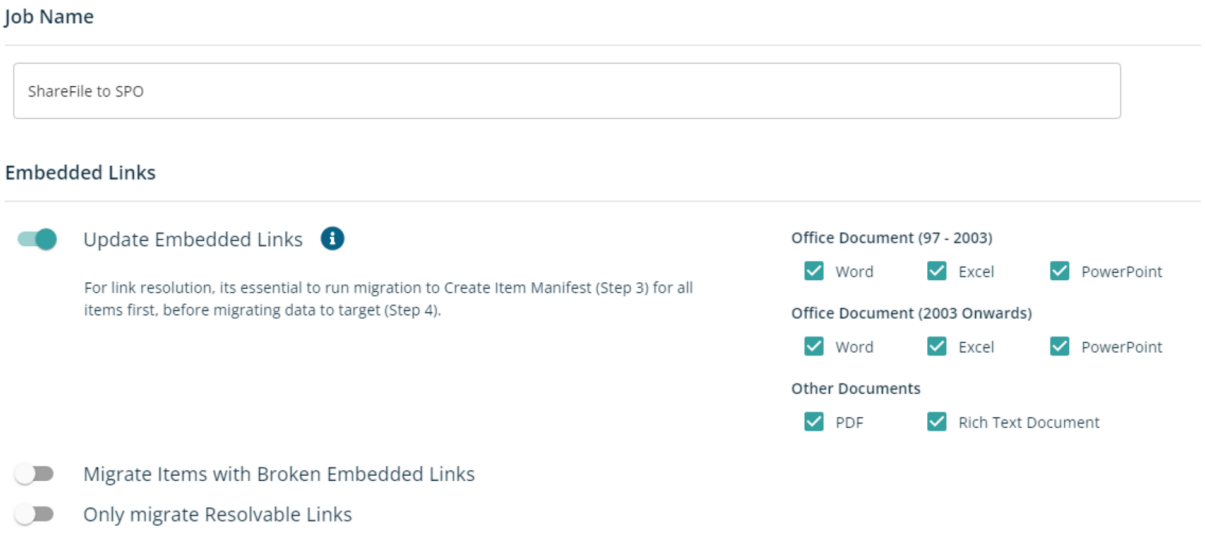


FIGURE 107 - GENERAL SETTINGS

NOTE: Above is a subset of all settings. Prior to the migration processing you should review all of the settings available to ensure they are correct.

Job Name: This was set during the initial job creation and can be changed here.

Update Embedded Links can be used to re-establish hyperlinks in well-known Source file contents with the new location of linked files in Target container while performing the migration. This feature is not mandatory for migrating content from Source to Target.

NOTE: Enabling this feature will make the migration significantly slower and requires Microsoft Office to be installed on the CPS server.

NOTE: When using this feature, it's essential to take a two pass approach to the migration. When in the migration phase, first execute **up to Step 3-Transform ONLY.** This will generate the necessary information about the location items will be migrated to. After this has been executed for **all items**, then Step 4 – Load can be executed and in doing so, this feature will replace the links within document types specified above with the new target information generated during step 3. The reason why Step 3 must be performed is to ensure that the target URL for all embedded items is known, before any migration which may refer to them takes place

The document types supported are:-

File Type	Extensions
Word	• .doc .docx
PowerPoint	• .ppt .pptx
Excel	• .xls .xlsx .xlsm .xlsb
Other	.pdf .rtf

NOTE: Embedded links in encrypted/password protected items cannot be processed

Migrate items with Broken Embedded Links

A document may contain a link to another item where the target URL is not/cannot be determined. For example, referring to another document that has been deleted in the source or potentially not captured during the Discovery process. Such links are termed Broken Links. This setting controls whether items should be migrated even if it refers to an item where the Embedded link will no longer function in the target. The default behavior is that such items would not be migrated.

Invalid Characters

Replace Invalid Characters: Targets often have different character limitations than the source so it may be necessary to replace such characters during migration to ensure the item is successfully migrated. The Pre-Check feature will identify items containing invalid characters.

The default list of invalid characters is suitable for migrations to SharePoint Online.

FIGURE 108 - CONFIGURING INVALID CHARACTER REPLACEMENT

Replace with allows you to specify the character to replace any of the invalid characters specified. You can specify no replacement character, if required in which case the character(s) are removed.

Example An invalid filename like **Help ? information.txt** becomes **Help _ information.txt**

File Handling

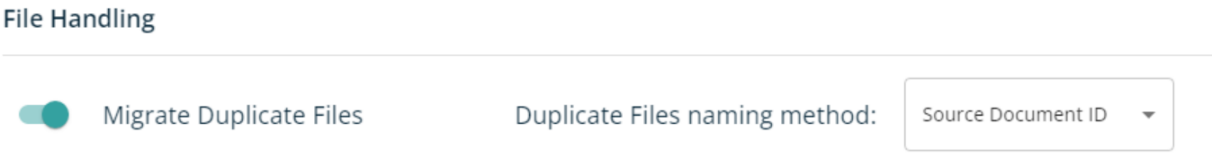


FIGURE 109 - CONFIGURING FILE HANDLING SETTINGS

Migrate Duplicate Files: When selected(default) then if an item with the same exists in the migration job, then during migration, the name will be changed to make it unique. If unselected, any duplicate items will not be migrated.

The Duplicate Files naming method allows you to control what naming scheme should be used. There are two options: -

Source Document ID

The Source Document ID will be appended to the item name as a suffix. E.g. Original item called ItemWithSameFileName.pdf would be named ItemWithSameFileName-090d08d78001f1b7.pdf

Incremental number

A unique document name will be created by appending an auto incrementing number as a file name suffix. If unselected, any duplicate items will not be migrated.

For example, 3 files with same name as Proventeq User Guide.pdf migrated to the same folder will be migrated with names of

- Proventeq User Guide.pdf
- Proventeq User Guide (1).pdf
- Proventeq User Guide (2).pdf

External Source for Insights Information

By enabling external insight Job (please refer section **14. Insights**), we will be able to refer an existing Insight Job, where link analysis is already performed. This will ensure that during migration process only those files that has embedded links will be picked for the link resolution, which will improve the migration performance significantly.

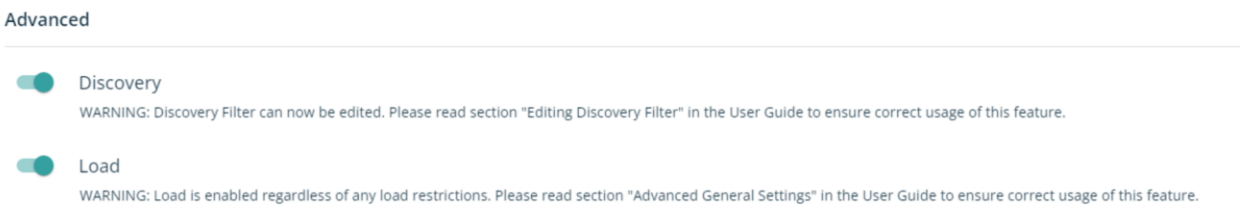
Pre-Requisite:

Ensure that the HasLinks and HasResolvableLinks column of the SourceItemInsights table on the Insights task is set to 1 for the items that should be link resolved

App.config setting:

Set InternalFileLinkResolutionVersionConstraint to “Any” (to resolve links in all versions) or “Latest” (to resolve links in latest version only)

Advanced General Settings



Discovery

By default, after a discovery has begun, it’s not possible to edit the discovery filter. This is because doing so could create inconsistent results. For example, having no filter and then adding a filter after initial discovery does not remove existing discovered items. In some scenarios it can be useful to edit the filter, for example to correct a wrong filter value. By selecting this option, it allows the Discovery filter to be edited even after discovery has started.

Load

By default, the processing stage ‘Step 4 – Migrate to Target’ cannot be selected in the Processing Screen if the “Update Embedded Links” option has been selected and not ALL items have been first processed up to the ‘Step 3 – Transform – Create Item Manifest’ step. Selecting this option allows the ‘Step 4’ option to become available even if all items have not been transformed. Note the consequences of selecting this option means that items will be migrated but may contain embedded links which have not been updated (because transform has not occurred for all items) and so will not point to the new location of migrated items.

13. PHASE 3 – MIGRATION PHASE

The migration phase deals with the final preparation and migration of items into the target. It consists of a set of 4 steps that are executed for each item, in doing so evaluating the settings specified in phase 2 i.e. structure mapping, content type and security information.

Each step will output data and during the pilot stage it's recommended to run each step separately for a subset of data to review the output in the Items List before then executing the next step for the items.

Executing a step will also execute any previous steps (if it hasn't previously been executed), so executing step 3 for example also executes step 1 and step 2 first.

Migration Steps

13.1.

Select the **Migration** option on the **Migration Dashboard** to display the migration steps.

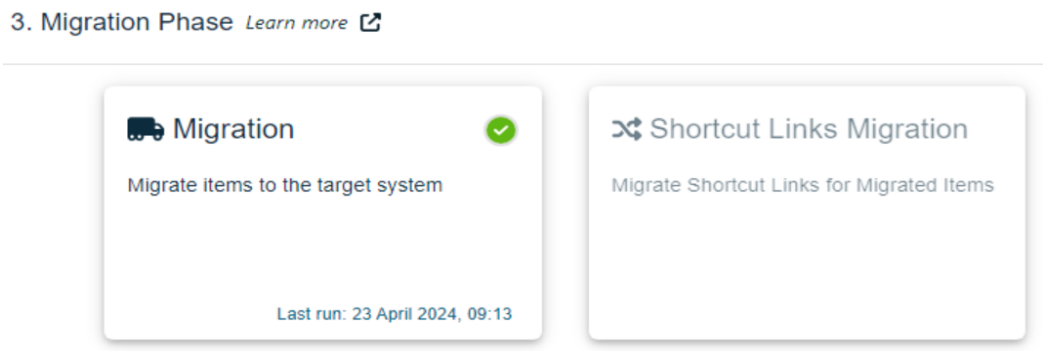


FIGURE 110 - STARTING DATA MIGRATION

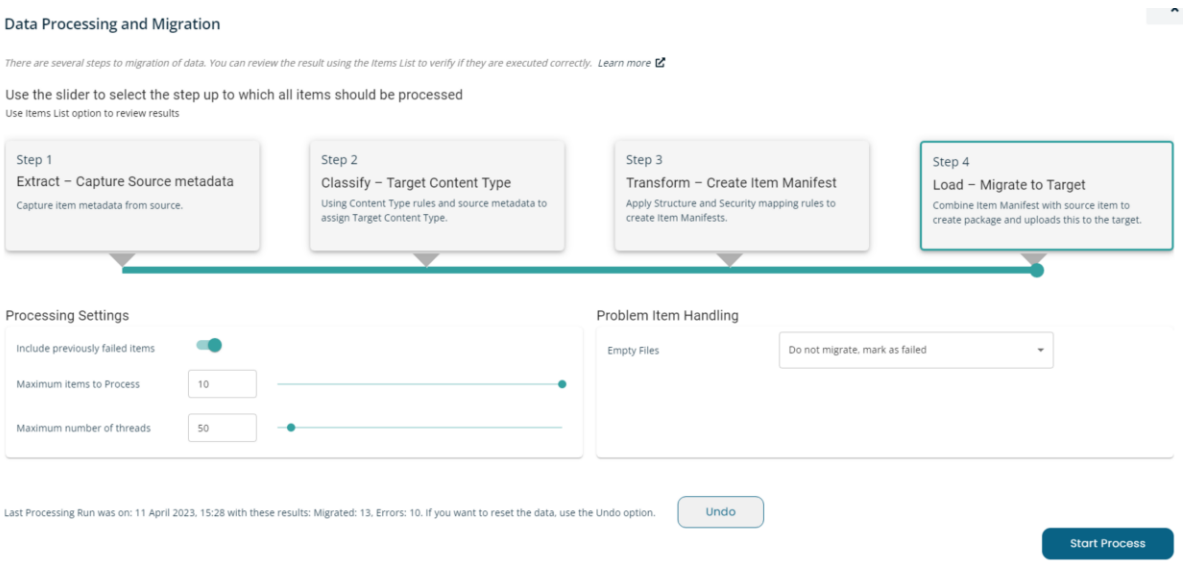


FIGURE 111 - LIST OF STEPS IN DATA MIGRATION

Migration is actually a set of 4 sequential steps that need to be executed for each item. Using the slider control you can select how many steps to execute. This allows you to review the output of that step for the processed items, before continuing.

The output generated from a step remains with that item and is not re-generated even if that step is executed again. This means if step 2 (Classify) has assigned content type or step 3(Transform) has built the Item Manifest they will remain with that item, even if subsequent changes elsewhere have been made to migration settings (e.g. Structure mapping, content and security mappings). It also means metadata captured in Step 1 is unchanged, even if such data is altered in the source system.

If steps 1 to 3 are run and the resulting output is not correct, then use 'Undo' option to delete this data. The 'Undo' option is only available once any step has been executed.

Use the **Item Details** screen to display the details of items in the **Items List** to view the output from these steps.

Step	Verification
1	Items list shows source metadata content. Ensure this is complete and correct.
2	On the Item details review the Source Content type
3	On item details, review target metadata properties. Target URL etc.
4	Review item in target location, content, metadata properties, access permissions if relevant.

To start processing, select the step by using the slider or clicking on the card and then select the **Start Process** option.

Migration Run Settings

- **Include previously failed Items Option**
  If selected, items that previously failed will be resubmitted for processing along with any items still awaiting migration.
- **Maximum number of items**
  Select number of items to be processed.
- **Maximum number of threads**.
  Defines how many items are processed concurrently. The optimal value for this setting depends on various factors including the number of documents going into each site collection. A higher number of threads gives better performance when you have fewer site collections with lots of documents in folders.

NOTE Processing can take several hours or more, depending on the amount of data and selected processing step.

Problem Item Processing

13.2. Problem Item Processing

13.2.1. Empty Files

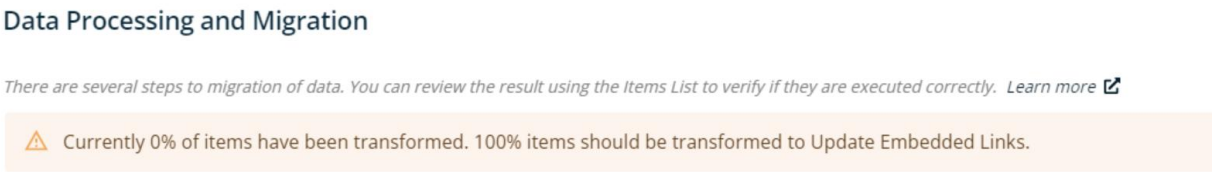
This allows control over the migration processing of source items that are empty (0 byte). There are three options: -

- **Do not migrate, mark as failed** – This will not attempt to migrate the item, instead immediately marking it as 'failed'.
- **Migrate item as is** – (default) This will attempt to migrate the item. Depending upon whether the item is a version of an item, the target system may allow the item to be migrated successfully.
- **Replace empty content with placeholder text** – This option allows the content of these 0 byte items to be replaced with placeholder content based upon its type so Word, Excel, PPT and text. You'll find existing template placeholder files in C:\Proventeq\Placeholders\default. When a 0 byte item needs to be migrated its content will be replaced with the placeholder content based upon the relevant file type so an empty Word document is replaced by a Word document containing placeholder content and so on. Using this feature allows all source to be migrated regardless of content size.

NOTE: These placeholders files can be edited if required. We recommend that if edits are made to these placeholder files that they are saved into C:\Proventeq\Placeholders\override and CPS will process these in preference to the default location. Otherwise, any changes to placeholders in the default location may be lost if the CPS software is upgraded as the installation would overwrite these existing files.

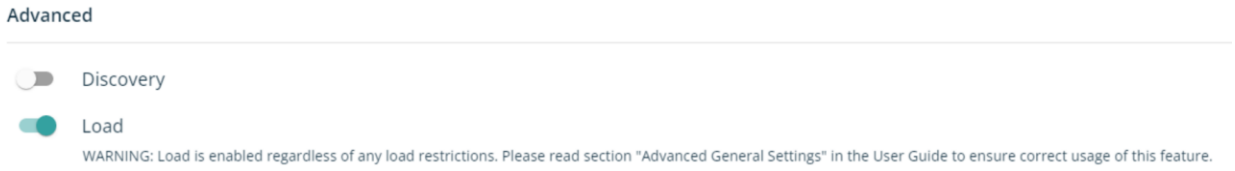
13.3. Transform Warning

If the option 'Update Embedded Links' is selected in General Settings then this warning message below may be displayed at the top of the processing screen and the option "Step 4 – Migrate to Target" cannot be selected. This is because for embedded link updates to be successful, the target URL for ALL items must first be determined by running up to Step 3- Transform for all items first, so that then as items are migrated these embedded links in the content can be updated to contain the new URL for these linked to items (even if at this stage these links refer to items that may not have been migrated yet).



Running Step 3 – Transform may not be successful for all items. If after investigation, it's confirmed that it's not possible for all items to be successfully Transformed then select the

'Load' option in the Advanced section within General Settings as this will allow the 'Step 4 – Migrate to Target' option to be selectable. Using this Load option obviously means that any embedded links to these items that have failed to be transformed will not be updated and therefore the documents will keep their existing legacy links which will not function once source system is unavailable.



Step 1 Extract – Extract metadata

13.4 The Discovery process will extract a standard set of properties from items in the source and record this information in the CPS SQL database which can then be viewed on the Item details screen accessible from the Items List

This step will retrieve further information, often specific to the source platform and once this data is extracted it can be used by subsequent processing steps. It's important to review these properties on the Item Details screen to ensure they are correct and complete.

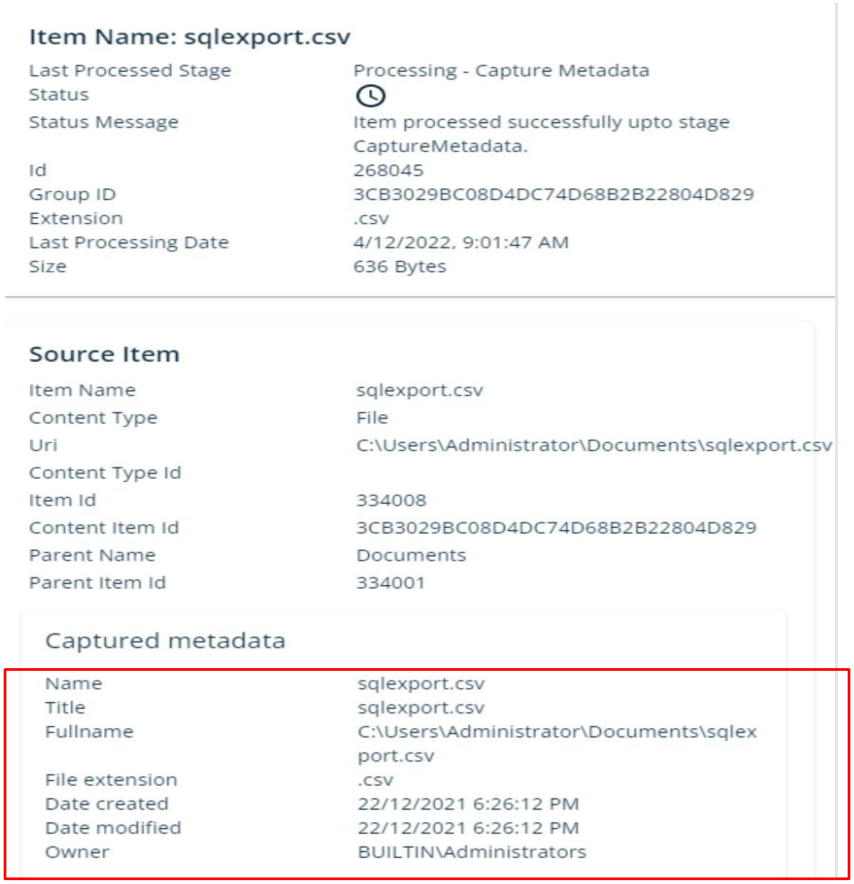


FIGURE 112 - VIEWING SOURCE METADATA RETRIEVED BY MIGRATION STEP 1

In Items List the status will be ‘Capture Metadata’

Step 2 Classify - Assign Target Content Type

The Content Type can be viewed in the Content Type field in the Source Item section of item details. This step will update the source content type of the item by execute any custom classification scripts, so running a migration to this step is unnecessary unless that feature is used. In Items List the status will be ‘Classify’.

13.5.

See Custom Classificatio for further details.

Based upon knowing the source content type you can determine what the Target Content Type will be by viewing content type mapping.

As a reminder, content types define how properties are moved between the source and target items so it’s critical that the correct Target Content Type is assigned based upon business requirements.

Step 3 Transform - Create Item Manifest

13.6. The Item manifest is a package of data that contains all necessary metadata associated with the item to be migrated. The manifest content is migrated along with the item content in Step 4.

This step populates the Item Manifest for each item by:

- Creating target metadata by using the assigned Target Content Type to apply appropriate metadata processing to the source fields.
- Creates structure mapping metadata by evaluating structure mapping to determine target location to migrate the item to.
- Creates security information metadata by evaluating security mapping.

In Items List the status will be ‘Transform’.

13.7. NOTE: If the option to ‘Update Embedded Links’ is enabled in General Settings, it is essential that **ALL ITEMS** are **first processed ONLY to Step 3-Transform**. This is because this step determines the target locations of items which is then used in step 4 (Migrate to target) to replace existing embedded links with links to the new location items will be in once actually migrated. See 13.3. Transform Warning

Step 4 Load – Migrate to Target

This is the default step and will execute the following: -

- Retrieve item contents from the source and combine that with Item Manifest for that item.
- Based upon package size settings, a set of these items will be placed into a package and a request sent to upload then to Azure Blob Storage.

- Send notification to SharePoint to request that package is ingested.
- SharePoint will process notification and ingest item packages into target.
- Client will display progress,

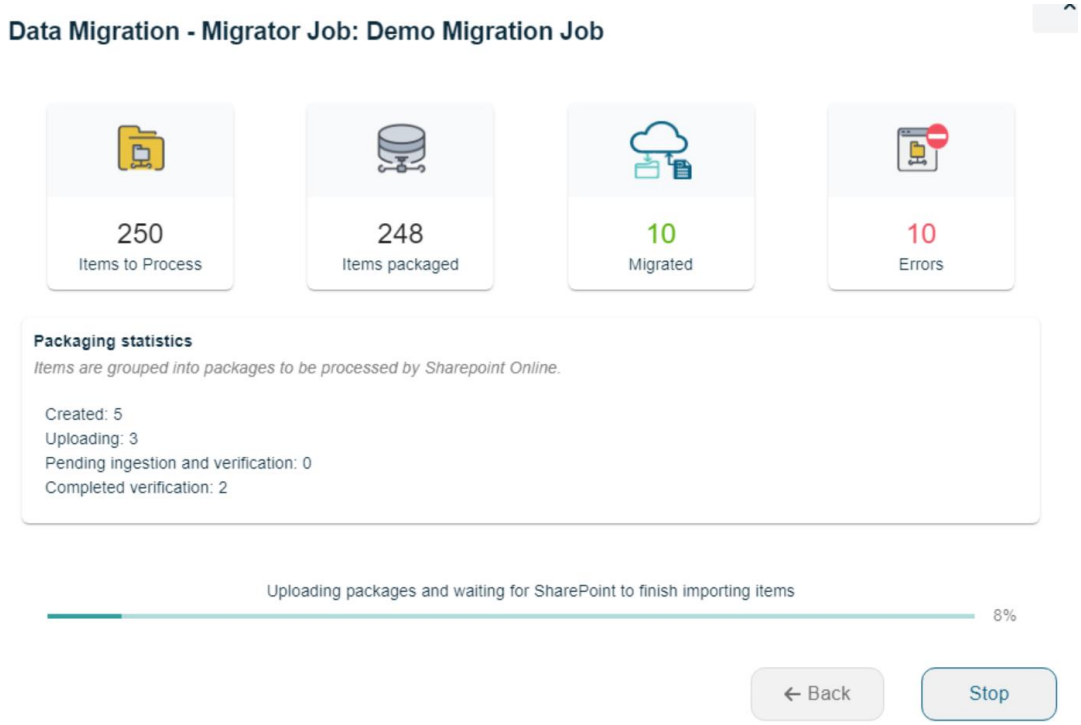


FIGURE 113 - MIGRATION PROGRESS SCREEN

Packaging is the terms used for a technique used by CPS to improve migration performance into SharePoint. During migration, items are grouped together into packages (**Created**) and these packages are then uploaded into Azure (**Uploading**). SharePoint will then process these packages asynchronous (**Pending ingestion and verification**). Package verification is a process that CPS executes from time to time during migration to check the current status of any uploaded packages till the processing of that package is verified (**Completed verification**).

13.8 Migration Considerations

Loading data into the target is often the area where problems occur because it's ultimately a verification that the previous steps were correct. In some cases, failure may mean that an error occurred when attempting to migrate the item, in other situations an item may migrate to the target successfully but there are issues with the item.

'Failure examples'

- Item is migrated to the wrong location due to Structure mapping
- Item may incomplete or incorrect in the target due to content type/metadata mapping issues
- Incorrect target users or groups specified in security mappings

Since Load is the stage where content is retrieved and copied into the target, it's obviously and area where performance issues can arise.

For information on SharePoint best practices see https://docs.microsoft.com/en-us/sharepointmigration/sharepoint-online-and-onedrive-migration-speed

13.9. Undo Migration

The Undo option will reset the system back to the state it was after discovery by deleting any of the data generated from migration steps 1 – 4, including deletion of the items from the target. This means all the records captured by discovery are still in the CPS database and a re-run of Discovery is not required.

> Note: Undo will 'reset' the license consumption for items undo is for steps 1-3 ONLY. Undo of migrated items does NOT reset the license. That means if an item is migrated to the target, then deleted and then re-migrated that would consume the license twice.

Undo can be used when a problem has occurred and so some of the steps that have already been executed need to be corrected and run again.

The undo option is displayed on the Data Processing and Migration screen once any of the steps in phase 3 (Extract, Classify, Transform or Load) are executed. If data has been migrated to the target undo option will delete them irrespective of whether they have been subsequently modified. The deletion may fail if retention plans do not allow items to be deleted.

Undo Scenarios

- Items are migrated to the incorrect location, due to structure mapping issues
- Migration items have incorrect metadata properties or content types
- Migrated items have incorrect security information
- Step 2 (Classify) or Step 3 (Transform) have been run but the results as shown in the Items List suggest the data is incorrect.
- Source item metadata has changed

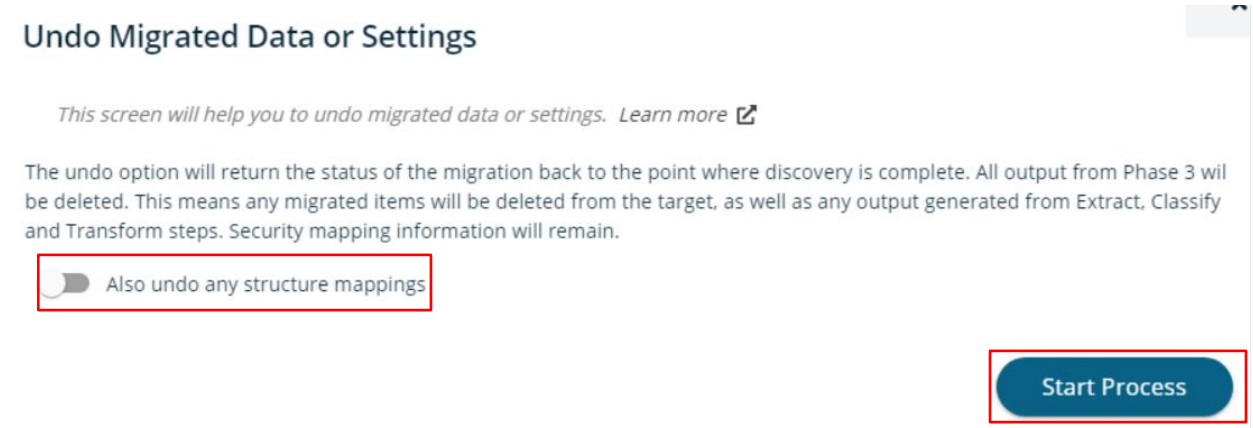


FIGURE 114 - UNDOING A MIGRATION

If the "**undo any structure mappings"** option is selected, Undo will delete any structure mapping information. Content Type settings Security mappings are not removed by Undo.

13.10. Resolving migration errors

Errors are mostly like to occur when executing the Transform or Load stages.

To view failed items, use **Reports->Failed Items Report**

The **Status Message** column will may contain a numeric error code to identify the type of issue.

e.g.

046 Transform Failure - Error: Failed: Get Mapped User: Did not find Mapping for User: [Domain\Username]

Clicking on the error will access the online knowledge base and display a relevant article to help you understand and resolve the error.

13.11. Shortcut Links Migration

Shortcuts are standalone items in the Source which when opened will open the item they refer to. Using this option will migrate shortcuts to the target but will ensure that the shortcut is updated to refer to the new target location. The maximum processing speed of Shortcut Link migration is around 2000-3000 per hour so this should be factored into any migration planning.

The Discovery Statistics panel on the Migration Dashboard displays the number of shortcuts found in the source. To migrate shortcuts the 'Links' content type must be assigned to a target content type.

Note: Microsoft do not recommend the use of shortcuts in SharePoint.

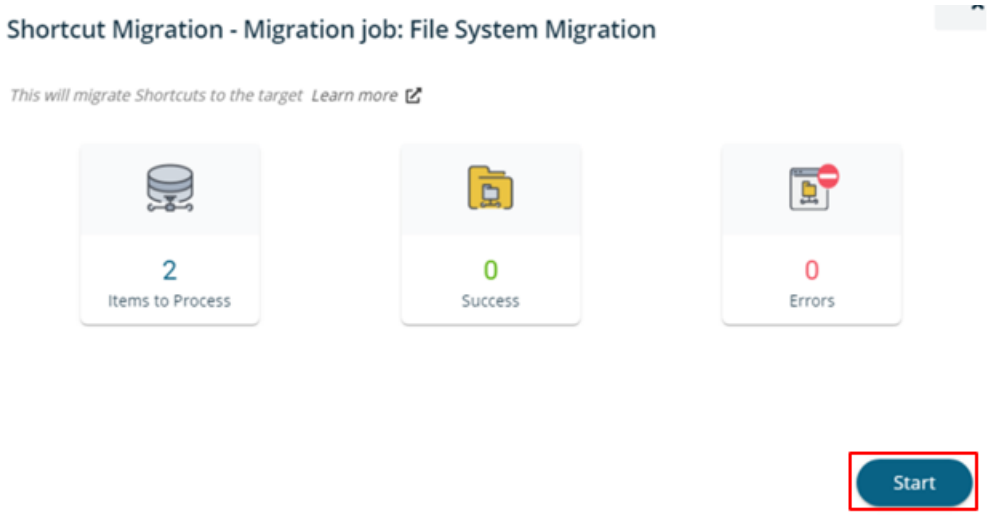


FIGURE 115 - STARTING SHORTCUT LINKS MIGRATION

Verify and communicate with end users

Ensure that the data has been migrated as expected using the mapping and security settings that have been configured. This is best done by logging into the Source and Target and ensuring the Folders and Files have been migrated and can be accessed.

13.12 Review the Source Data and the Target Data

- Using the structure mapping to confirm data has been moved to the intended location
- For each content type, review item and meta data is correct and that the items operate as expected in the target.
- Where different permissions mappings have been carried out, verify they achieve desired outcome

14. Power BI Reports

Power BI Reports can be used to access the information in the Proventeq Migration Accelerator database to provide intuitive and interactive drill down reports. The reports can be viewing by using the Visualize option on the migrated dashboard after Pre-Checks option has been run.

Two warnings displayed when opening Power BI : -

1. Potential security risk

Displays warning about 'Potential security risk' due to the file using multiple data source. Click **OK** to continue.

2. Native Database Query

Displays warning to approve running native query. Click **Run**

14.1 General Distribution of Content

This report shows the general distribution of content within the system. This includes content type distribution, mime type distribution, version documents and documents under folders.

FIGURE 116 - POWER BI - GENERAL CONTENT DISTRIBUTION

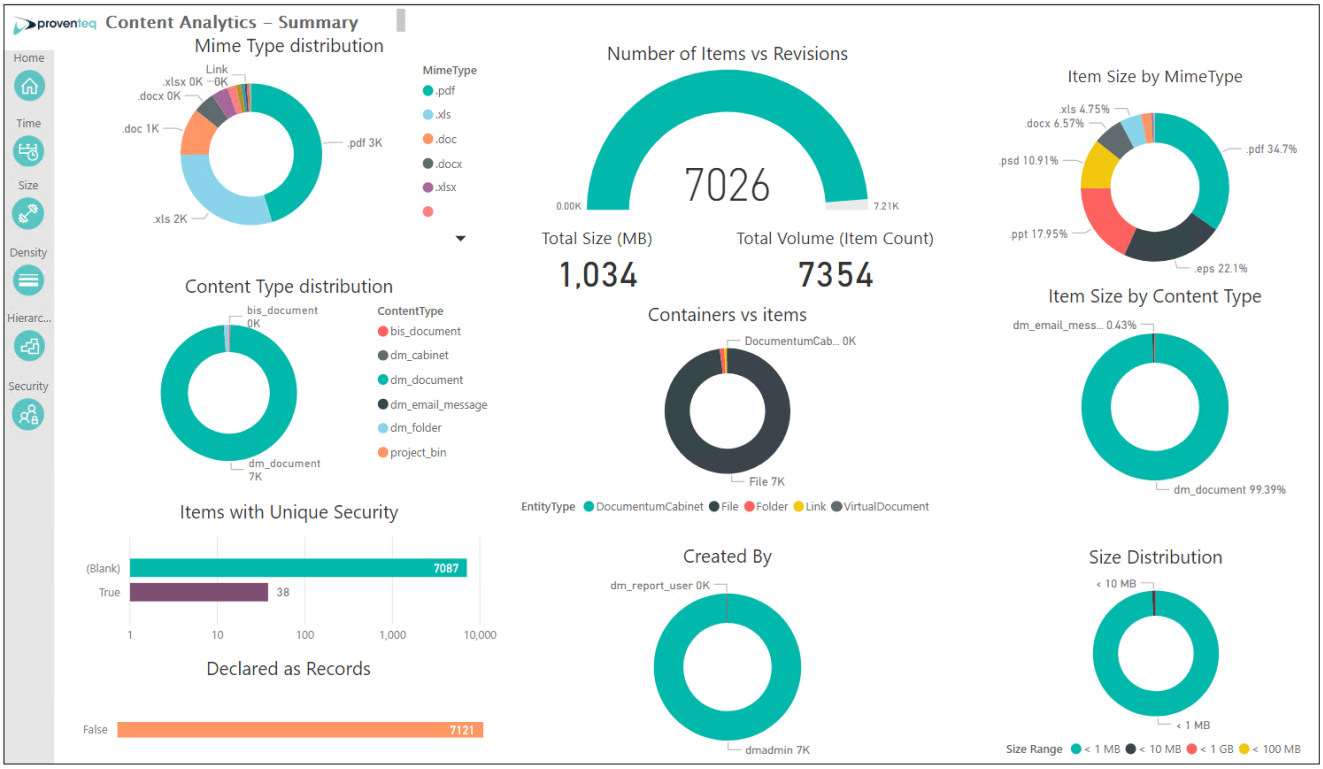


FIGURE 117 - ANALYTICS – CONTENT SUMMARY

To navigate different reports using Navigation, please use CTRL+CLICK on the menu button on the left Menu panel.

Pre-Migration Checks- Version Continuity

Power BI Version Continuity Reports shows the items which are qualified in different categories where Proventeq will apply certain rule and configuration to resolve these target constraints.

14.2.

FIGURE 118 - POWER BI – PRE-MIGRATION CHECK – VERSION CONTINUITY

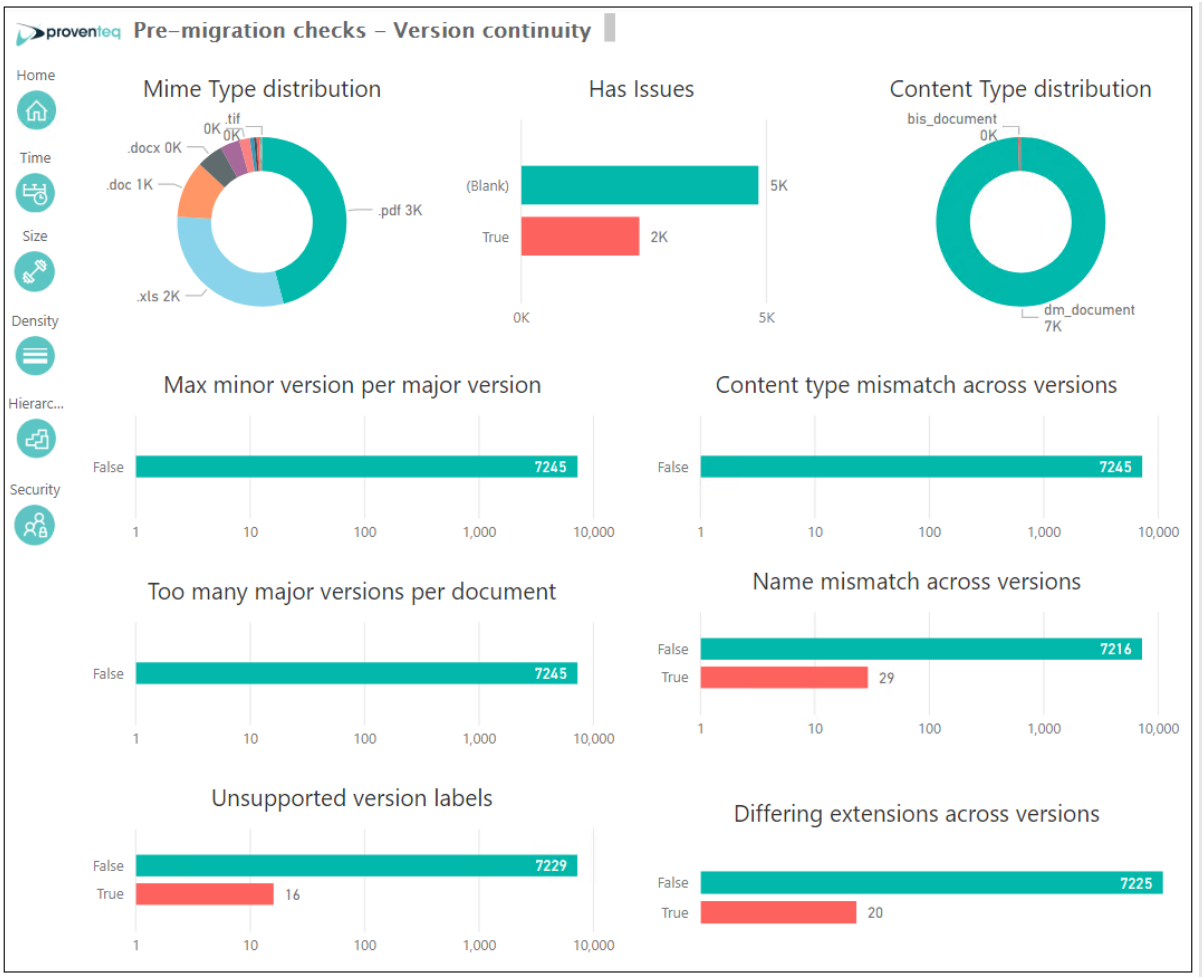


FIGURE 119 - ANALYTICS – VERSION CONTINUITY

Pre-Migration Checks- Paths and Names

Power BI Paths and Names Reports shows the items which are qualified in different categories where Proventeq will apply certain rule and configuration to resolve these target constraints.

FIGURE 120 - POWER BI – PATHS AND NAMES

14.

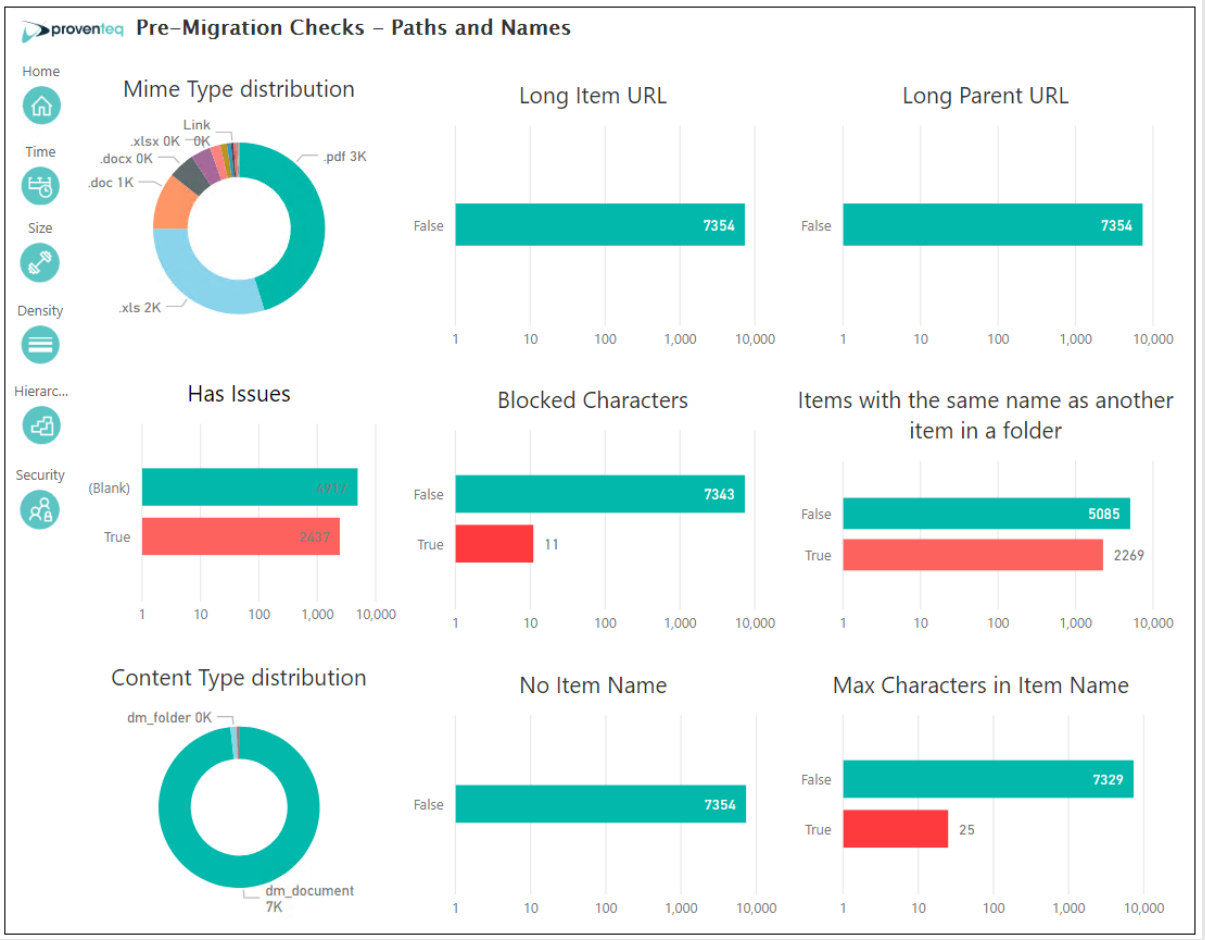


FIGURE 121 - ANALYTICS – PATHS AND NAMES

Pre-Migration Check – Folders & Files

Power BI Files Reports shows the results of several checks for folders and folders.

FIGURE 122 - POWER BI – PRE-MIGRATION CHECK – FILES

14.

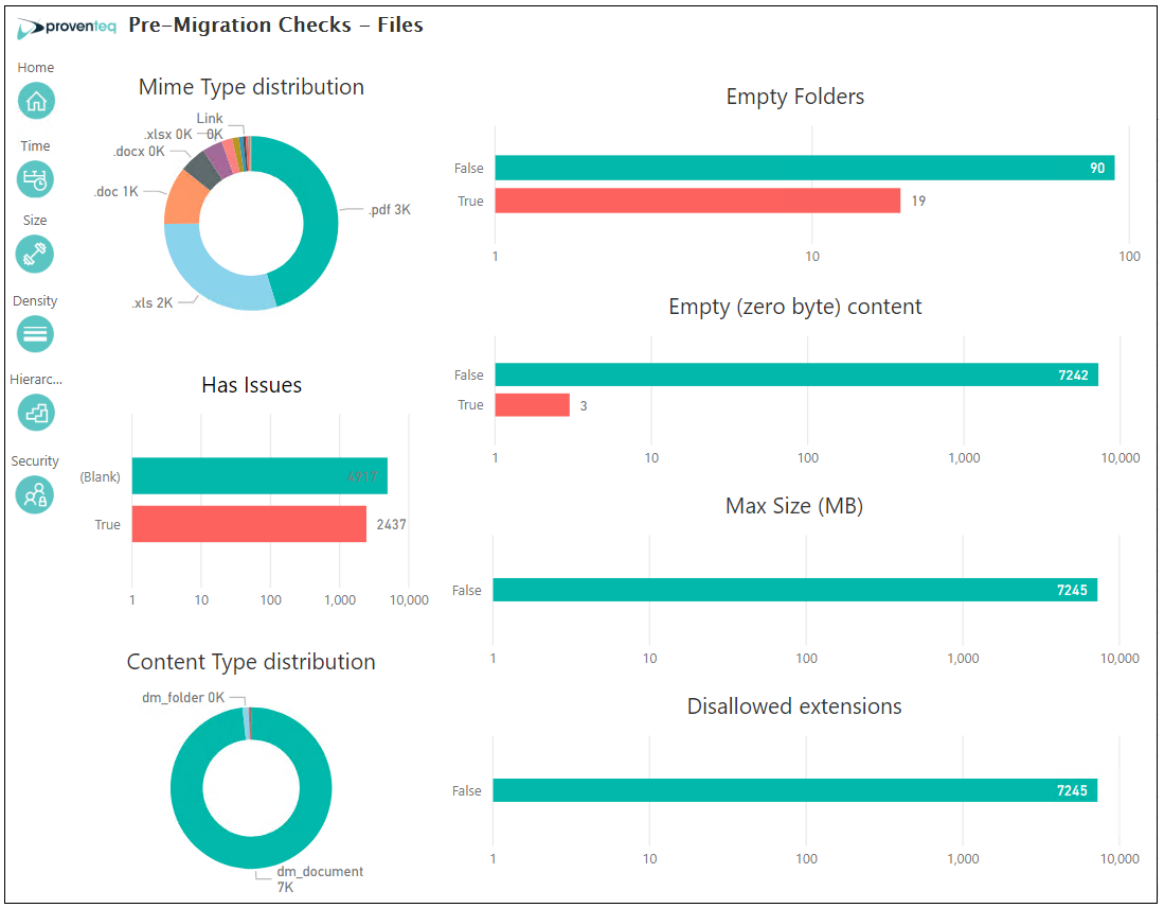


FIGURE 123 - ANALYTICS – FILES

Note: This report will highlight Max Size files, as Migration Accelerator supports migration of files up to 2GB

General Usage of Content

Power BI usage report shows the history of documents such as- when it has been created, modified, created date range, modified date range, created by, and modified by users.

14.

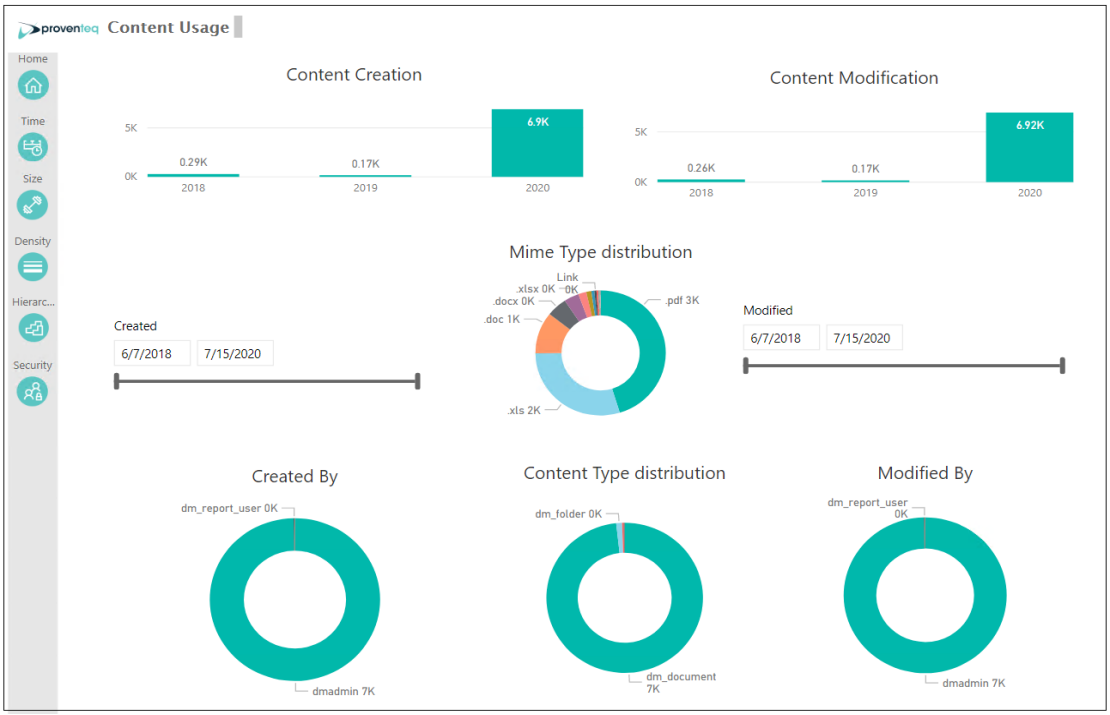


FIGURE 124 - ANALYTICS – CONTENT USAGE

14.6.

Container Analysis

Power BI Container report shows the hierarchy of the Source system.

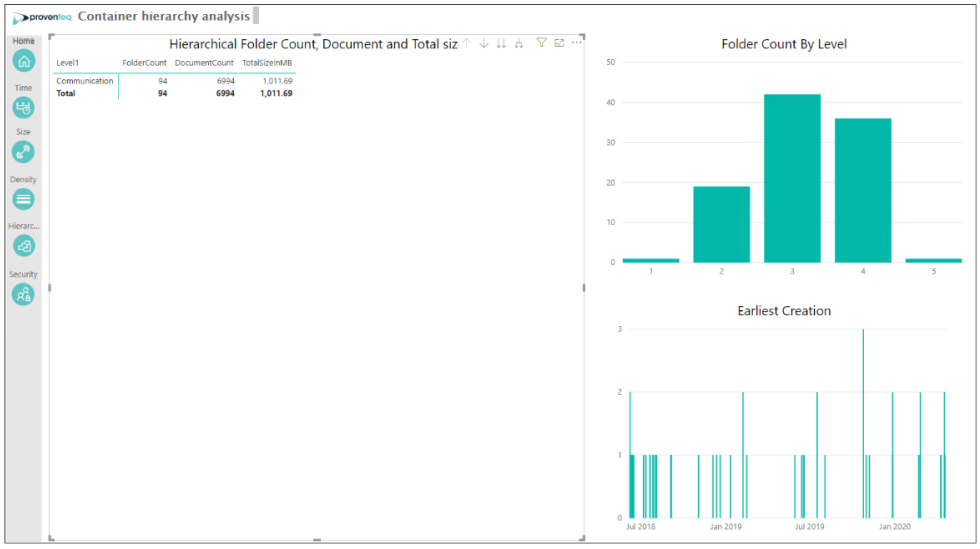


FIGURE 125 - ANALYTICS – CONTAINER ANALYSIS

As container analysis is an intensive and complex operation, it is disabled by default. To enable, change the application configuration setting below to True and rerun discovery.

<add key="PerformContainerAnalysis" value="false" />

15. INSIGHTS

Proventeq Insights is a component within the Proventeq Content Suite which provides customers with the opportunity to gain a better understanding of their ECM system prior to considering actual migration. Until Migration Accelerator, no target system needs to be configured or specified for Insights.

The analysis provides the following capabilities: -

- Migration readiness precheck – Will check data against known limitations of SharePoint as well as present information data that may be useful for customers to be aware of prior to further processing.
- ROT - Redundancy, Obsolete and Trivial (ROT) to determine data which customers may wish to delete from the source or exclude from possible migration.
- Embedded link analysis – Examine Office documents for embedded links to provide details of links within documents.

15.1. Installation and configuration

The pre-requites and installation instructions are contained within the CPS Installation Guide. Note no pre-requites related to the target need to be followed since Insights does not require any target configuration for it to run.

Configuring Insights - Creating an Insights Job

NOTE: If you have used CPS Migrator, you'll find Insights uses a very similar UI framework.

A connection to the source needs to be created which can be done separately, or during the Insights job creation process. Specify system and authentication details for the relevant source system and then further details of the source path to be processed need to be specified. See **Error! Reference source not found.**for details on adding source connections.

Once the Insights job is created the Insights dashboard is displayed

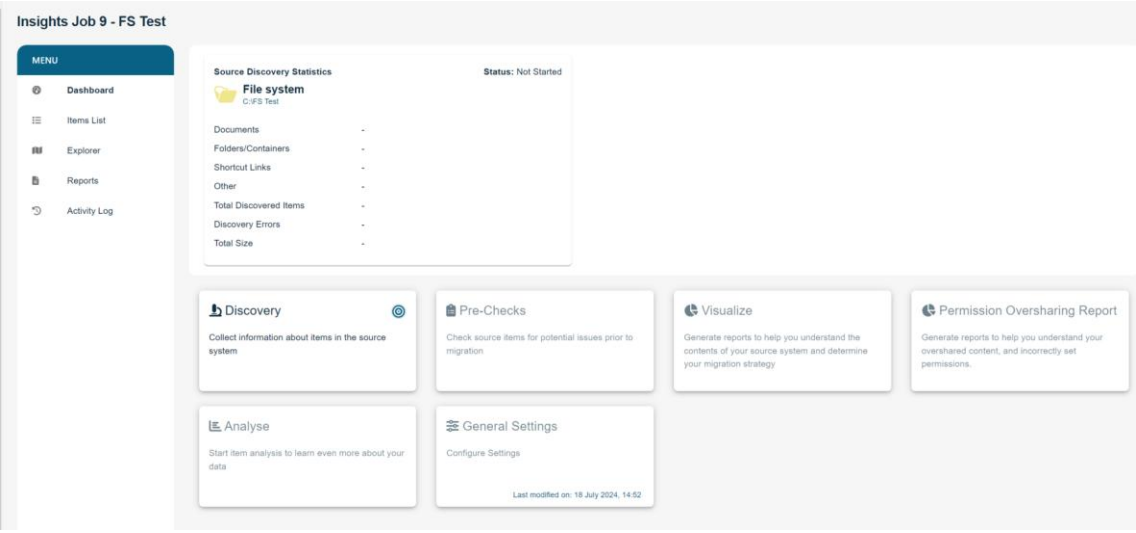


FIGURE 126 - INSIGHTS DASHBOARD

Discovery

The first step is to run Discovery which will connect to the source system and capture details of each item and record this into the CPS staging database (Note no item content is stored in the database).

Select the Discovery option and review the discovery settings. Once these are confirmed, select the **Start Discovery** option. The speed of discovery depends upon a number of factors including source system health, CPS hardware specification and number of items to be discovered.

Once discovery is complete other options within the Insights Dashboard then become available.

Delta Discovery

Often Insights is a single pass exercise of running discovery and analysis to allow customers to understand their ECM system before deciding on what strategy to take going forward. However, in some cases the source system may continue to change enough that another discovery is required.

15.2. Pre-Check

Migrations between different sources and targets often have incompatability issues due different capabilites,limitations and behaviours in the target which could impact the migration.

The Pre-check option allows a set of checks against the source data to be executed, based upon the known limits of SharePoint Each can be enabled/disabled and some have additional configuration parameters which can be customised as required.

Analyse

The Analysis option is used to carry out a detailed analysis of discovered items.

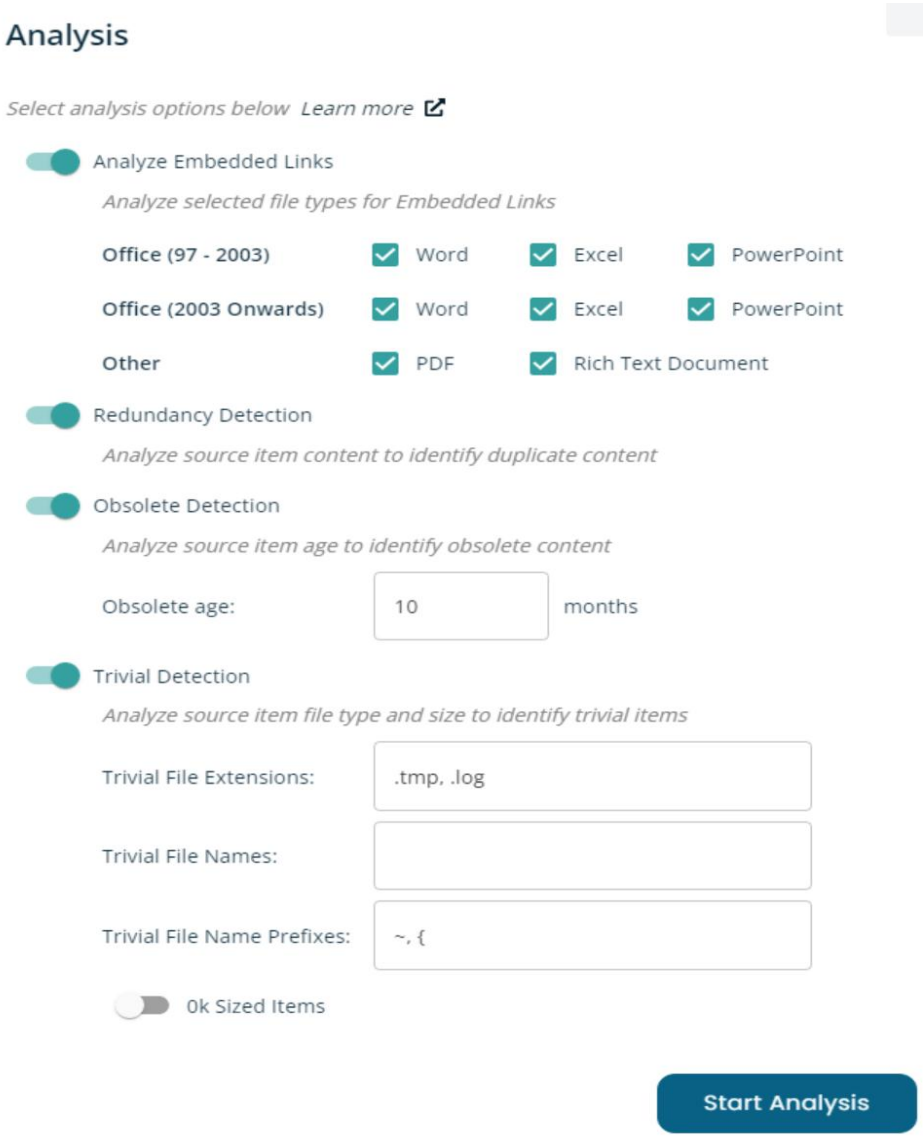
15. 

FIGURE 127 - INSIGHTS ANALYSIS OPTIONS

This table explains the scope for each of the analysis options.

Analysis Option	Folder	Item	Versions supported
Embedded Links	N/A	Y	Latest version only
Redundancy	N/A	Y	Latest version only
Obsolete	Y	Y	All versions
Trivial	N/A	Y	All versions

FIGURE 128 - INSIGHTS ANALYSIS SCOPE

So, for example, the Obsolete option will check the age of both folders and items and mark them as obsolete.

Embedded Link Analysis

For items which are of the selected file type, this option will retrieve the item for the source and examine the content for embedded links. The results of this analysis can be seen in the Embedded Links report.

Redundancy

The definition of redundancy is to find two items with the same content. This option will retrieve items from the source and generate a MD5 content hash. At the end of processing run, all hashes are checked for duplication. If more than 1 item has the same hash value then they have identical content. The item with the earliest date is considered the 'master' and any other instances of that item are considered the duplicate. The results analysis of this can be seen in the Redundancy Report.

Obsolete

Obsolete items is defined as items where the Last Modified date of the most recent version is older than the age as specified in Obsolete Age field. It is NOT the case that some older versions of an item will be identified as 'obsolete' and more recent versions are not. The results of analysis of this can be seen in the Obsolete Report.

Trivial

Trivial items are those that match any one of the criteria specified i.e. those with specified file extension or have a specific name or have a filename that begins with specified character. One or more single comma separated characters can be specified.

0k sized items

A toggle option can also be used to identify items with 0k size as Trivial. Note this information comes from the source system itself so this size information in the source system database much be trusted and corrected.

15.4. Insights Reports

Redundant

This report displays a list of items identified as Redundant based upon the criteria specified during Analysis.

Obsolete

This report displays a list of items identified as Obsolete based upon the criteria specified during Analysis.

Trivial

This report displays a list of items identified as Trivial based upon the criteria specified during Analysis.

Embedded Link

For each embedded link in documents, this report displays a detail of the source document and the link and its transformation status.

Pre-check

This report displays a list of items that fail relevant pre-check criteria after the pre-check option has been executed. A drop-down field can be used to see which particular pre-checks failed and how many are impacted.

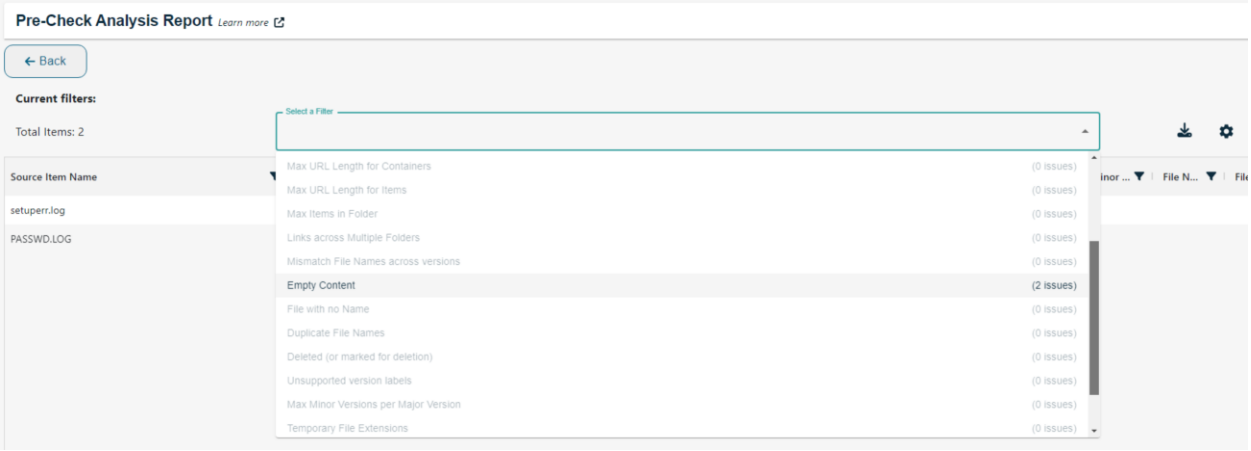


FIGURE 129 - EXAMPLE PRE-CHECK REPORT

16. ORCHESTRATOR

Proventeq Orchestrator is a component within the Proventeq Content Suite which provides customers with the opportunity to automate the migration.

Installation and configuration

The pre-requites and installation instructions are contained within the Orchestrator Installation Guide.

16.1. Setup New Orchestrator Job

To create a new orchestrator job, click on Create Automation Schedule action card. It will redirect on Setup screen of Create Automation Schedule wizard

16.2.

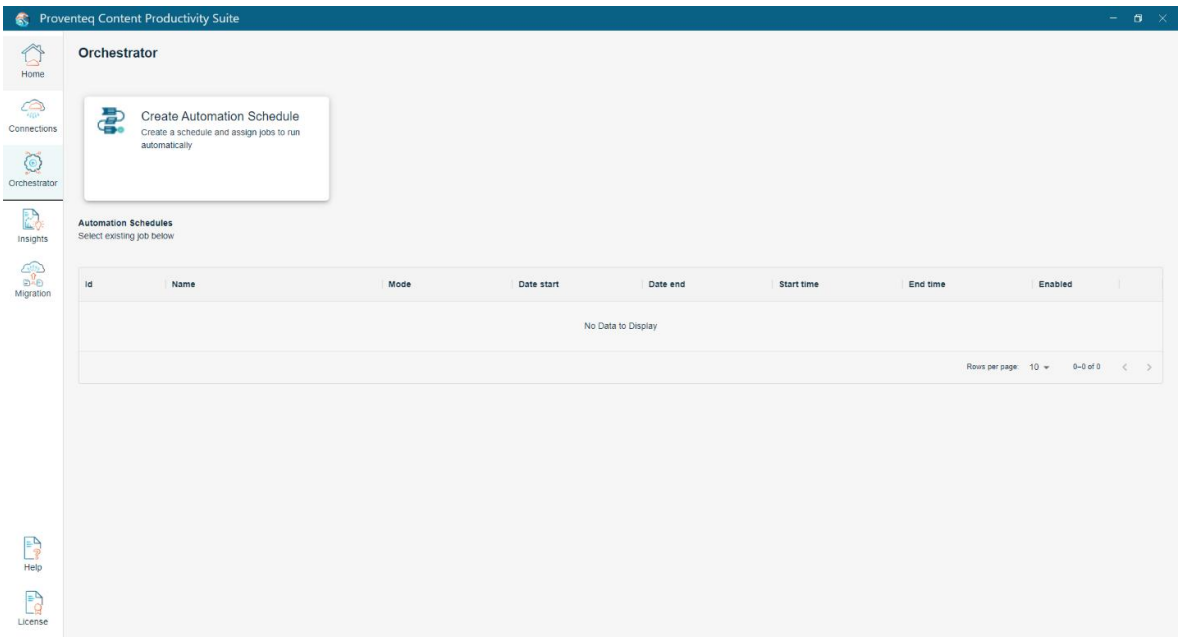


FIGURE 110 - ADD NEW ORCHESTRATOR JOB

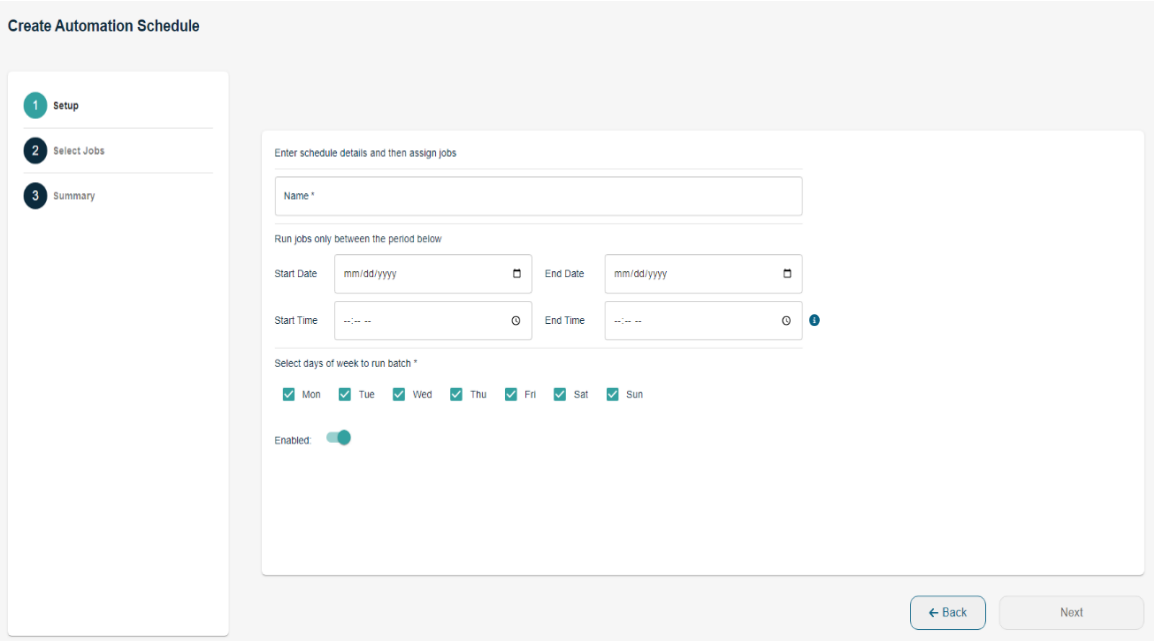


FIGURE 111 - SETUP AUTOMATION SCHEDULE

On the screen, following information required to fill up

Name: Name of orchestrator job to uniquely identify the different scheduled jobs

Start Date: This is a date field to configure start date of schedule.

End Date: This is a date field to configure end date of schedule.

Start Time: This is a time field to configure the start time of schedule. Based on selected time, schedule will be start on/after that time.

End Time: This is a time field to configure the end time of schedule. Based on selected time, schedule will be end on that time.

Select Days of Week: This is a set of check boxes to configure on which day of week schedule will run. In ideal scenario, selected day(s) should be in sync with selected date range.

Enable: This is toggle to mark created schedule to Enable for execution. If want to add a schedule but not to execute it, set this toggle OFF

> **NOTE:** Orchestrator Job will run between selected Start time and End Time each day for selected date range.
>
> **NOTE**: If schedule date range is 1st July 2024 (Monday) to 4th July 2024 (Thursday), select days of week to configure between Monday to Thursday.

After adding all required field when click on Next button, it will move on next wizard screen to select job to add into schedule

16.3. Select Job

This screen will show a list of available job(s). Jobs which are not added into another schedule will be displayed on the list. Use checkbox control display for each job to select one or multiple jobs to add into schedule. After selecting one or multiple jobs click on Next button, it will move to next wizard screen to show summary.

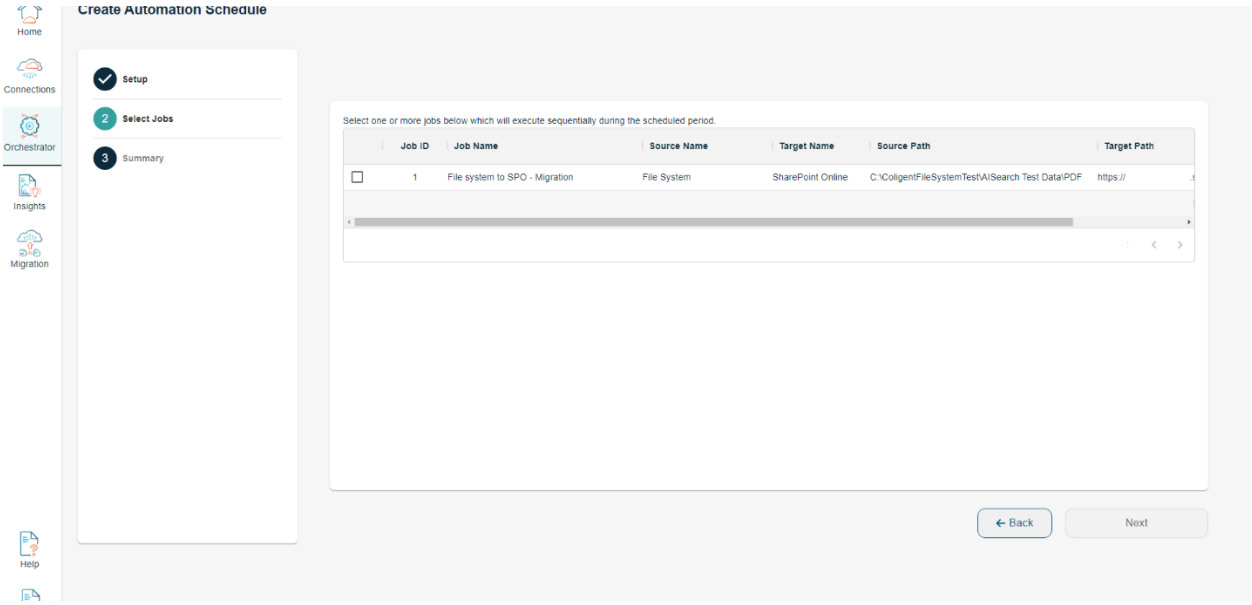


FIGURE112 - SELECT JOB

16.4. Orchestrator Summary

Before creating the job, review the details shown on the Summary screen and alter the job, if necessary, by clicking back button.

After reviewing details, click on Create Automation button which will create automation schedule.

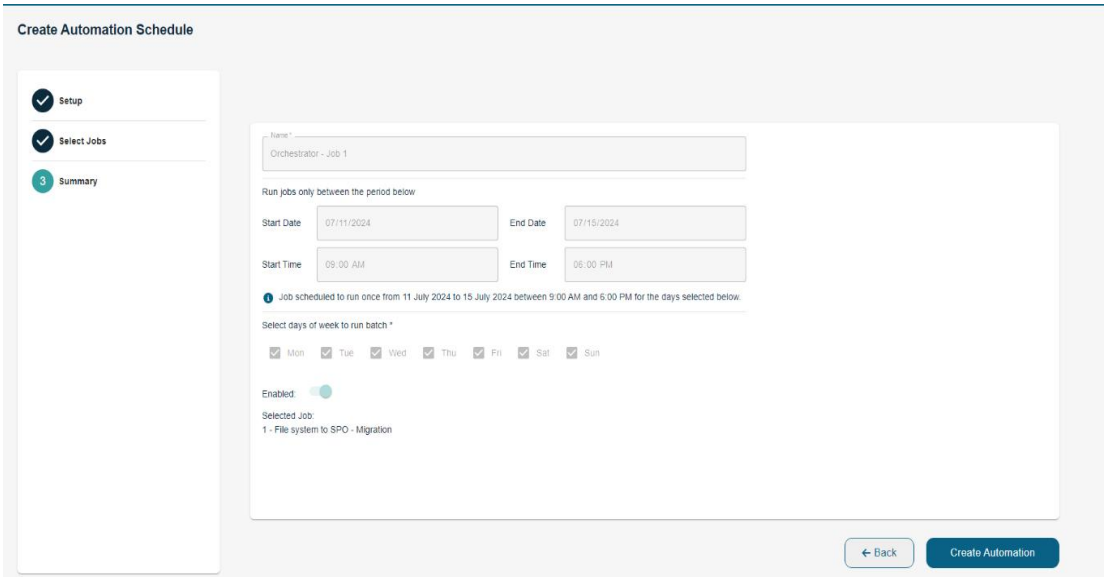


FIGURE 113 - SUMMARY

16.5 Orchestrator Dashboard

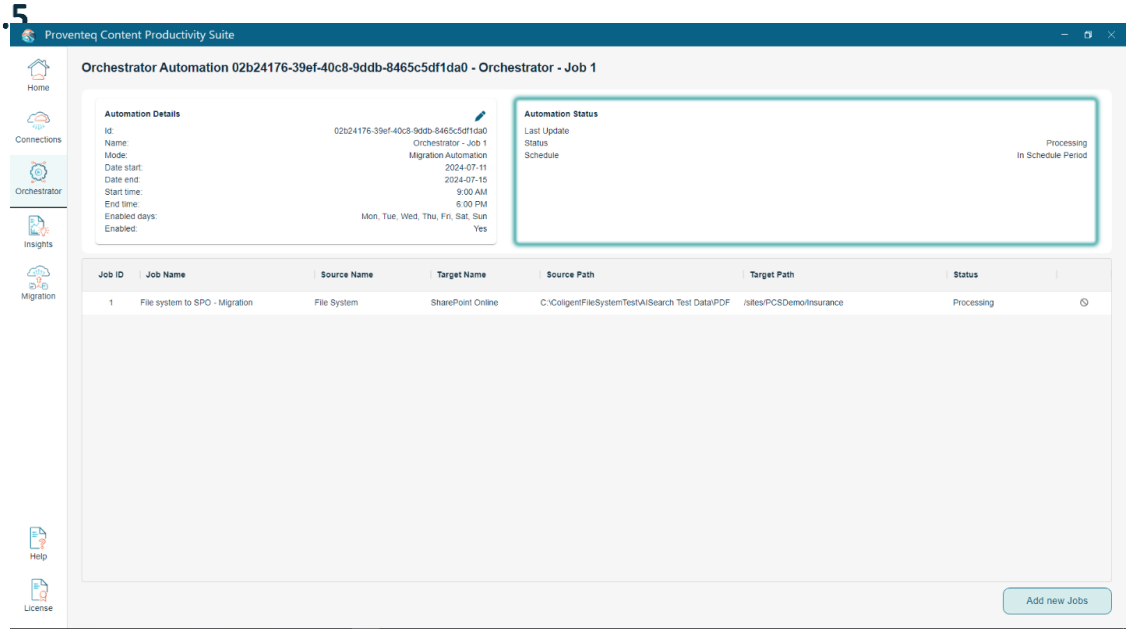


FIGURE 114 - ORCHESTRATOR DASHBOARD

16.5.1. Automation Details

This section shows information related to orchestrator job like Name, Schedule details (Start Date/time, End Date/Time), Days selected for execution and job enabled status.

We can also edit the orchestrator job details by selecting the edit icon on the top right corner of the section.

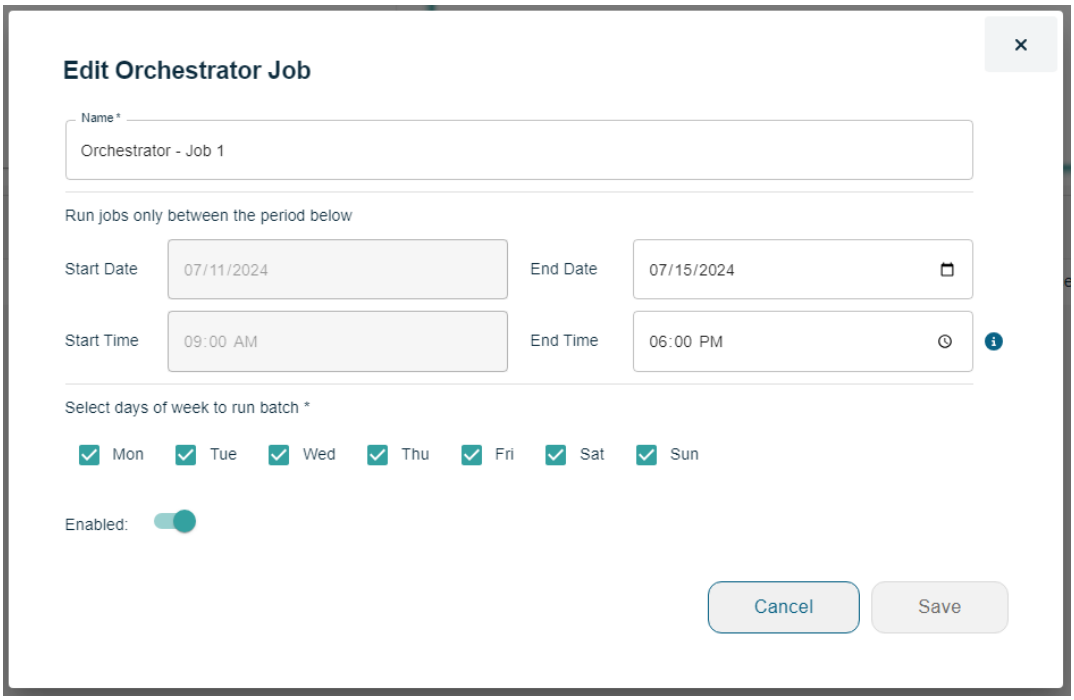


FIGURE 115 - EDIT ORCHESTRATOR JOB

Click on Save once the changes are made so they can reflect on the job.

Click Cancel if no changes required.

16.5.2. Automation status

This section shows information related to orchestrator job like Status of execution, status of schedule.

- Status: This shows the overall status of the selected job. Status is defined based on the following status of the jobs.

Job(s) status	Status
If All jobs are Pending	Pending
At least one job is processing	Processing
If all of the jobs are Completed Successfully	All jobs Completed Successfully
If all jobs are completed but one or more having any error	All jobs Completed, some with errors
If one or more jobs completed with errors and some are pending as well	Completed with errors, jobs pending
If some jobs are pending but processed job get completed without any error	Completed Successfully, jobs pending

- Schedule: This shows status of current schedule based on the following conditions
    - In Schedule Period - If Start date/time of schedule < current date/time
    - Out of Schedule Period - If Start date/time of schedule > current date/time

16.5.3. Remove or add more jobs

For all added jobs, there is a Remove icon displayed at the end of each row. User can remove added jobs until it's not picked up for process.

We can also add more jobs to a batch by clicking on the 'Add new job' button. When clicking on that, it will open a popup with list of other jobs available to add in current schedule.

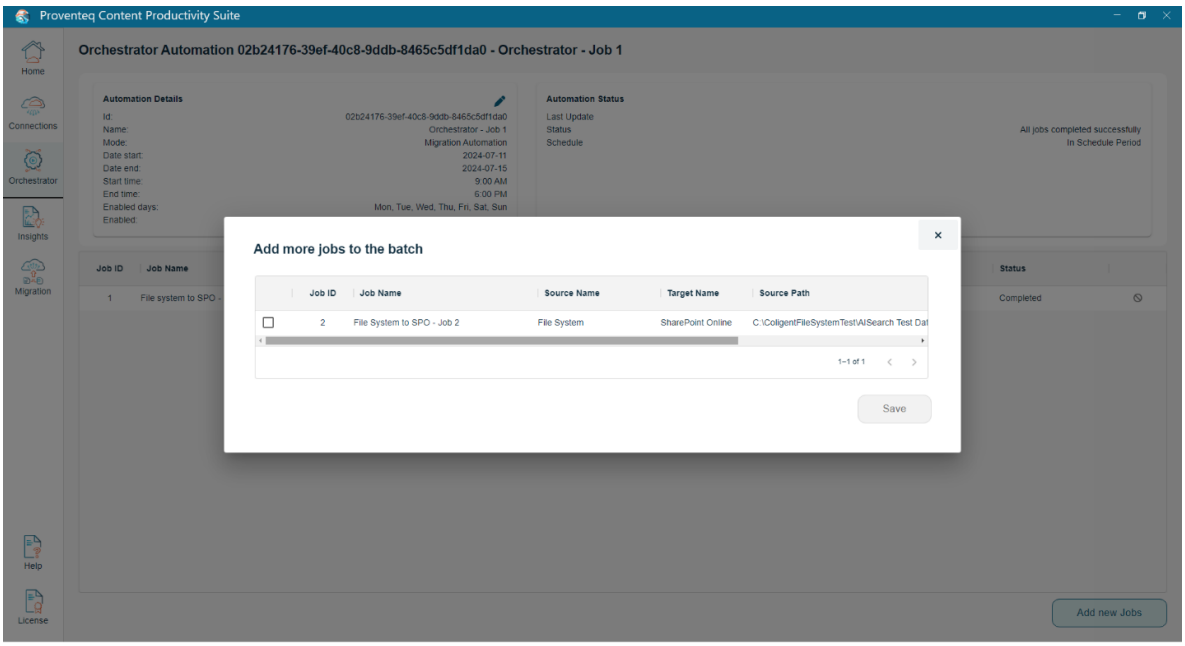


FIGURE 116 - ADD MORE JOBS

17. MONITOR

Proventeq Monitor is a component within the Proventeq Content Suite to monitor source content and permission oversharing analysis.

17.1. Installation and configuration

The pre-requites and installation instructions are contained within the CPS Installation Guide. Note no pre-requites related to the target need to be followed since Monitor does not require any target configuration for it to run.

Configuring Monitor - Creating a Monitor Job

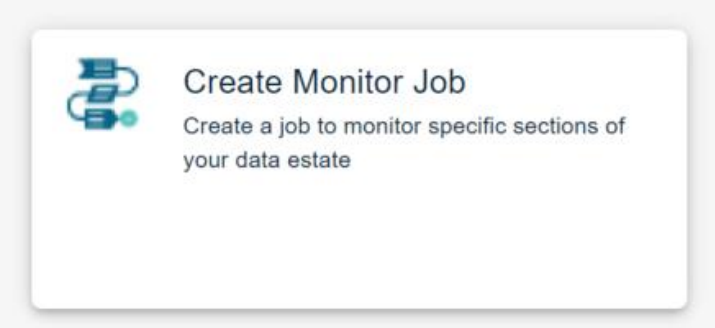
NOTE: If you have used CPS Migrator, you'll find Monitor uses a very similar UI framework.

A connection to the source needs to be created which can be done separately, or during the Monitor job creation process. Specify system and authentication details for the relevant source system and then further details of the source path to be processed need to be specified. See **Error! Reference source not found.**for details on adding source connections.

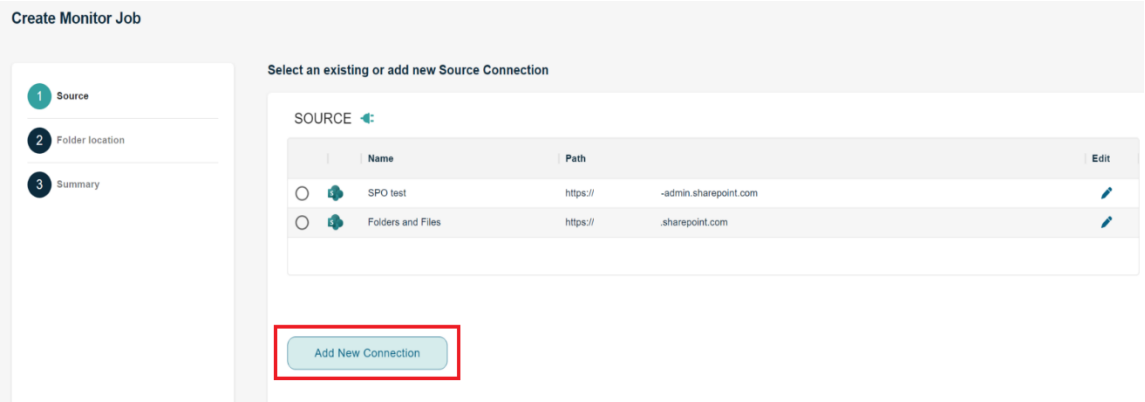
Once the Monitor job is created the Monitor dashboard is displayed

17.2. Add a New Monitor Job

To add a new monitor job, click on the Create Monitor job wizard



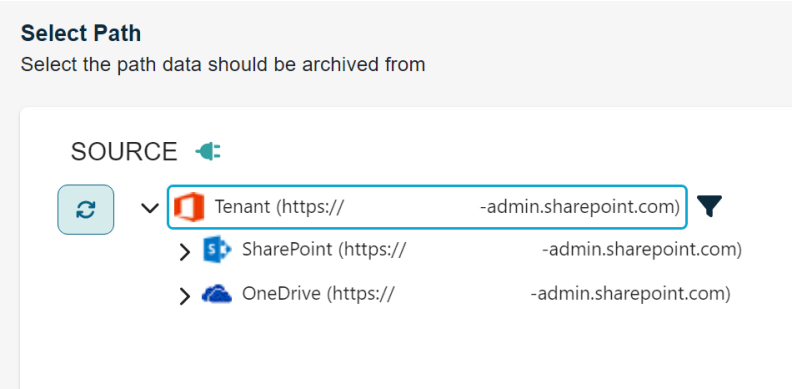
Now, click on the Add New connection, relevant security credentials will be required.



When we click on the Add New connection, the UI will be directed to the source connection screen. Please follow the steps provided in section 4.1.12 for adding SharePoint Online as a source.

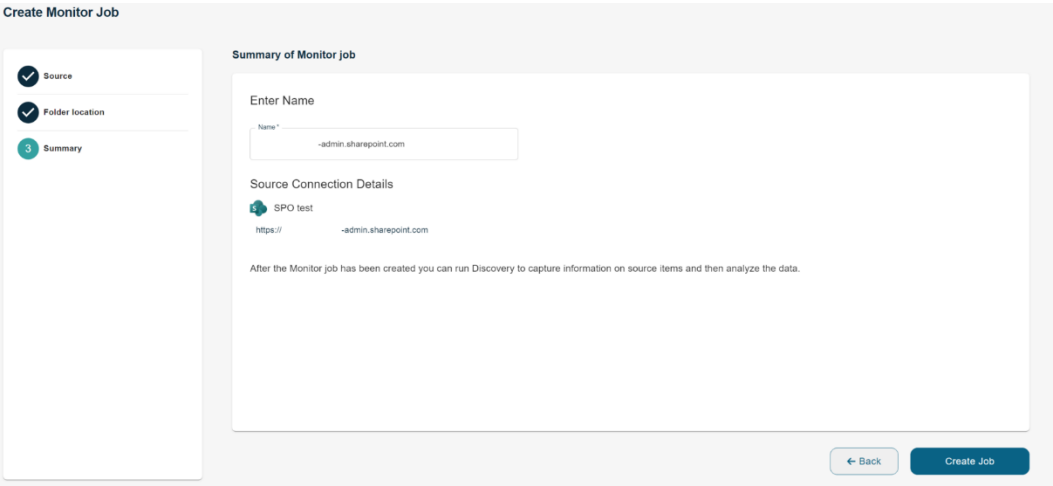
Once the connection to SPO is added as a source, you will be directed to select the Connection path as shown below.

Select Source Connection Path



Job Summary

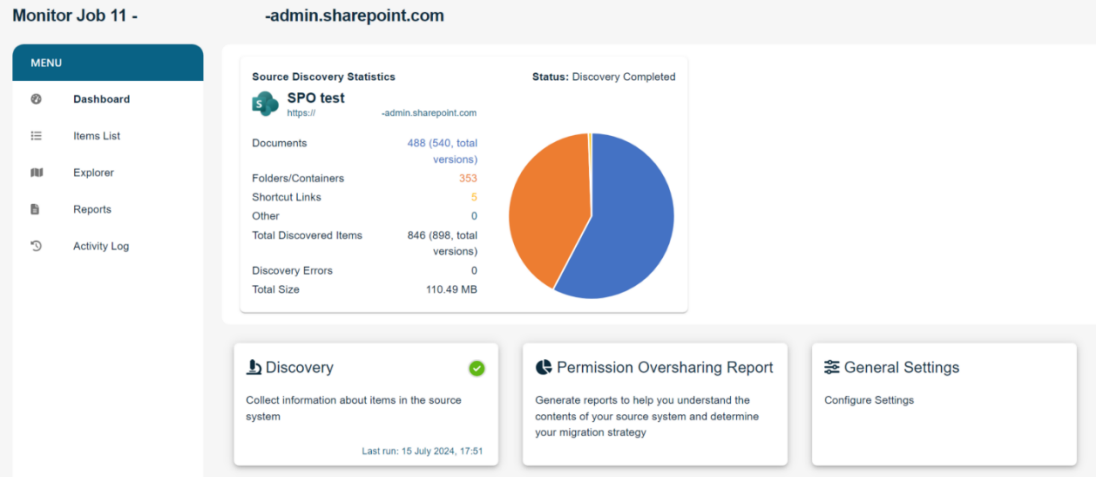
Before creating the job, review the details shown on the Job Summary screen and alter the job name if necessary else Click on Create Job.



Clicking **Create Job** will create the job and the Archive Dashboard will be displayed.

17.3. Monitor Dashboard

The Monitor Dashboard allows access to configuration and status information and execution of all operations. It is the main screen displayed when a job is selected from Monitor job and is also displayed after a job is created using the Monitor Job Wizard



17.3.1. Discovery Status

This section displays information gathered during the Discovery process.

Folders/Containers	Number of folders/containers found
Documents	Number of documents found with version and without version
Links	Number of shortcuts links found
Other	Any item types not covered by the above types e.g. ShareFile Virtual Collections

Total Discovered Items	Number of items successfully captured during discovery (excludes discovery errors)
Discovery Errors	Number of errors during discovery
Total Size	Total size of items discovered (includes all versions of the items)
Status	Status of Discovery

17.4. Monitor Item list

To launch the Items List, select the Items List menu option or launch from the Discovery dialog box. The Items List will display details of each item captured from the source during Discovery and will be updated by operations executed during Monitor.

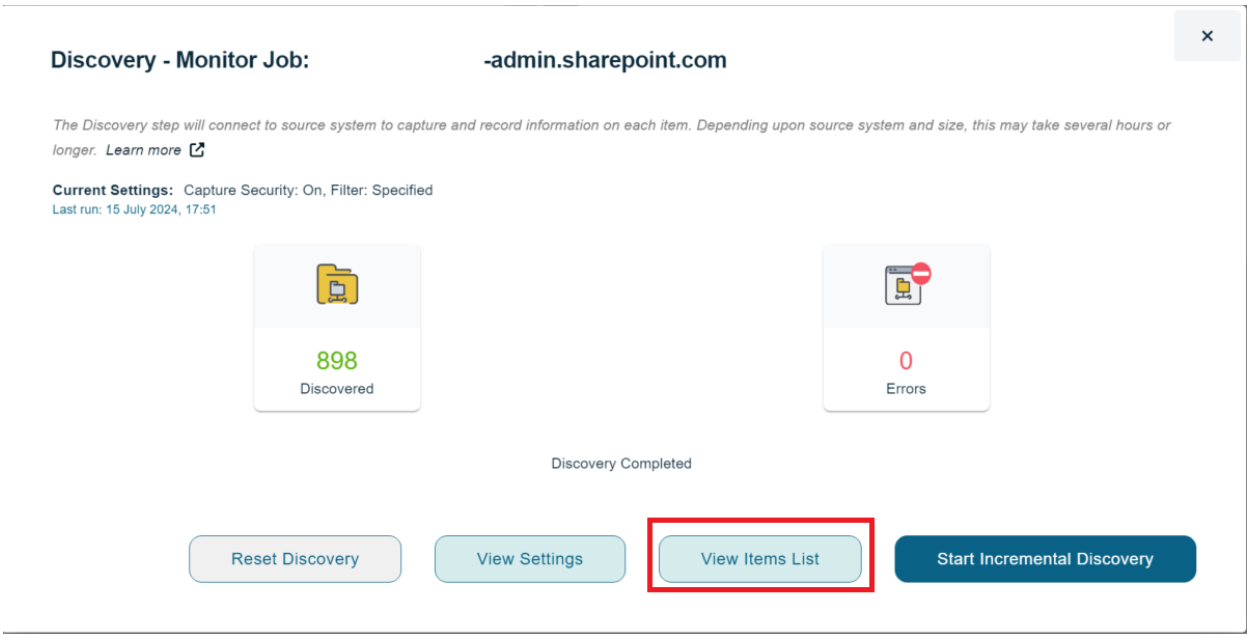


FIGURE 121 - LAUNCHING ITEMS LIST FROM DISCOVERY PROGRESS SCREEN

Columns displayed will vary depending upon the source being migrated. Filters can be applied to gain more Insights into the data being migrated and cleared using the × icon. Additional columns can be added or removed using the settings icon ⚙.

FIGURE 122 - FILTER

Export to CSV

The Item List can be exported into a .CSV file using the option. If a filter is applied to the items list, it will apply to the exported list also. All columns will be exported.

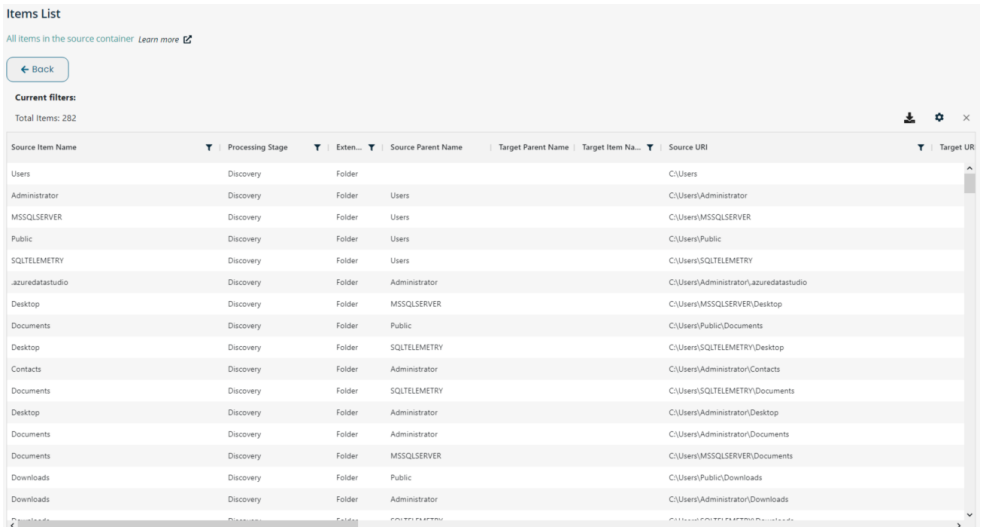


FIGURE 123 - ITEMS LIST

17.4.1. Item level information

Select the item to display the Item Details screen.

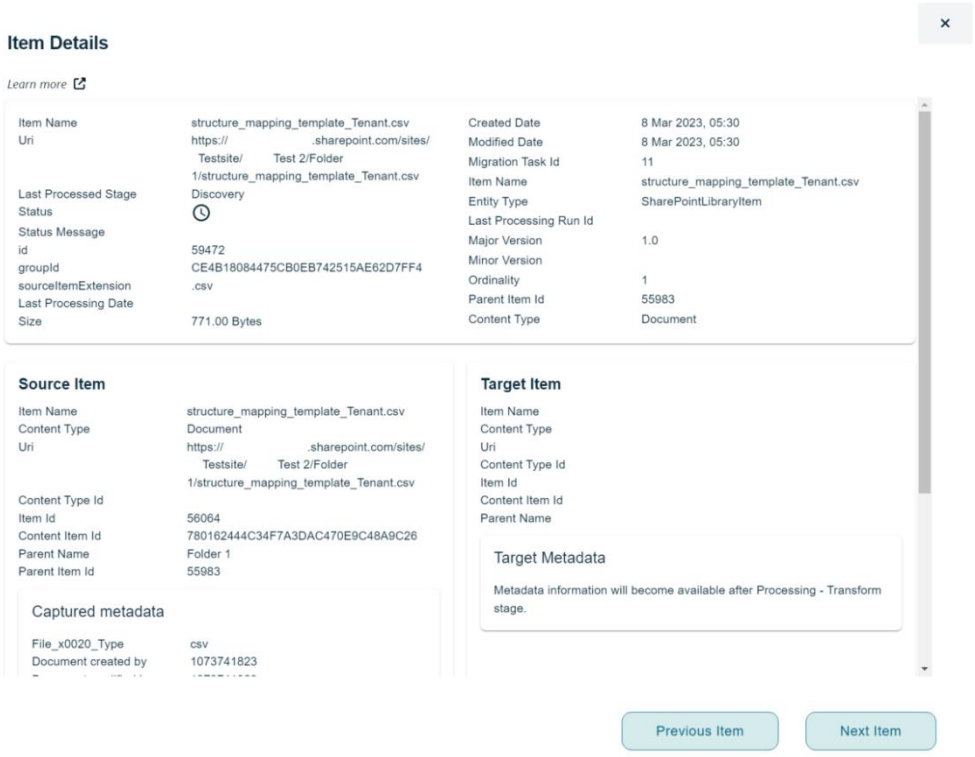


FIGURE 124 -: EXAMPLE ITEM DETAILS

17. Monitor Activity log

The activity log is accessible from the Menu option and displays the history of job processing. Each time Discovery or archive Step are executed the activity log is updated to reflect this.

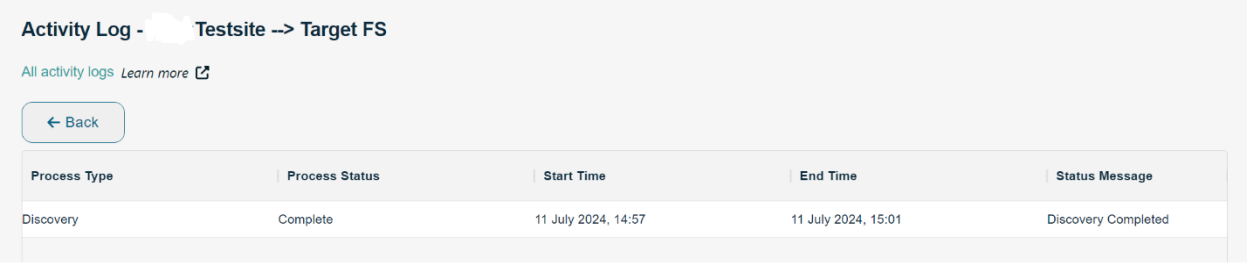
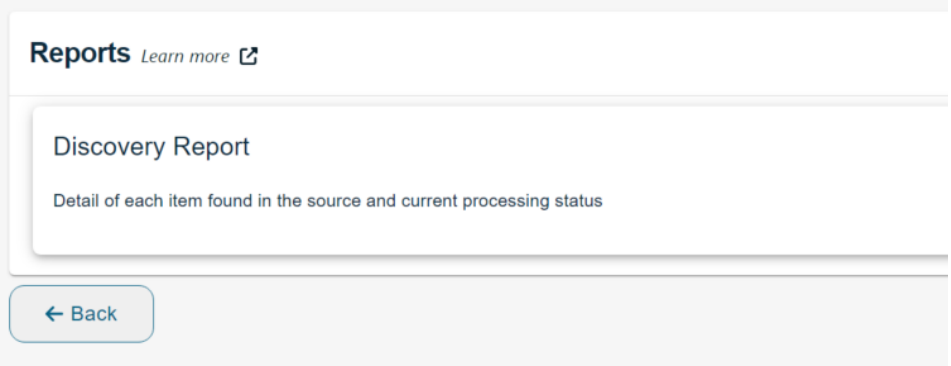


FIGURE 125 - ACTIVITY LOG

17.6. Monitor Reports

The reporting option provides access to the following reports.

- Discovery Report



17.6.1. Discovery Report

Displays a list of all items that have been discovered in the source along with additional metadata information and overall status of the item.

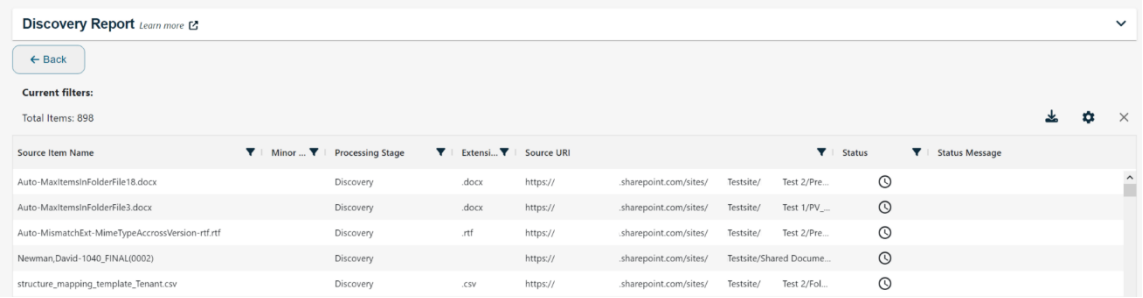


FIGURE 126 - EXAMPLE DISCOVERY REPORT

Monitor Discovery

After selecting a job the Monitor Dashboard is displayed, Select **Discovery.**

17.

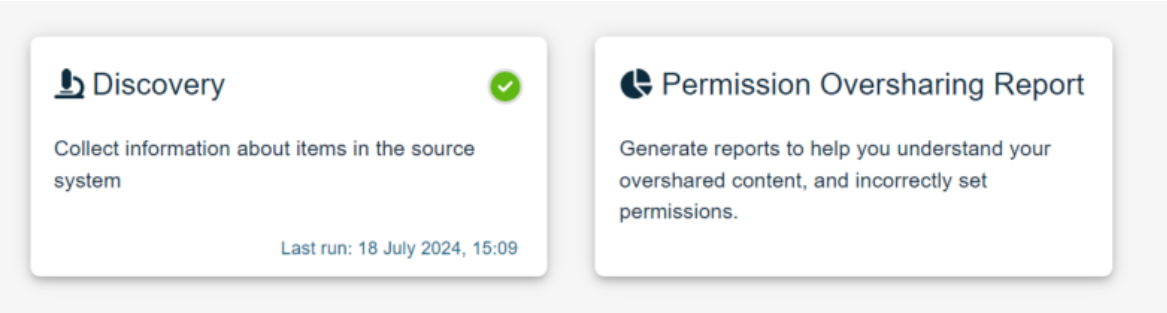


FIGURE 127 - DISCOVERY

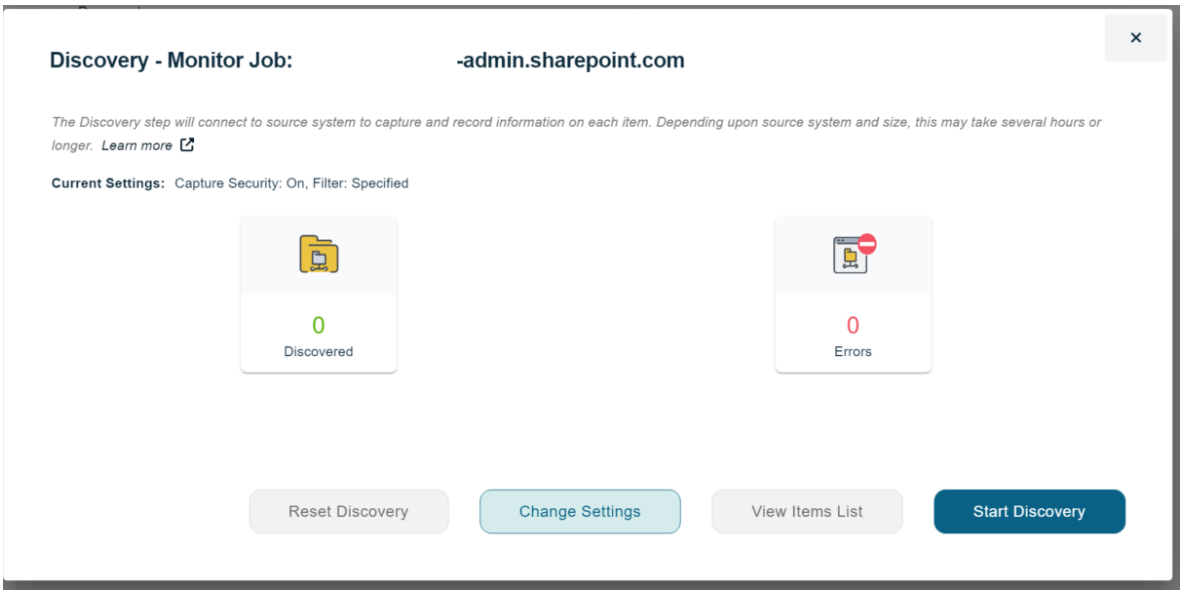


FIGURE 128 - STARTING DISCOVERY

Select **Start Discovery** option to begin the process of capturing information about the source environment into CPS. For many sources, the discovery progress will be displayed in the above screen.

After Discovery is complete, the discovery statistics information on the Monitor Dashboard will be updated to reflect the information captured during the discovery.

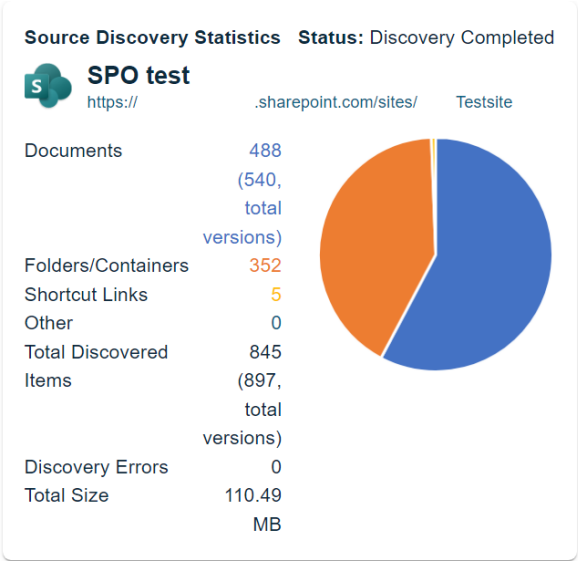


FIGURE 129 - DISCOVERY STATISTICS

17.8. Permission Oversharing report

Select the **Permission Oversharing report** option on the **Monitor Dashboard** so the PowerBI is opened.

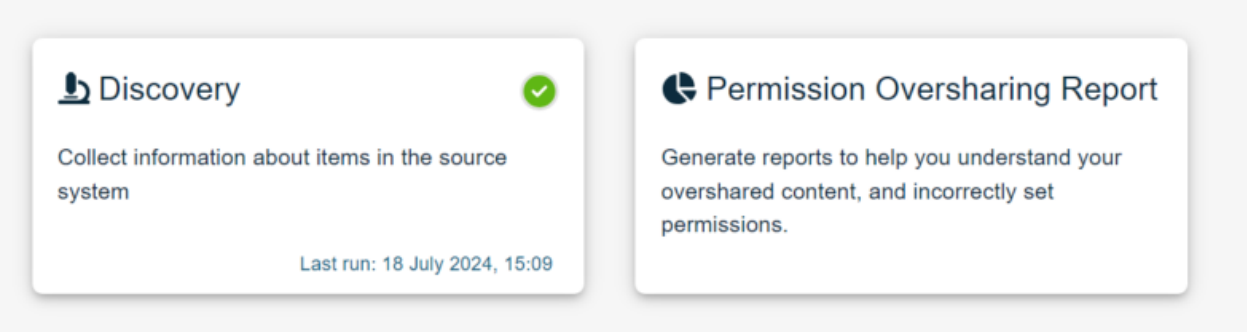


FIGURE 130 - PERMISSION OVERSHARING REPORT

Power BI Reports can be used to access the information in the Proventeq Migration Accelerator database to provide intuitive and interactive drill down reports. The report has 4 sections:

17.8.1. Introduction

This section gives the introduction to the report, Key concepts, how to use the report.

proventeq

Permission Oversharing Report

The Permissions Oversharing Report allows you to view and investigate the different types permissions assigned to content in your environment.

KEY CONCEPTS

The report highlights where permissions are specifically assigned to an object (Site / Library/ Folder/ Document). This can happen when a permission is set on a object by breaking the permission inheritance, or by creating a sharing link to share that document.

Permission Type

The permissions can be assigned **Direct** to a user or group by breaking inheritance. It can also be assigned as part of Sharing Links, which can be created to be accessible by the entire organisation with **view (SharingLinks.OrganizationView)** or **edit (SharingLinks.OrganizationEdit)** access, or assigned to **specific people (SharingLinks.SpecificPeople)**, or created as accessible by anyone i.e. Anonymous. **(SharingLinks.Anonymous)**

Risk profile

The permission are categorised into 3 different risk levels based on the type of permission

High risk : Permission assigned to *Everyone* or Everyone except external users, or shared with Anonymous user or any external user

Medium risk : Shared with anyone in the organization with ability to read or edit.

Low risk : If a permission was specifically assigned to a user or group

HOW TO USE THIS REPORT

Start with the **Permission Summary** report to understand the risk profiles and sites that need further investigation. Select a site from the main table, right click and "Drill Through" to the **Permissions Details** report which will provide more insight into permissions on each object in the site.

The **Permission Decomposition** report will allow you to see things from a users perspective, and understand what types of permissions are assigned to a user and which sites they have access to.

Introduction | Permission Summary | Permission Details | Permission Decomposition

FIGURE 131 - INTRODUCTION

17.8.2. Permission summary

This gives the overall summary for all the site discovered through CPS. It also highlights the Permission risk and site risk levels.

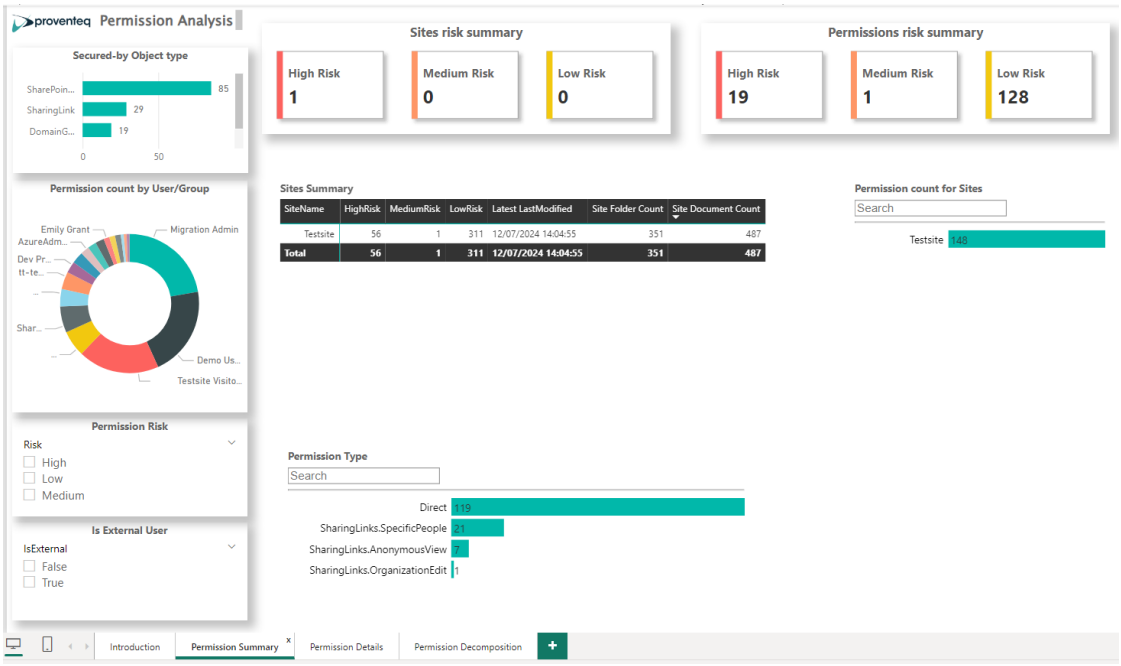


FIGURE 132 - PERMISSION SUMMARY REPORT

17.8.3. Permission Details

Select a site from the permission summary report, right click and drill through to view the details of the site in this report.

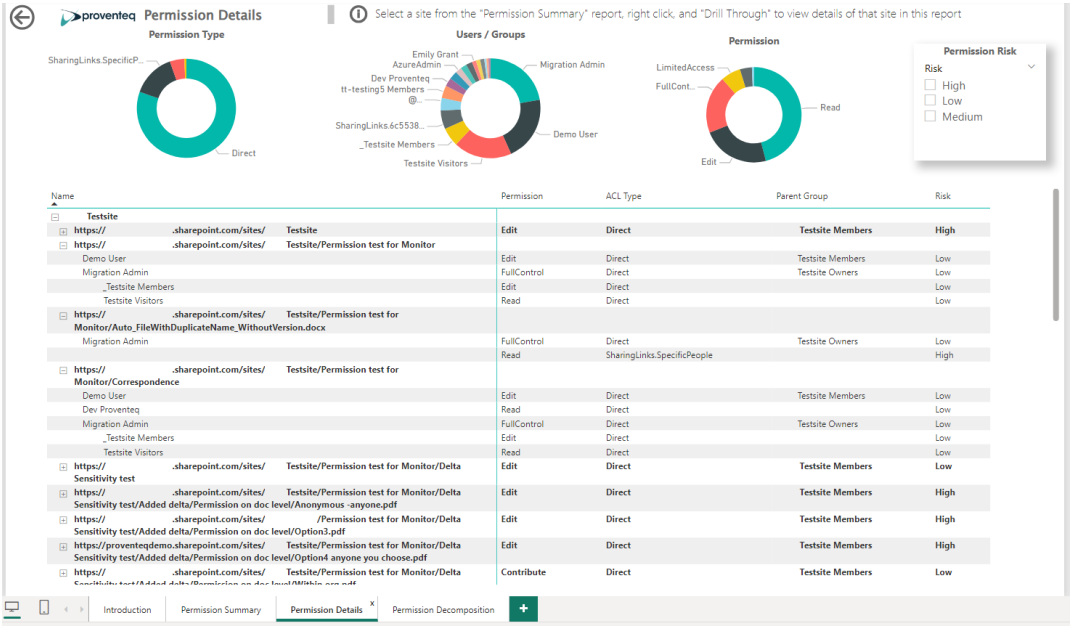


FIGURE 133 - PERMISSION DETAILS

17.8.4. Permission Decomposition

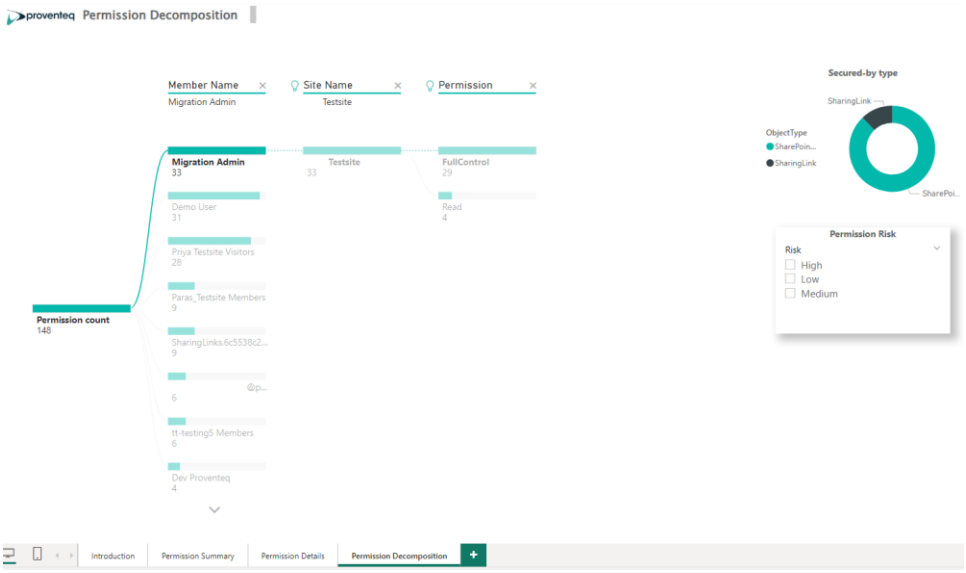


FIGURE 134 - PERMISSION DECOMPOSITION

18. APPENDIX

18.1. Licensing

Each CPS server will require an activated license file to allow the product to operate. A license can either be activated online (presuming there is an internet connection) or offline where an activation request file can be generated and once emailed and processed by Proventeq, an activation response file can be upload to complete the license activation.

A License can support several migration paths e.g. ShareFile and File Shares to SharePoint Online.

License consumption is based upon the volume of data, so customers may purchase a 200GB license etc. An item will consume license as soon as any Migration Phase operations, such as Extract metadata, Classify etc. So even if no migration into the target has taken place, license will still be consumed. Discovery does not consume license.

> NOTE: If the migration requires more than one CPS server then it's essential to understand the volume of data each server is going to migrate as license volumes are controlled on a per server basis. Incorrect amounts could lead to project delays due to license being fully consumed.

Delta migrations will consume license.

If the Undo operation is used, undo will free up license consumption for items processed up to steps 1-3 but NOT for migrated items. That means the undo of migrated items will effectively consume license twice.

CPS has a trial license and this is not based upon volume but the number of items discovered or migrated.

18.2. Discovery Filters

> Note: CSV based filters are only supported when running in non SaaS mode.

18.2.1. Discovery Filter for ShareFile

A filter can be applied to the Discovery. Using a filter can speed up the Discovery process, as well as reduce the size of the databases created by discovery and ensures that the database doesn't contain any items which do not have to be migrated.

Filters allow a list of folder locations to be specified in a CSV file, so only these folders (and their child) will be captured in the discovery process. Anything not discovered will not be processed for migration.

Discovery Filter Syntax:

[{"ObjectType":"CSV","Filter":"**<FilePath>**"}]

<FilePath> - File path of the CSV containing hierarchy path and app URI. It is mandatory to provide double backward slash '\\' in the file path.

E.g.

[{"ObjectType":"CSV","Filter":"F:\\Proventeq\\Configuration\\Product\\ShareFile\\Discovery Filter.csv"}]

Example Discovery Filter for ShareFile

FIGURE 130 - SAMPLE CSV FILE

	A	B
1	HierarchyPath	ApiUri
2	Virtual Root/Shared Folders/General Share/Test Data with Many items/Folder 03/Folder 02/Folder 01	https://mycompanydemouser.sf-api.com/sf/v3/Items(fo6a707a-6562-47ad-a4ee-c73dce9f890f)
3	Virtual Root/Shared Folders/General Share/Test Data with Many items/Folder 03/Folder 02/Folder 02	https://mycompanydemouser.sf-api.com/sf/v3/Items(fo4f4a49-31f2-44ec-be48-de5b65e8ceaa)
4	Virtual Root/Shared Folders/General Share/Test Data with Many items/Folder 01/Folder 01/Folder 02	https://mycompanydemouser.sf-api.com/sf/v3/Items(fo54e657-e647-4647-b491-94cd09f02bfa)

The AppURI can be found on the details screen on the selected folder in ShareFile administration portal.

18.2.2. Discovery Filter for FileShare

A filter can be applied to the Discovery. Using a filter can speed up the discovery process, as well as reduce the size of the databases created by discovery and ensures that the database doesn't contain any items which do not have to be migrated. Anything not discovered will not be processed for migration.

It possible to limit the discovery using a filter.

- Capture only information in folders specified in a specified .CSV file
- Capture items CREATED before specified date
- Capture items CREATED after specified date
- Items of specified file extensions
- All items, except those of a specified file extension

Syntax:

[{"FilterType":"<FilterType>","ObjectType":"<File or Folder>","Filter":"<Filter>"}]

Supported Filter Types

FilterType	Description	Object Type Supported
CSVInclude	Will discover folders provided in the CSV	Folder
DateBefore	Will discover files before the given date	File
DateAfter	Will discover files after the given date	File
ExtensionInclude	Will discover files which includes the given extension in CSV	File

ExtensionExclude	Will discover files which excludes the given extension in CSV	File

<ObjectType> - Folder or File depending upon FilterType.

<Filter> -

Example Limiting File Share discovery using CSVInclude

- [{"FilterType":"CSVInclude","ObjectType":"Folder","Filter":"F:\\Proventeq\\Configuration\\Product\\ShareFile\\Discovery Filter.csv"}]

Note: It is required to provide a double backward slash '\\' in any file paths (e.g. C:\\Proventeq\\MimeType.csv).

Example format for CSVInclude file.

A
HierarchyPath
E:\PMA Test Documents\Root\F2\F2-1
E:\PMA Test Documents\Root\F2\F2-3\F2-3-1
E:\PMA Test Documents\Root\F2\F2-4
E:\PMA Test Documents\Root\F2\F2-2
E:\PMA Test Documents\Root\F2\F2-3\F2-5

Example Limiting File Share discovery using CSVInclude using DateBefore

[{"FilterType":"DateBefore","ObjectType":"File","Filter":"7/5/2021"}]

For date filters, the date specified must be in the correct format as per the regional settings. This filter applies to the created date of the file.

Note: The date filter is not supported in delta discovery or processing.

Limiting Discovery using File Extensions - Using ExtensionInclude or ExtensionExclude

If the filter is an ExtensionInclude or ExtensionExclude list, a CSV file of relevant file extensions needs to be created. This is a simple list with each extension on a new line with the header "Type" e.g.

```
Type
.pdf
.txt
```

Extensions must begin with "."

Using multi file properties to control Discovery

Several filter criteria can be used together if required. The criteria act as an 'AND' so for items to be discovered they must meet all criteria.

- [{"FilterType":"ExtensionInclude","ObjectType":"File","Filter":"
  C:\\Proventeq\\MimeType.csv
  "}, {"FilterType":"DateAfter","ObjectType":"File","Filter":"7/5/2021"}}

18.2.3. Discovery Filter for Documentum

Use the below discovery filter options to filter out cabinets, documents or folders while performing discovery. There are three types of discovery filter options supported:

- REST Service based Filter
- DQL Query based Filter
- CSV based Filter
- SQL based Filter

> **Note** Documentum DB connect type supports CSV and SQL based filter. Documentum API connection type only supports Rest Service and DQL Query based filters.

REST Service Filter

REST service filter can be used to discover the folders or documents by iterating through give cabinets and folders hierarchy structure. So, it will discover the filtered documents or folders along with cabinet or folder structure.

Filter Format: [{"FilterType":"**REST**","ObjectType":<ObjectType>,"Filter":"<Filter Criteria>"}]

Object Types:
The following Object Types are supported (Use numeric values):

- RootFolder = 2
- SubFolder=3
- Item = 4

Filter:
The following functions are supported in the filter

- contains
- starts-with
- between

Rest Filter Examples

Filter Criteria template for RootFolder with Contains function:

[{"FilterType":"**REST**","ObjectType":2,"Filter":"contains(object_name, 'Di') or contains(object_name, 'A')"}]

By using this filter expression, the request only returns objects whose object_name contains A.

Filter Criteria template for Items with starts-with function:

The starts-with function checks the starting string of a property.

[{"FilterType":"**REST**","ObjectType":4,"Filter":"starts-with(object_name, 'D')"}]

By using this filter expression, the request only returns **Items** whose object_name starts with D.

Filter Criteria template for Items with between function:

The between function returns items between a specified range. Check the format of the date provided in the filter.

[{"FilterType":"**REST**","ObjectType":4,"Filter":"between(r_modify_date, date('**<date1>**'), date('**<date2>**'))"}]

Eg:

[{"FilterType":"**REST**","ObjectType":4,"Filter":"between(r_modify_date, date('2014-03-02'), date('2017-09-10'))"}]

By using this filter expression, the request only returns objects whose r_modify_date is between 2014-03-02 and 2017-09-10.

DQL Query Filter

The DQL filter can only be used if connecting to Documentum API.

The DQL filter can be used to filter the documents across repository or in a particular cabinet/folder in a flat structure i.e. without iterating through cabinets/folders structure. So, discovery result will discover only filtered documents (similar to Documentum Search) without any cabinet or folder and its hierarchy/structure. DQL filter is not supported at virtual source containers like public or private cabinets level.

Filter Format: [{"FilterType":"DQL","ObjectType":"<ObjectType>","Filter":"<Condition1> **AND** (<Condition2> **OR** <Condition3>)"}]

Here Filter Type will be as DQL, Object Type is the content or object type in the Documentum and Filter will be suitable DQL filter to discover specific set of documents.

Filter Criteria template for Items within specific Public/User Cabinet:

This filter will bring all the files under respective public/user cabinet id.

[{"FilterType":"**DQL**","ObjectType":"**dm_document**","Filter":"**i_cabinet_id = '<Cabinet_ID>'**"}]

Filter Criteria template for specific document and its versions:

This filter will bring given document id with all the versions.

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"i_chronicle_id ='**<Chronicle_ID>**'"}]

Filter Criteria template for Items between dates:

This filter will bring all the files creation date and modified date in between Date1 and Date2.

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"<Creation Date> > date('<date1>') and <Creation Date> < date('<date2>')"}]

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"<Modified Date> > date('<date1>') and <Modified Date> < date('<date2>')"}]

Filter Criteria template for specific document name:

This filter will bring the files with specific File Name.

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"object_name ='**<File Name>**'"}]

Filter Criteria template for Items containing repeatable attribute values:

This filter will bring the files with repeating attributes like tags having <Tag1> and <Tag2> values.

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"any **<Repeating_Attribute>** IN ('<Tag1>','<Tag2>') "}]

Eg:

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"any keywords IN ('KeywordA','KeywordB') "}]

Filter Criteria template for Items containing given name:

This filter will bring the files and its versions where <attribute> starts with <FileName>, ends with <FileName> and contains <FileName> respective to the Percentage(%) symbol specified.

A. This will bring <FileName> starts with given name.

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"<attribute> like '<FileName>**%**' "}]

B. This will bring <FileName> ends with give name.

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"<attribute> like '**%**<FileName>'"}]

C. This will bring <FileName> containing given name.

[{"FilterType":"DQL","ObjectType":"dm_document","Filter":"<attribute> like '%<FileName>%'"}]

CSV based Filter

A CSV based filter can be user to restrict the discovery based upon Documentum Cabinet ID for DocumentumDB connections. It will return all folders and items contained within the specified cabinets.

Filter Format:[{"FilterType":"**CSV**","ObjectType":"dm_cabinet","Filter":"**<FilterCriteria>**"}]

FilterCriteria: One or more cabinet ids(i_cabinet_id) separated by comma.

Example: Filter Criteria template for single cabinet:

[{"FilterType":"**CSV**","ObjectType":"**dm_cabinet**","Filter":"0c0007c580009548,0c0003f270008498"}]

SQL based Filter

The SQL filter can be used to filter the documents across repository or in a particular cabinet/folder based on document's standard properties like Name, Id, Object Type, Created or Modified date etc.

Filter Format: [{"FilterType":"SQL","ObjectType":"" Filter":"<Condition1> AND (<Condition2> OR <Condition3>)"}]

Here Filter Type will be SQL, Object Type is not required, and Filter will be suitable SQL filter conditions to discover specific set of documents based on below Documentum View columns and their values.

Database View Name	Alias	Columns
dm_sysobject_sp	SP	i_chronicle_id, r_object_id, r_object_type, a_content_type, object_name, r_version_label, r_creation_date, r_modify_date, owner_name, r_creator_name, r_modifier, i_is_deleted, r_is_virtual_doc, i_latest_flag, a_storage_type, r_access_date, acl_name
dmr_content_sp	C_SP	storage_id, full_content_size, rendition

- Filter Criteria to discover files after certain creation and modify dates:

This filter will bring all the files creation date and modified date.

[{"FilterType":"SQL","ObjectType":"","Filter":"(C_SP.STORAGE_ID is null OR SP.r_creation_date > '<Creation_Date>')"}]

[{"FilterType":"SQL","ObjectType":"","Filter":"(C_SP.STORAGE_ID is null OR SP.r_modify_date > '<Modify_Date>')"}]

- Filter Criteria to skip 0KB documents:

This filter will skip the files having no content associated.

[{"FilterType":"SQL","ObjectType":"","Filter":"(C_SP.STORAGE_ID is null OR C_SP.full_content_size > 0)"}]

- Filter Criteria to skip deleted items:

This filter will skip the files having no content associated.

[{"FilterType":"SQL","ObjectType":"","Filter":"sp.i_is_deleted = 0"}]

- Filter Criteria for specific document and its versions:

This filter will bring given document id with all the versions.

[{"FilterType":"SQL","ObjectType":" ","Filter":"(C_SP.STORAGE_ID is null OR sp.i_chronicle_id ='<Chronicle_ID>')"}]

- Filter Criteria to discover certain object type files:

This filter will bring given document id with all the versions.

[{"FilterType":"SQL","ObjectType":" ","Filter":"(C_SP.STORAGE_ID is null OR sp.r_object_type ='<Object_Type>')"}]

Flatten discovery using CSV based Filter

Flatten discovery using CSVFile based filter is used for discovering documents underneath the provided r_object_id of the folder and its parent id (optional) in the CSV file

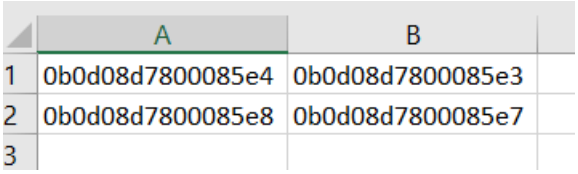
Filter Format: [{"FilterType":"CSV","ObjectType":"<ObjectType>","FlatDiscovery":true or false, "Filter":"<Path of the file>"}]

Filter: Filter contains the path of the file containing r_object_id of the folder and its parent id (optional) separated by comma.

Examples:

- Filter Criteria template:

[{"FilterType":"CSVFile","ObjectType":4,"FlatDiscovery":true,"Filter":"C:\\Documents\\Flat_discovery_build\\flat_discovery -modified.csv"}]



	A	B
1	0b0d08d7800085e4	0b0d08d7800085e3
2	0b0d08d7800085e8	0b0d08d7800085e7
3		

The r_object_id should be provided in column A and its parent folder id in column B. Note that the \ must appear twice in folder paths.

18.2.4. Discovery Filter for OpenText

Filters can be applied for OpenText systems using Oracle or Microsoft SQL databases. Filtering can be applied to folders and documents.

These are list of tables and fields that can be used within filters.

Table Alias	Field
A for DTreeCore	DATAID, NAME, USERID, CreatedBy, ModifiedBy, CreateDate, Createdate, Modifydate, SubType, VersionNum
V for DVersData	VersionID, DocID, Version, VersionName, Owner, VerCDate, VerMDate, FILENAME, FileType, MimeType
K for KUAF	CreatedBy

K2 for KUAF	ModifiedBy
VK for KUAF	Owner

Below are the Sub Types available in OpenText for different entities which can be used in filters:

- 0 - Folder
- 144 - Documents
- 749 - Email
- 751 - Email Folder
- 136 - Compound Document
- 298 – Collection
- 141 - Enterprise Workspace
- 142 – User Workspace
- 848 – Business Workspace
- 849 – Business Workspace Root
- 202 – Project
- 1 – Shortcut links

Examples:

Filter based on Name from DTREECORE Table

To discover files where the A.Name property begins with specified value.

Product	Example Query
SQL	[{"ObjectType":"SQL","Filter":"(A.Name like '<Name>%' AND A.SubType in (144)) OR (A.SubType in (0,751,298,136))"}]
Oracle	[{"ObjectType":"SQL","Filter":"(A.Name like '<Name>%' AND A.SubType in (144)) OR (A.SubType in (0,751,298,136))"}]

Or even
[{"ObjectType":"SQL","Filter":"(A.Name like '<Name>%' AND A.SubType in (0,144,751,298,136))"}]

Just like standard SQL query syntax, use %. For example, <Name>% to find records that end with specified value or %<Name>% to find records that contain specified value.

Filter criteria based on CreateDate field from DTREECore Table:

To discover files with Created Date greater than or equal to specified date. Note: Date should be in ISO format.

Product	Example Query
SQL	[{"ObjectType":"SQL","Filter":"A.CreateDate >'2021-01-01' AND (A.SubType = 144 OR (A.SubType in (0,751,298,136)))"}]
Oracle	[{"ObjectType":"SQL","Filter":"A.SubType In (0, 144) AND A. CreateDate > TO_DATE('2021-01-01', 'yyyy-mm-dd') OR (A.SubType in (0,751,298,136))"}]

Filter criteria based on VerMDate field from DVersData Table:

To discover versions of documents with a modified date greater than specified date.

Product	Example Query
SQL	[{"ObjectType":"SQL","Filter":"(V.VerMDate>'2021-05-06' AND A.SubType in (144)) OR (A.SubType in (0))"}]
Oracle	[{"ObjectType":"SQL","Filter":"(V.VerMDate> TO_DATE('2021-05-06', 'YYYY-MM-DD') AND A.SubType in (144)) OR (A.SubType in (0))"}]

Filter criteria based on MimeType field from DVersData Table:

To discover document using MimeType of documents like docx. MimeType value need to get from oracle DB as storing of mimetype in orcle DB id not conventional

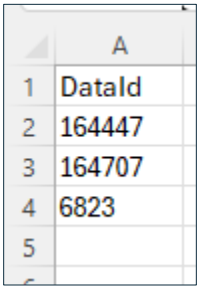
Product	Example Query
Oracle	**ORCALE:** [{"ObjectType":"SQL","Filter":"(V.MimeType='**application/vnd.openxmlformats-officedocument.wordprocessingml.document'** AND A.SubType in (144)) OR (A.SubType in (0))"}]

CSV Based Filter

To discover folders CSV based discovery filter can be used. This will be Folder based discovery only and so it will not work for files

Product	Example Query
SQL	[{"ObjectType":"csv","Filter":"D:\\LocalFiles\\Insurance.csv"}]
Oracle	[{"ObjectType":"csv","Filter":"D:\\LocalFiles\\Insurance.csv"}]

User below format to create CSV file and includes DATAID of the folder which need to discover.



	A
1	DataId
2	164447
3	164707
4	6823
5	

18.2.5. Discovery Filter for Filenet P8

The filter capabilities depend upon whether the FileNet P8 connection is configured to use API or Database.

API Filter

Filter Format:

[{"ObjectType":"2","Filter":"<Filter Criteria>"}]

The Object type determines the type of objects that can be found in the filter. The value of "2" represents "Files", as this is the only type supported.

Examples:

- **Filter criteria based on Document Title:**

  [{"ObjectType":"2","Filter":"(d.DocumentTitle like '<DocumentTitle>%')"}]

- **Filter criteria based on File Creator:**

  [{"ObjectType":"2","Filter":"(d.Creator like '<FileCreatorName>%')"}]

Just like standard SQL query syntax, use %. For example, <FileCreatorName>% to find records that end with specified value or %<FileCreatorName>% to find records that contain specified value.

- **Filter criteria based on Date Modified**

  To discover the files based on Date created that is greater than or equal to the given <Date>. The date should be in ISO format.

  [{"ObjectType":"2","Filter":"d.DateLastModified >= <DateModified>"}]

- **Filter criteria based on Class or Content Type:**

  To discover files based on Class or Content Type.

  [{"ObjectType":"2","Filter":"(IsOfClass(d, <ContentType>))"}]

  Example a list of items with a Content Type/Class of “Communication” or “Policy”

  E.g. [{"ObjectType":"2","Filter":"(IsOfClass(d, Communication) OR IsOfClass(d, Policy))"}]

- **Filter criteria based on Metadata:**

  To discover the files based on metadata properties.

  [{"ObjectType":"2","Filter":"(<PropertyName>='PropertyValue')"}]

  E.g. [{"ObjectType":"2","Filter":"(BankName='Indian Bank' OR BankName='Overseas Bank')"}]

Database Filter

Note: The Database filter requires table alias and database column/fields name. e.g. to apply filter on DocumentTitle column name in database must use table alias “d” and column name “u1708_documenttitle”.

Filter Format: [{"ObjectType":<ObjectType>,"Filter":"<Filter Criteria>"}]

Table Alias	Fields
D for DocVersion	object_id, mime_type , version_series_id, major_version_number , minor_version_number, create_date, creator, modify_date,

	modify_user, compound_document_state , 1708_documenttitle , Modify_user and custom property fields
CR for Current DocVersion	object_id, mime_type , version_series_id, major_version_number , minor_version_number, create_date, creator, modify_date, modify_user, compound_document_state , 1708_documenttitle , Modify_user and custom property fields
CD for ClassDefinition	symbolic_name

D queries against all versions, **CR** queries against latest version of documents.

Object Type:

The following Object Type is supported (Use numeric value):

- File = 2

Examples:

Filter criteria based on Document Title:

To discover the files that starts with given <DocumentTitle>.

[{"ObjectType":"2","Filter":"(D.u1708_documenttitle like '<DocumentTitle>%')","Uri":"","IsRecursive":true}]

Just like standard SQL query syntax, use %. For example, <DocumentTitle>% to find records that end with specified value or %<DocumentTitle>% to find records that contain specified value.

- **Filter criteria based on File Modifier:**

To discover the files based on Modifier that starts with given <FilemodifierName>.

[{"ObjectType":"2","Filter":"(D.modify_user like '%<FilemodifierName>')","Uri":"","IsRecursive":true}]

- **Filter criteria based on Date Last Modified:**

To discover documents were the latest version has a modified date greater than specified date.

[{"ObjectType":"2","Filter":"CR.modify_date>'<ModifiedDate>'","Uri":"","IsRecursive":true}]

The date can be in the format depending on FileNet datetime format.

E.g. [{"ObjectType":"2","Filter":"CR.modify_date>'2020-06-01'","Uri":"","IsRecursive":true}]

- **Filter criteria based on Class or Content Type:**

To discover the files based on Class or Content Type.
[{"ObjectType":"2","Filter":"(CD.symbolic_name=’<ContentType>')","Uri":"","IsRecursive":true}]

E.g. [{"ObjectType":"2","Filter":"(CD.symbolic_name='Communication' OR CD.symbolic_name='Policy')","Uri":"","IsRecursive":true}]

- **Filter criteria based on Metadata:**

  To discover the files based on metadata properties.

  [{"ObjectType":"2","Filter":"(D.<PropertyName> ='<PropertyValue>')", "Uri":"", "IsRecursive":true}]

  E.g. [{"ObjectType":"2","Filter":"(D.u06a8_bankname='Indian Bank' OR D.u06a8_bankname='Overseas Bank')","Uri":"","IsRecursive":true}]

18.2.6. Discovery Filter for IManage

Filters can be used to limit discovery to workspaces, container(folders) and documents by allowing queries against RootContainer, Container and Item tables respectively.

Filter Type	Description	Filter Query Syntax
Root Container (Workspace)	To filter particular workspace	{"RootContainer" : "Root.PRJ_NAME ='<Workspace Name>'"}
	To filter multiple workspaces	{"RootContainer" : "Root.PRJ_NAME in ('<Workspace Name>', '<Workspace Name2>')"}
		{ "RootContainer”: "Root.PRJ_ID IN (Select PRJ_ID from Temptable where Department ='Natural Resources and Environment')", "Item" : "Item.C4ALIAS = 'Natural Resources and Environment'" }
Container	To filter specific folder.	{"RootContainer" : "Root.PRJ_NAME ='<Workspace Name>'","Container" : "(Container.ItemName = '<Folder

		Name>' AND Container.EntityType ='Folder')"}
Item	To filter specific item inside the specific folder.	{"RootContainer" : "Root.PRJ_NAME ='<Workspace Name>'","Container" : "(Container.ItemName = '<Folder Name>' AND Container.EntityType ='Folder')","Item" : "Item.DoNum = '<Document Number>'"}
	To filter items with date created filter	{"Item" : "Item.EntryWhen > '<Date>')"}
	To filter items with date modified filter	{"Item" : "Item.EditWhen > '<Date>')"}

The **<Workspace Name>, <Folder Name>, <Document Number>** and **<Date>** should be replaced with the respective values as mentioned in above query. Standard SQL operators can be used in these filters.

18.2.7. Discovery Filter for M-files

Filters can be used to limit discovery to specific classes

Filter Criteria template:

Filter Syntax:
[{
"FilterType":"ClassList",
"Filter":"C:\\mfilesclasslist\\discoveryfilter.csv"
}]

Example File: discoveryfilter.csv .Note: ClassName is a mandatory column heading.
ClassName
Contracts

Filter Syntax: [{FilterType":"FileList","Filter":"C:\\temp\\mfiles_file_list.csv"}]

Example File: mfiles_file_list.csv. Note: FileId is a mandatory column heading and Type can be optional
FileId,Type
1
2
44444
7
3,101
4,101

18.2.8. Discovery Filter for Alfresco

Filters can be used to limit discovery to specific sites.

Filter Criteria template for sites:

Filter Syntax:

[{"ObjectType":"Site","Filter":"<**SiteName**>"}]

Example - Discovery all folders and documents within "Insurance" site.

[{"ObjectType":"Site","Filter":"Insurance"}]

Note: To filter multiple sites a comma separated list can be provided. E.g.

[{"ObjectType":"Site","Filter":"Insurance,HR"}]

Filter Criteria template for filelists:

Filter Syntax:

[{FilterType":"FileList","Filter":"<CSV File> "}]

Example –

[{FilterType":"FileList","Filter":"C:\\temp\\mfiles_file_list.csv"}]

Example File: mfiles_file_list.csv. Note: FileId is a mandatory column heading and Type can be optional

FileId,Type
1,
2,
44444,
7,
3,101
4,101

The FileID values can be found by exporting from a search view in M-Files.

Type is the M-Files object type (not class), leave it blank unless they know that its different from the standard.

18.2.9. Discovery Filter for Oracle UCM

Filter is only support for items not contained within folders, i.e. no parent folder. Filtering on Contribution Folders and Framework Folders is not supported.

Filter Syntax:

dDocName='<DocumentName>'
dDocType='<DocumentType>'

Examples

dDocName = 'PROV_PROV_090007C280000583'
dDocType = 'Document'

18.2.10. Discovery Filter for Hummingbird/EDocs

You can filter upon documents and folders based on a variety of criteria like Matter, SubMatter, DOCNUMBER, etc. The filter needs to contain the SQL clause filter string like below:

Dynamic Filter	Description
DP.E_DIVPROJ =’<FolderID>’	This filter will get folders and documents of respective folder ID.
DP.E_DIVPROJ IN (<FolderID>,<DocumentID>)	This filter will get folders and documents of respective folder ID and particular Document ID.
DP.CREATION_DATE BETWEEN ‘<Date1>' AND '<Date2>'	This filter will get the documents created between the Date1 and Date2.
DP.CREATION_DATE > '<Date>'	This filter will get the documents created from the given date.

18.2.11. Discovery Filter for BOX

A filter can be applied to the Discovery. Using a filter can speed up the discovery process, as well as reduce the size of the databases created by discovery and ensures that the database doesn’t contain any items which do not have to be migrated. Anything not discovered will not be processed for migration.

Discovery Filter Syntax:

It is required to provide a double backward slash ‘\\’ in any file paths (e.g. C:\\Proventeq\\.csv).

Inclusion Filter - Include only specified users/folders:

[{"ObjectType":"CSV","Filter":"path\\to\\infil.csv"}]

The inclusion filter requires a URL of a folder to include in the discovery (FolderURL), and the username or login email of the owner (UserName, this is required due to how Box handles permissions and service accounts).
A user's content root (the "All Files" folder) can be included by providing the user content root URL (e.g. https://app.box.com/folder/0) with the user's info, as in the example below.

FolderURL,UserName
https://app.box.com/folder/158718117193,Corp Admin
https://app.box.com/folder/158718322420,dev@acmecorp.com
https://app.box.com/folder/0,other_user@acmecorp.com

Exclusion Filter - Exclude only specified users/folders:

[{"ObjectType":"CSV","FilterType":"CSVExclude","Filter":"path\\to\\exfil.csv”}]

The exclusion filter only needs the URLs of the folders to exclude. No user information needs to be provided. The exclusion filter does not currently support excluding user root content folders.

FolderURL
https://app.box.com/folder/159258028738
https://app.box.com/folder/159259572184

18.2.12. Discovery Filter for Meridio

CPS allows to filter out documents and folders based on a variety of criteria like Container [ObjectType:0] & Items [ObjectType:1].

So, to discover folders and documents, users need to provide filter string as explained below:

Filter Type	Filter String
Container with DocID	[{"ObjectType":0,"Filter":"KContainer.CategoryID=<CategoryID>"}]
File Name with DocID	[{"ObjectType":1,"Filter":"D. DocID=<DocID>"}]
File Name by CategoryID	[{"ObjectType":1,"Filter":"D. CategoryID=<CategoryID>"}]
Container and File Name	[{"ObjectType":0,"Filter":"KContainer.CategoryID=<CategoryID>"}, {"ObjectType":1,"Filter":"D. CategoryID=<CategoryID>"}]

The <**CategoryID**> and <**DocID**> should be replaced with the Category ID and Document ID as mentioned in above query.

19. APPENDIX 2 - PROGRAMMING

Target metadata fields can be populated using the output of a programming function. The programming function can be a single built-in function, or from the output of a customer written PowerShell script.

Target Field	Parameter Type	Value
Author	• Function	• TOUPPER(SrcAuthor)

In the example above the Author field in the target is populated using a function which takes the value of the source metadata field SrcAuthor field and applies the TOUPPER function to it to set all characters to uppercase.

Built-in Functions

19.1. The full list of functions available to populate Target Metadata fields.

19.1.1. String Functions

Function Name	Parameters	Description
• ConcateWithSeparator	• Separator, stringOne, stringTwo	• This function allows you to combine two different strings with your separator.
Concate	stringOne, stringTwo	This function will join two strings.
StringContains	stringToCheck, orgString	This function will check one string contains other supplied string or not. It will return value as true or false.
ReplaceString	orgString, wantToReplace, replaceString	This function will replace sting.
SubstringByStartPosition	orgString, startPosition	This function will return sting from your supplied position.
SubstringByPostionAndLength	orgString, startPosition, length	This function will return sting from your supplied position and length.
ToUpper	Name	Changing text case to upper
ToLower	Name	Changing text case to lower

Compare	stringOne, stringTwo	Comparing two string values

19.1.2. Math Functions

Function Name	Parameters	Description
• SUM	• stringOne, stringTwo	• This function will return sum of given two numbers.
• Round	• Value, RoundUpto	• Returning a decimal value to the nearest integral value
• Absolute	• Value	• Returning the absolute value of a Decimal number
• Power	• Value, PowerUpto	• Returning a specified number raised to the specified power

19.1.3. Date Functions

Function Name	Parameters	Description
• Now()	•	• This function will return current datetime value.
• Get-Day	• Date	• This function will return day value from given date.
• Get-Month	• Date	• This function will return month value from given date.
• Get-Year	• Date	• This function will return year value from given date.
• Get-Time	• Date	• This function will return time value from given date.
• Get-DateFormat[yyyy-MM-dd]	• Date	• This function will return given date in yyyy-MM-dd format.
• Get-DateFormat[dd-MM-yyyy]	• Date	• This function will return given date in dd-MM-yyyy format.

19.1.4. Misc Functions

Function Name	Parameters	Description
• GetMappedUser	• srcMetaValue	• This function will give target mapped for given source user based on user mapping.

• Substitute_Invalid_Characters	• inputString	• This function will substitute invalid characters in given input string that are not supported by target system.
• GetTargetTaxonomy	• sourceTaxonomy	• This function will give target taxonomy value for given source taxonomy value.
• Get-Newname	• Name, Ext	• This function will return a new name after combining name and extension of file.
• Get-PathSegment	• Url, SegmentNumber	• This function will return path segment from given Url.
• GetTargetContentItemIdBySourceUri	• srcMetaValue	• This function will return target content item id based on source item Uri.

19.1.5. OCR Functions

Function Name	Parameters	Description
• Get-OCRText	• srcMetaValue	• This function will read text field from Image file by rendering text pattern.
• Get-OCRDateField	• srcMetaValue	• This function read date field from Image by rendering through date pattern.
• Get-OCRWordField	• srcMetaValue	• This function will return specific word from string by rendering image
• Get-OCRSpaceDelimitedField	• srcMetaValue	• This function will return field which is space separated with field value by rendering image
• Get-OCRTerminatedField	• srcMetaValue	•
• Get-OCRNumberField	• srcMetaValue	•

• Get-OCRFourColumnNumber	• srcMetaValue	• This function will return four column values separated '/'.
• Get-OCRMonthYear	• srcMetaValue	• This function will return Month year value from Date field.
• Get-OCRYear	• srcMetaValue	• This function will return Year value from Date field.

19.2. Creating Lookup tables for Custom Functions

Often target metadata may need to be supplemented with additional data from a CSV file. For example, a table of information for each employee where a lookup using the EmployeeID from the source item is then used to return the user's department, line manager, grade level etc. from a .CSV file.

In such cases, a Target metadata property can be mapped to a PowerShell script which can lookup in a CSV file for relevant information. If the lookup file(.CSV) were to occur every time an item is processed, this approach would severely impact migration performance, especially when using large lookup tables.

To resolve such issues, CPS has the ability for a .CSV file to be loaded only once per migration thread instead of for each item.

Note: This feature is only supported on non SaaS installations of CPS.

1. Design CSV file

The CSV file must contain a unique Item ID column (effectively a column that acts as a key), and columns for all the Metadata associated with the item.

```
EmployeeID,     FirstName,   Surname,   DOB
0001,           Jane,        Potter,    16/2/1962
0002,           Sarah,       Mahoney,   28/4/1995
0003,           Rosemary,    Sharkey,   3/3/1975
```

Example EmployeeDatabase.CSV

EmployeeID is the key and there are three columns FirstName, Surname and DOB

2. Creating new PowerShell module

A new PowerShell module file (.PSM1 file) needs to be created which will be executed once by each thread basis to load the data.

e.g. Datatable_Load.psm1

3. **Use PowerShell to import CSV.**

Edit the new PowerShell script Datatable_Load.psm1 to create a script to load the .CSV file.

Syntax to create a global variable and import the CSV

```
$global:<VariableToHoldCSV> = Import-CSV  -path '<My csv path>'
```

e.g.

```
$global:Employee_Metadata = Import-CSV  -path 'c:\datamigration\EmployeeDatabase.CSV'
```

When the script is executed the Employee_Metadata global variable will contain the contents of the CSV file.

4. **Add reference to new PowerShell module**

CPS needs to be configured to load this new PowerShell module.

Open Notepad in Administrator mode and open the main Proventeq PowerShell Module

Proventeq.MigrationAccelerator.PowerShell.Module.psd1 (typically located in C:\Program Files\Proventeq\Proventeq Migration Accelerator\PowerShell)

Update the **NestedModule** property

of **Proventeq.MigrationAccelerator.PowerShell.Module.psd1** file to include this new module (e.g., Datatable_Load.psm1). Typically

```
NestedModules =
@("Proventeq.MigrationAccelerator.SharePointModule.psm1","Proventeq.MigrationAc
celerator.PowerShellModule.psm1","c:\datamigration\Datatable_Load.psm1")
```

After this change, starting CPS will execute the PowerShell script.

When the threads start, they will execute the module Datatable_Load.psm1 which will execute the above PowerShell to load the CSV into a global variable DOB_Metadata for the Custom Function to use.

19.2.1. Using the Imported Data

The PowerShell script executed by the Custom Function can now use the data as follows.

```
$EmployeeDatabase = $global:Employee_Metadata | ?{$_.EmployeeID -eq  $EmployeeID}
Return $EmployeeDatabase.DOB
```

In the example above, a variable $EmployeeID is used to find a matching entry by looking up the .EmployeeID field in the Employee_Metadata global variable. The script then returns the .DOB field for that record.

Testing PowerShell Functions

It is recommended to test any changes made to the PowerShell module prior to launching CPS. The following steps can be executed to validate the changes.

Open PowerShell console and import the Proventeq PowerShell module

```
"C:\Program Files\Proventeq\Proventeq Content Suite
\PowerShell\Proventeq.MigrationAccelerator.Module.psd1" -Force
```

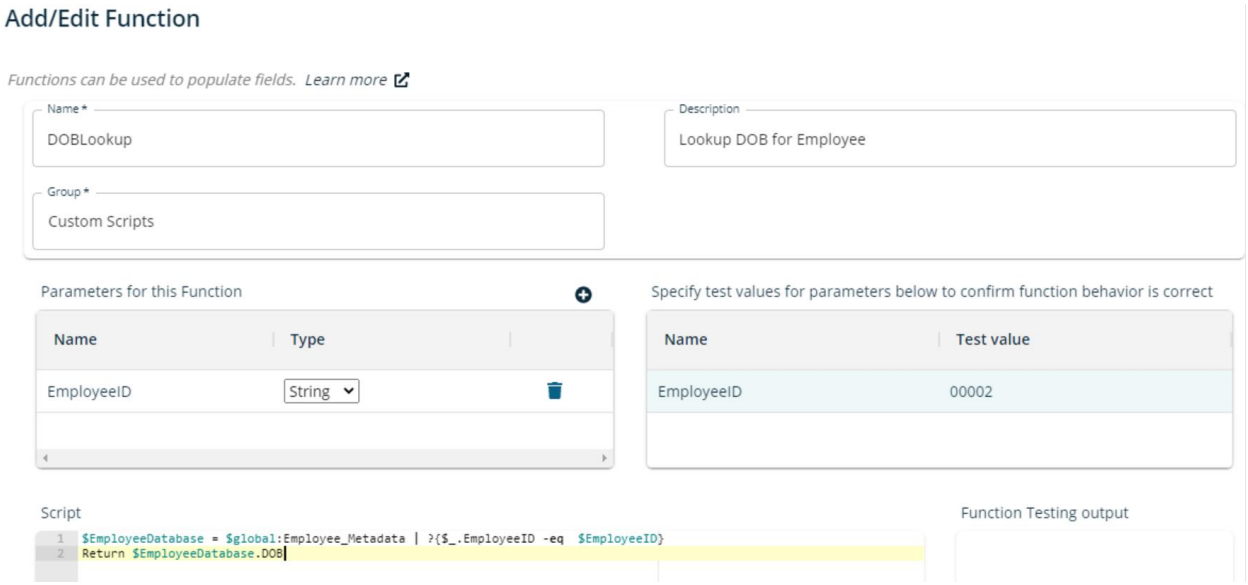


FIGURE 131 - ANALYTICS – EXAMPLE FUNCTION USING IMPORTED DATA

In the example above you can see the EmployeeID variable defined as a string and below that the Script to execute. On the right is Test Data window where the value 0002 is passed to the function and it correctly returns in the Text View the DOB of 28/4/1995 based upon the values in the CSV EmployeeDatabase.CSV which is held in Employee_Metadata global variable.

Controlling how parameters are passed to the Custom Function

In the properties of the function, specify how the parameter value will be passed to the customer function. In the example below, the EmployeeID value comes from the Name metadata field.

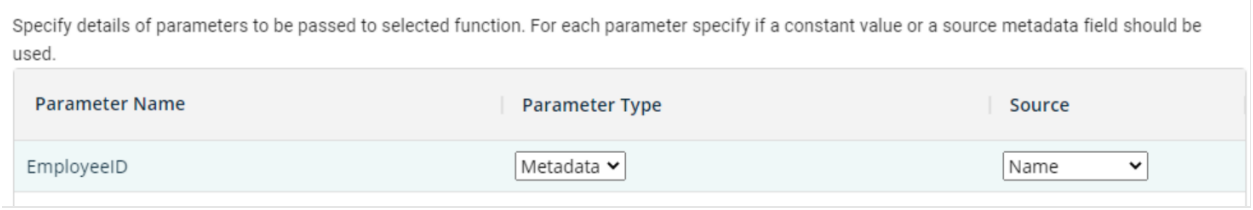


FIGURE 132 - CONFIGURE HOW DATA IS PASSED IN AS PARAMETER TO FUNCTION

19.2.2. BOX API Consumption

Below are details of how CPS consumes BOX API resources. This can give an indication of how much BOX API licence consumption may occur during a migration. It's essential to discuss BOX API limits with your BOX Account Manager.

- **Discovery**
  1 API call per folder to retrieve a list of subfolders and files.
  1 API call per file to discover the sharing permissions (if security discovery is enabled and the file has been shared with other users).
  1 API call per file to discover the version history (if version discovery is enabled and the file has previous versions).
  A small number of calls per discovery run for setup (get user ID mappings, discover the initial folder / discovery filter entries, etc).
- **Migration**
  1 API call per file version to download the file to the migration server.
- **Misc**
  A small number of calls when adding Box as a source connection to populate the folder selector.

> Note: Some of these API methods are paginated at 1000 items, so it takes 1 call per 1000 entries. For example, if a folder contains 2500 files, it will take 3 API calls to get the full list of items.