

SharePoint Deduplication

The SharePoint Deduplication module helps administrators identify and remove duplicate files from SharePoint Online document libraries. It uses hash-based scanning to accurately detect duplicate content and ensures unique files are retained in the destination environment.

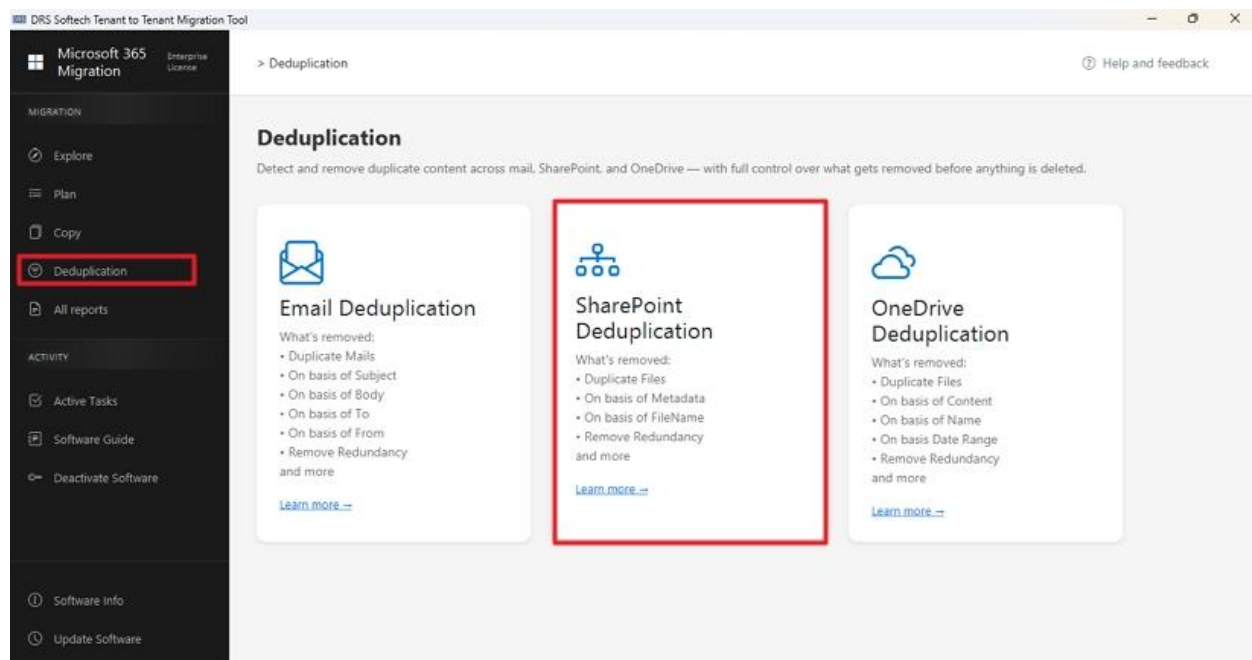
Note:

- The Demo Version allows only a preview of duplicate files. Duplicate removal is available in the licensed version.
- SharePoint deduplication removes files only. It does not affect lists, list items, permissions, workflows, or site configurations.

Steps for SharePoint Deduplication

Step 01: Select SharePoint Deduplication

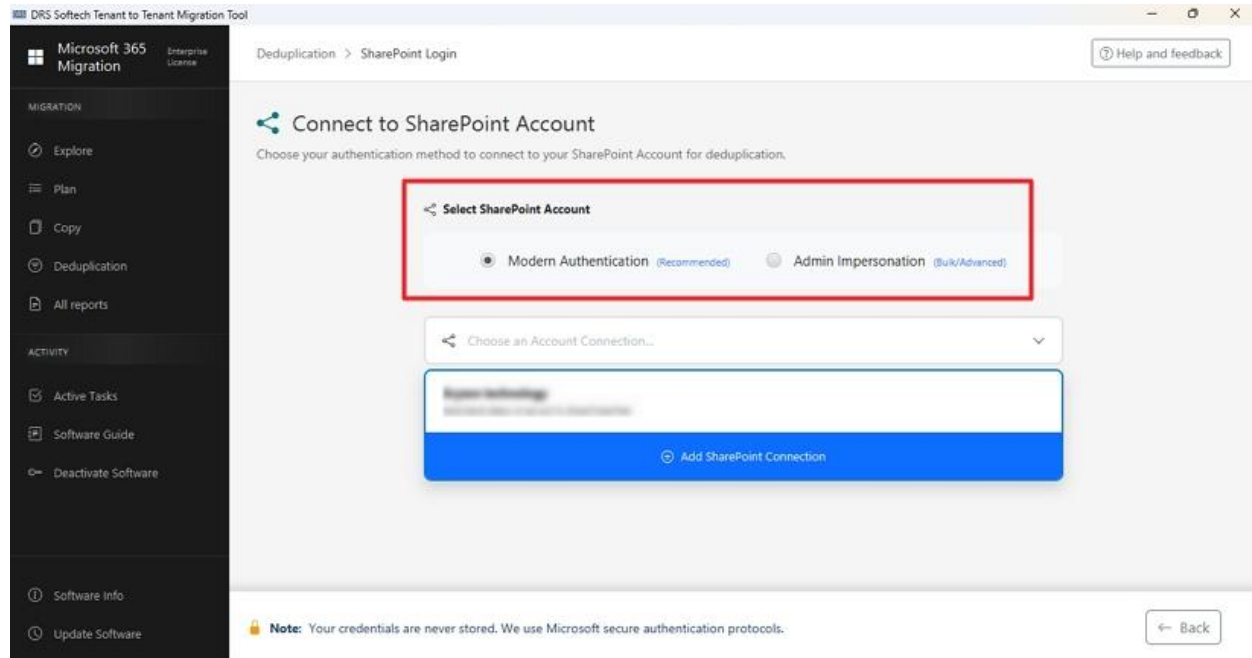
From the Deduplication module, click **SharePoint Deduplication** to begin the process.



Step 02: Choose Source Authentication Mode

Select the authentication method to connect to the source Microsoft 365 tenant:

- **Modern Authentication:** Sign in securely using your Microsoft 365 Administrator account.
- **Admin Impersonation:** Connect using Azure App Registration credentials, including Client ID, Client Secret Key, and Tenant ID.

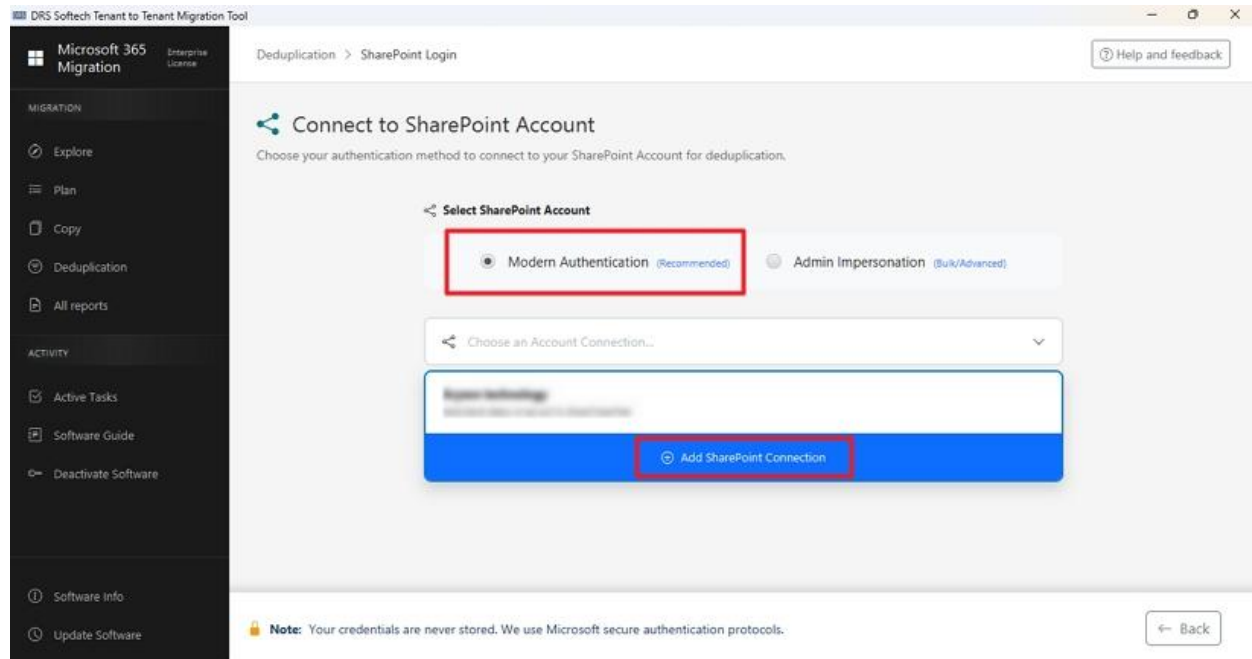


Step 03: Sign In Using Modern Authentication

If you choose Modern Authentication, click **Choose an Account Connection**. Select an existing connection (if available) or click **Add Connection** to create a new one. You will be redirected to the Microsoft sign-in page.

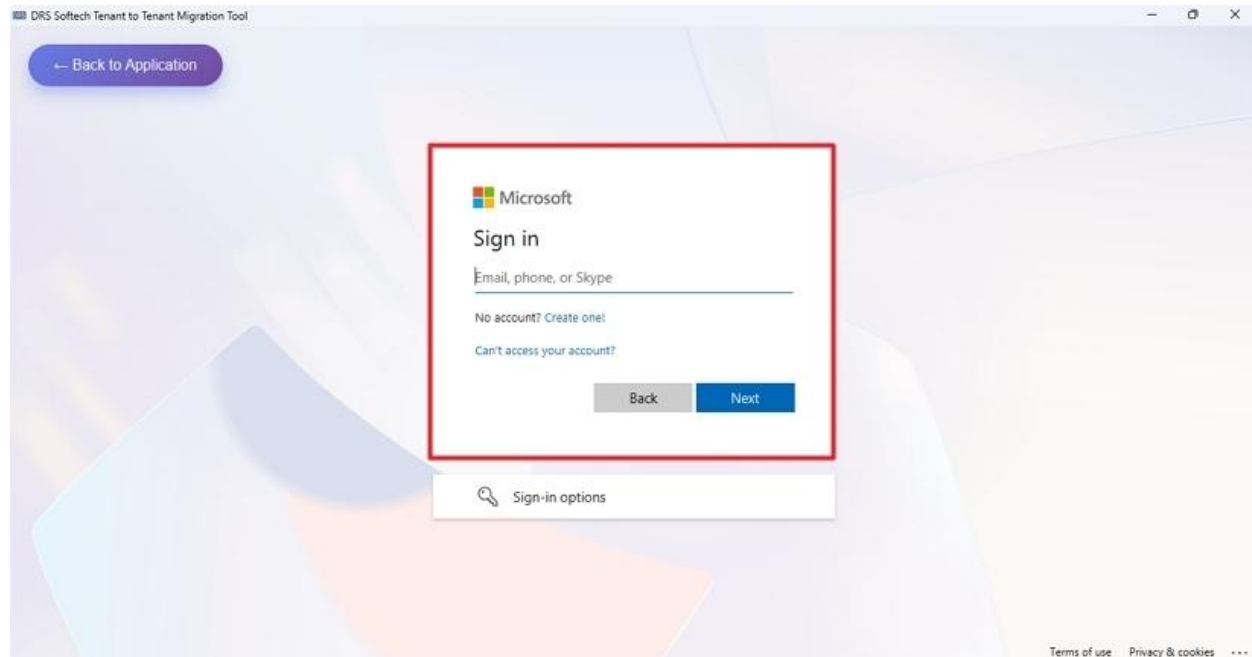
Note:

- The software securely saves your configured login connections. For future migrations, simply select a saved connection instead of entering the credentials again.
- If Modern Authentication fails, you can manually grant the required Microsoft permissions by following the Microsoft Entra configuration guide. [Modern Authentication Fails](#)



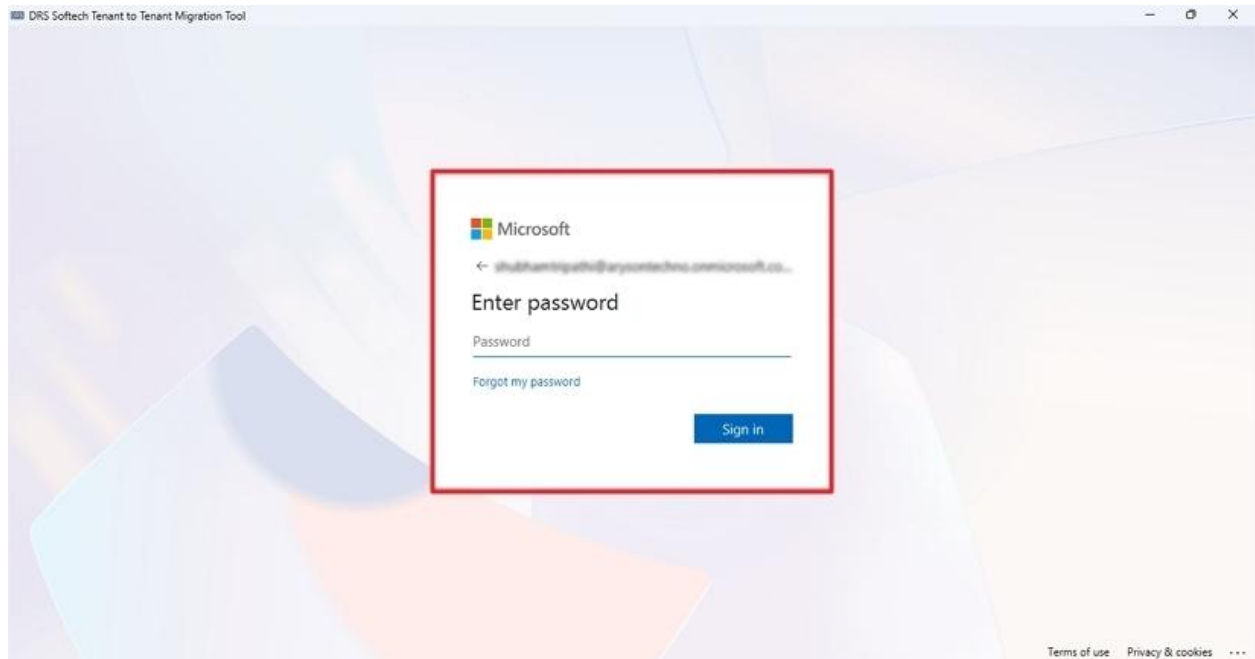
Step 04: Enter Your Microsoft 365 Administrator Account

On the Microsoft sign-in page, enter or select your **Microsoft 365 Administrator email address** and click **Next**.



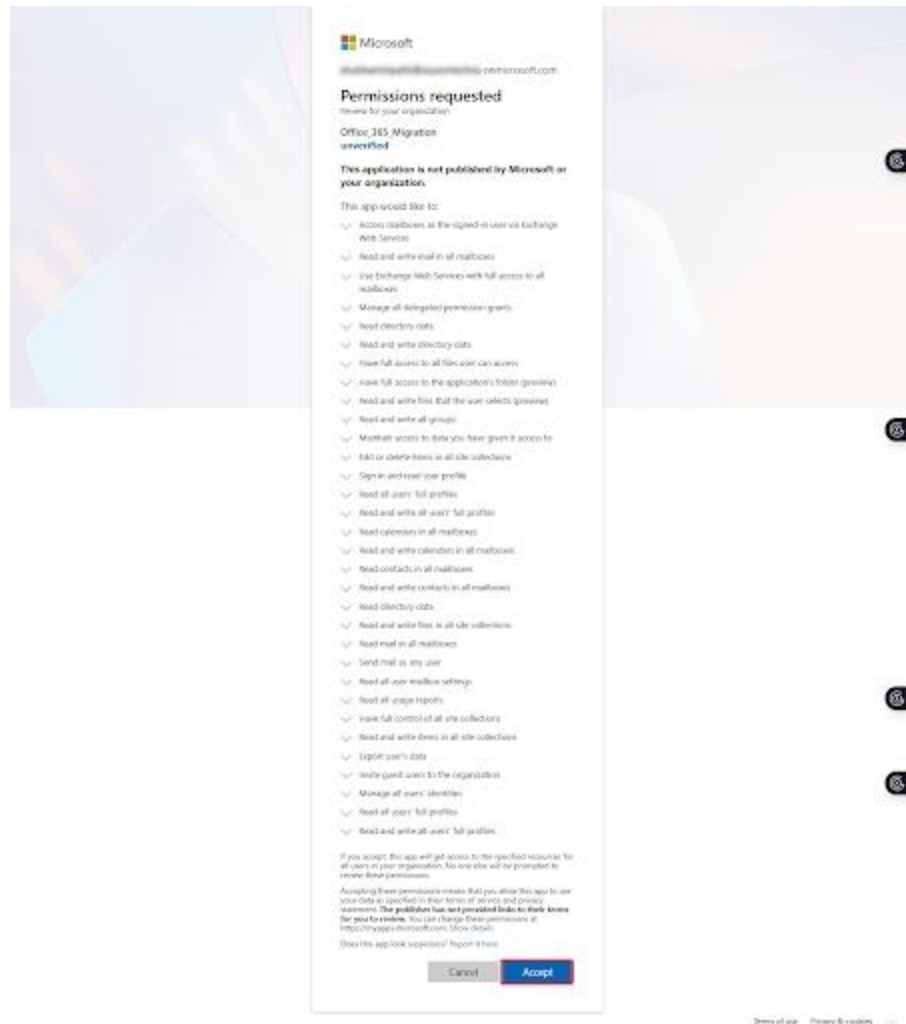
Step 05: Authenticate Your Account

Enter your administrator account **password** and click **Sign In**. If Multi-Factor Authentication (MFA) is enabled, complete the verification process to continue.



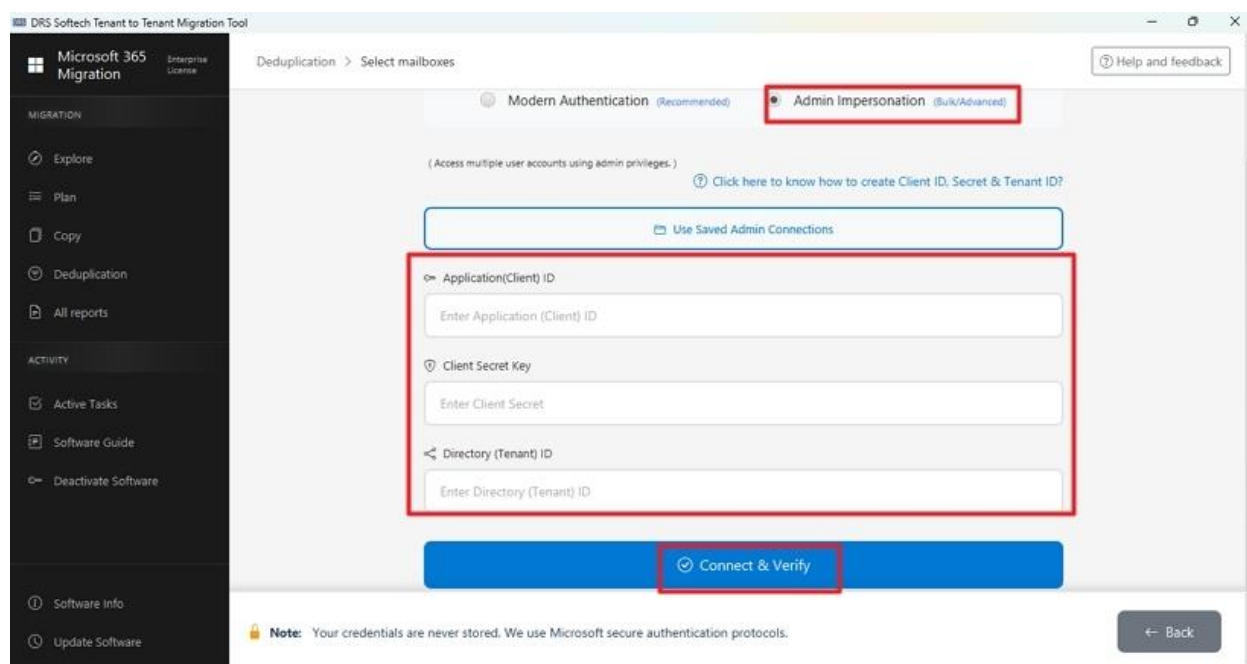
Step 06: Grant Microsoft Permissions

After successful authentication, Microsoft displays a permission consent window. Review the requested permissions and click **Accept** to authorize the application and continue.



Step 07: Sign In Using Admin Impersonation

If you choose Admin Impersonation, enter the following Azure App Registration credentials: **Client ID**, **Client Secret Key**, and **Tenant ID**. After entering the required information, click **Connect & Verify**.



Note: If you have not yet created an Azure App Registration or generated the required credentials, follow this guide:

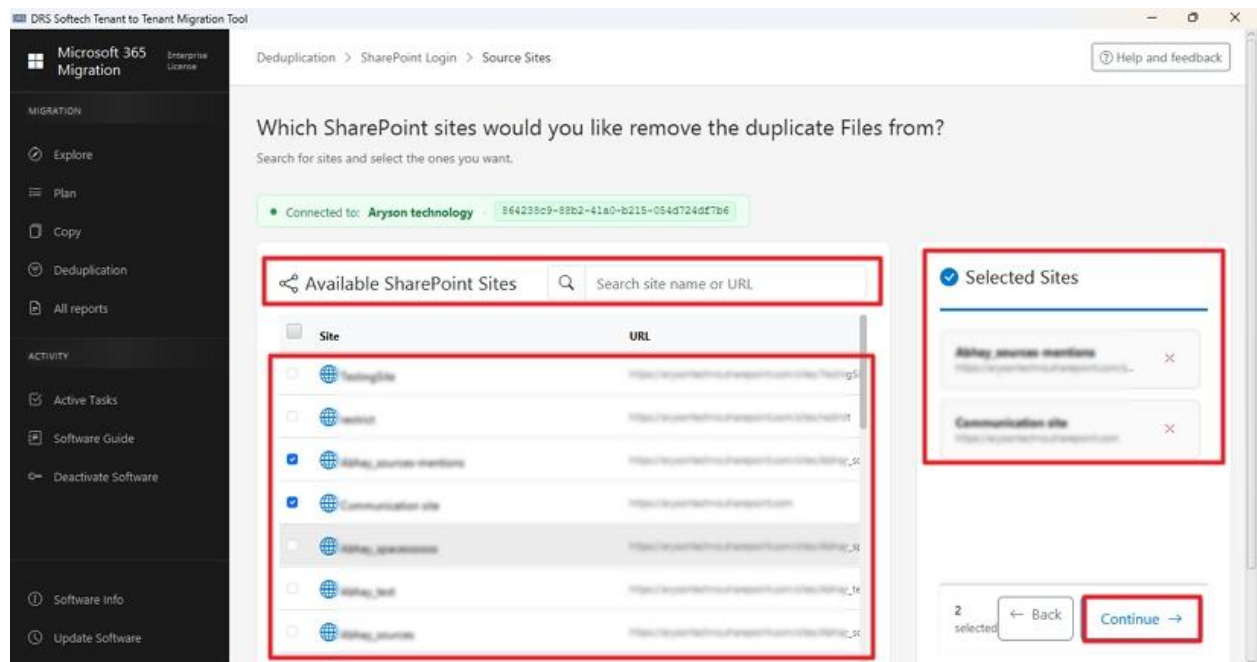
[https://www.data-recovery-solutions.com/how-to/create-azure-active-direct
ory-and-enable-api-permission/](https://www.data-recovery-solutions.com/how-to/create-azure-active-directory-and-enable-api-permission/)

Once the Azure App Registration is configured correctly, the required Microsoft Graph API permissions are automatically assigned. Therefore, no additional consent is required during the connection process. The screenshot below shows the Microsoft 365 API permissions required during Azure App Registration.

API / Permissions name	Type	Description	Admin consent requ...	Status
▼ Microsoft Graph (30)				
Calendars.Read	Application	Read calendars in all mailboxes	Yes	✔ Granted for BUSINESS ...
Calendars.ReadWrite	Application	Read and write calendars in all mailboxes	Yes	✔ Granted for BUSINESS ...
Channel.ReadBasic.All	Application	Read the names and descriptions of all channels	Yes	✔ Granted for BUSINESS ...
ChannelMessage.Read.All	Application	Read all channel messages	Yes	✔ Granted for BUSINESS ...
Contacts.Read	Application	Read contacts in all mailboxes	Yes	✔ Granted for BUSINESS ...
Contacts.ReadWrite	Application	Read and write contacts in all mailboxes	Yes	✔ Granted for BUSINESS ...
Directory.Read.All	Application	Read directory data	Yes	✔ Granted for BUSINESS ...
Files.ReadWrite.All	Application	Read and write files in all site collections	Yes	✔ Granted for BUSINESS ...
Group.Read.All	Application	Read all groups	Yes	✔ Granted for BUSINESS ...
Group.ReadWrite.All	Application	Read and write all groups	Yes	✔ Granted for BUSINESS ...
GroupMember.ReadWrite.All	Application	Read and write all group memberships	Yes	✔ Granted for BUSINESS ...
Mail.Read	Application	Read mail in all mailboxes	Yes	✔ Granted for BUSINESS ...
Mail.ReadWrite	Application	Read and write mail in all mailboxes	Yes	✔ Granted for BUSINESS ...
Mail.Send	Application	Send mail as any user	Yes	✔ Granted for BUSINESS ...
MailboxSettings.Read	Application	Read all user mailbox settings	Yes	✔ Granted for BUSINESS ...
Reports.Read.All	Application	Read all usage reports	Yes	✔ Granted for BUSINESS ...
Sites.FullControl.All	Application	Have full control of all site collections	Yes	✔ Granted for BUSINESS ...
Sites.ReadWrite.All	Application	Read and write items in all site collections	Yes	✔ Granted for BUSINESS ...
Team.Create	Application	Create teams	Yes	✔ Granted for BUSINESS ...
Team.ReadBasic.All	Application	Get a list of all teams	Yes	✔ Granted for BUSINESS ...
TeamMember.Read.All	Application	Read the members of all teams	Yes	✔ Granted for BUSINESS ...
TeamMember.ReadWrite.All	Application	Add and remove members from all teams	Yes	✔ Granted for BUSINESS ...
TeamsActivity.Read.All	Application	Read all teams' settings	Yes	✔ Granted for BUSINESS ...
TeamworkTag.CreateAll	Application	Create chat and channel messages with anyone's identity	Yes	✔ Granted for BUSINESS ...
TeamworkTag.ReadWrite.All	Application	Read and write tags in Teams	Yes	✔ Granted for BUSINESS ...
User.Export.All	Application	Export user's data	Yes	✔ Granted for BUSINESS ...
User.Invite.All	Application	Invite guest users to the organization	Yes	✔ Granted for BUSINESS ...
User.ManageIdentities.All	Application	Manage all users' identities	Yes	✔ Granted for BUSINESS ...
User.Read.All	Application	Read all users' full profiles	Yes	✔ Granted for BUSINESS ...
User.ReadWrite.All	Application	Read and write all users' full profiles	Yes	✔ Granted for BUSINESS ...

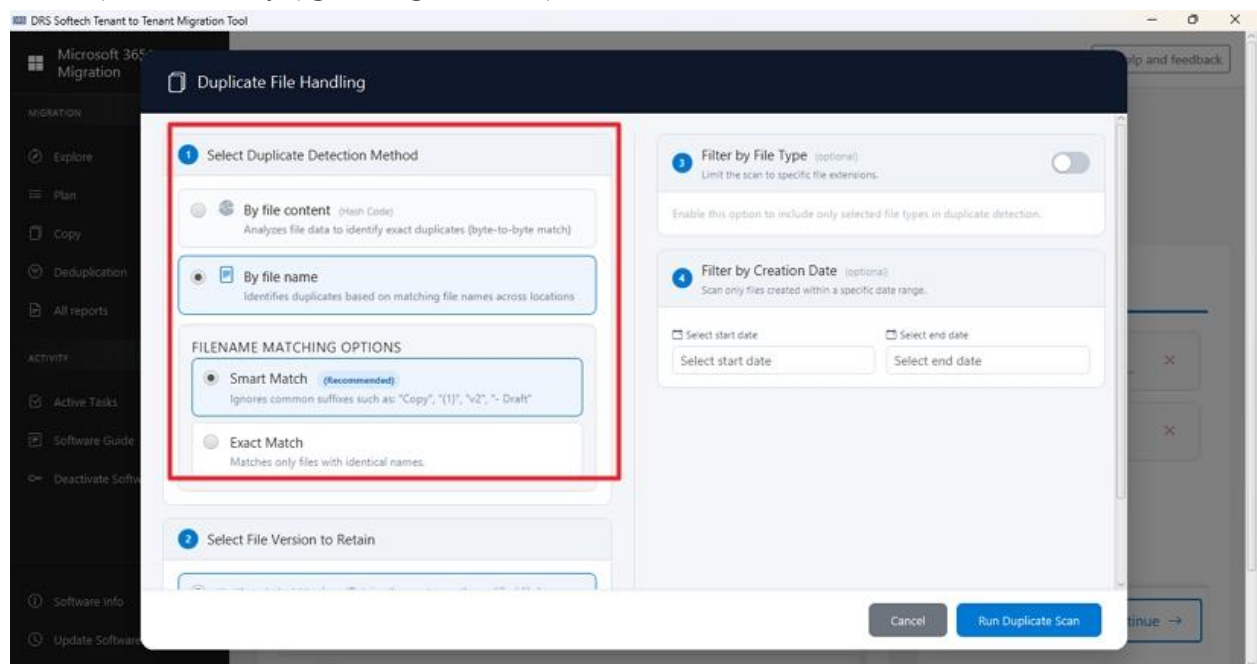
Step 08: Select SharePoint Sites

Search SharePoint sites by Site Name or URL, then select the required sites for deduplication. Verify the selected sites in the Selected Sites panel and click **Continue**.



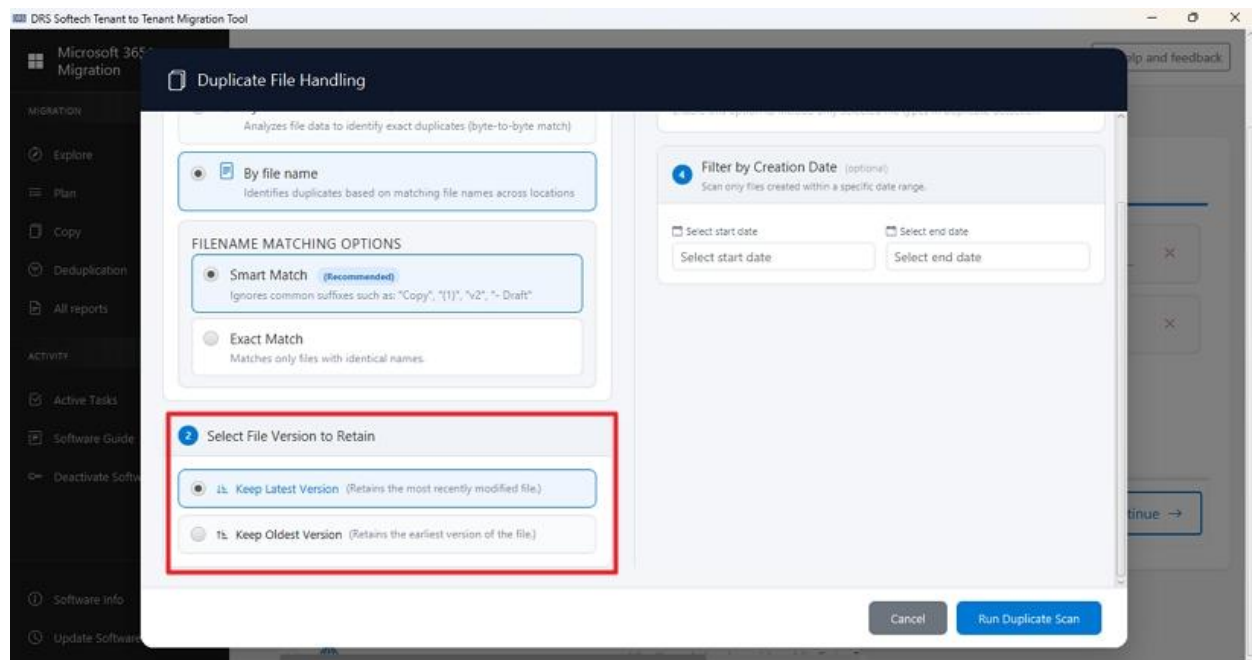
Step 09: Duplicate Detection

Choose how you want to detect duplicate files—either by **File Content** or by **File Name**. If you select **File Name**, choose whether to **match files Exactly** (identical names) or **Smartly** (ignoring suffixes).



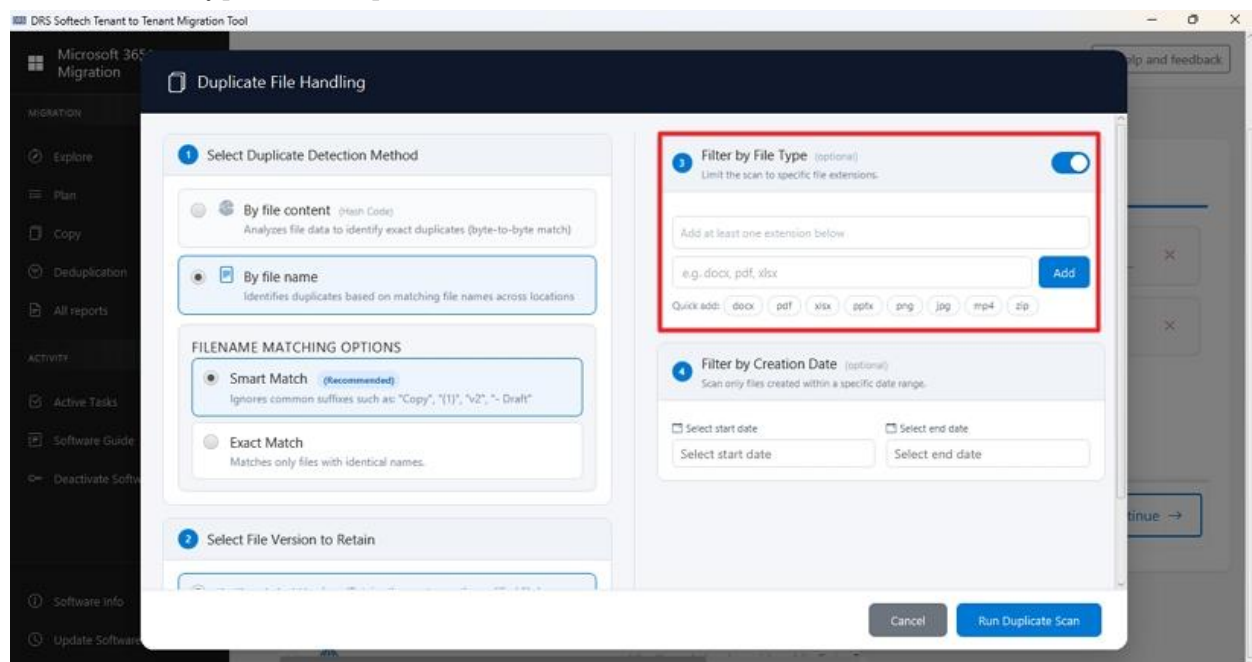
Step 10: Choose the File Version to Retain

After selecting the duplicate detection method, choose which file version to keep. You can retain either the **Latest Modified Version** or the **Oldest Version**.



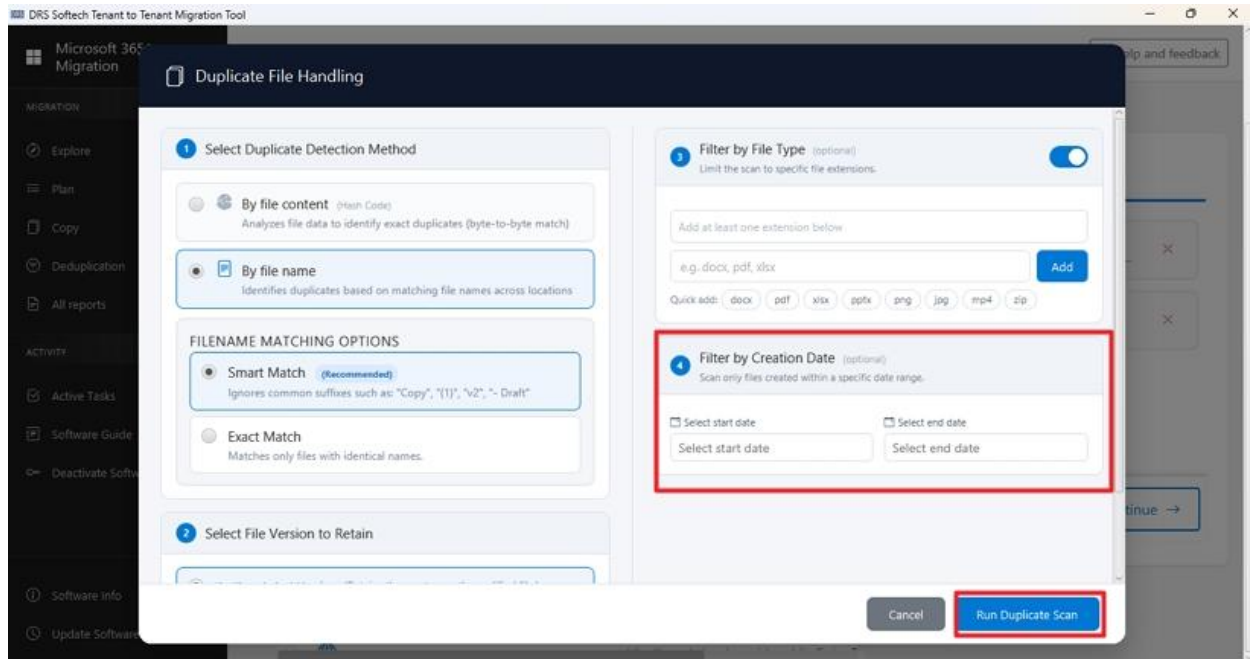
Step 11: Filter Duplicates by File Type

Select the file extensions using the **Quick Add** option or manually enter extensions (such as .csv, .docx, .pst, .bmp, etc.) and click **Add**. The tool will scan only the selected file types for duplicates.



Step 12: Filter Duplicates by Date

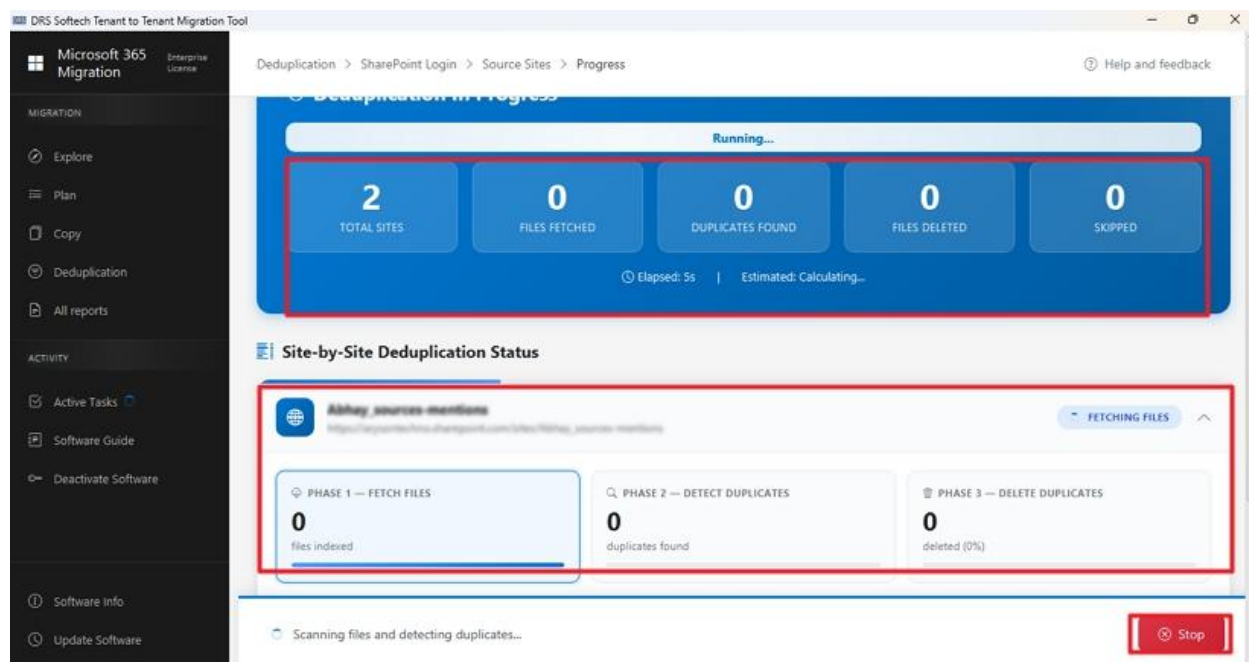
Use the **Creation Date** filter and specify the **Start Date** and **End Date** to scan duplicate files created within a particular time period. Then, click **Run Duplicate Scan** to begin the process.



Step 13: Deduplication Progress

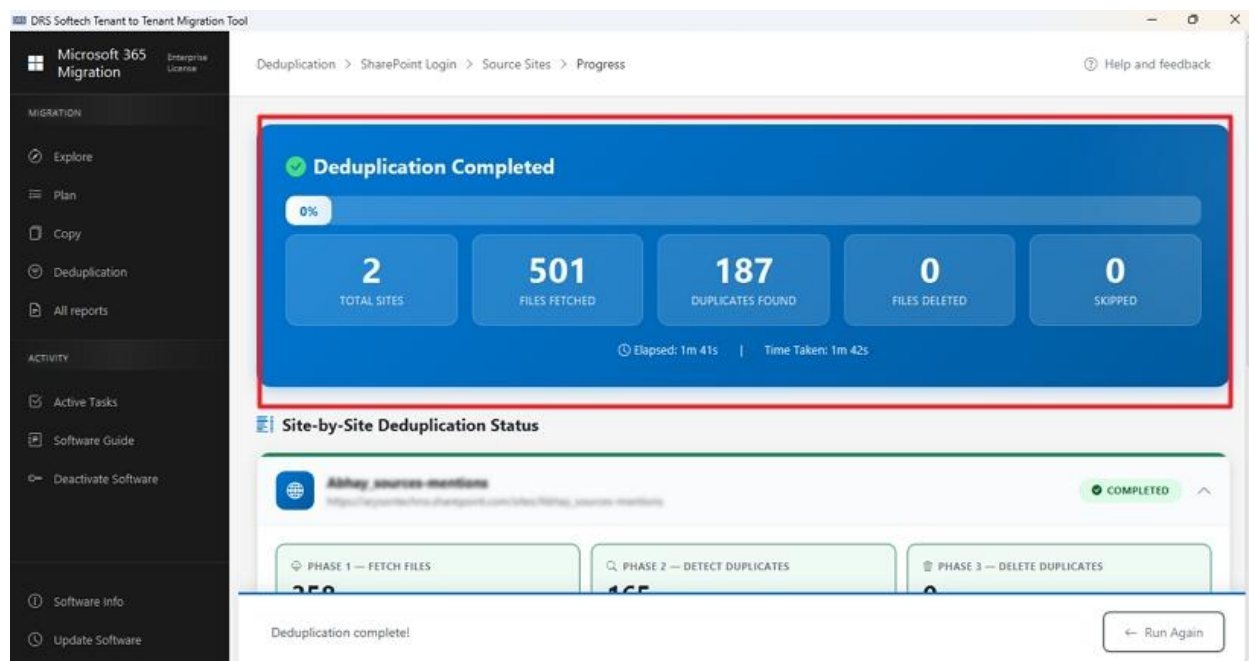
The progress window displays the real-time status of the SharePoint deduplication process. It shows how many sites and files are being fetched, along with the progress of duplicate detection, skipped files, and deletion. **The process is divided into three phases:**

- **Fetch Files:** Scans the selected SharePoint sites and indexes all files available in the document libraries.
- **Detect Duplicates:** Analyzes the fetched files and identifies duplicate items based on the selected criteria.
- **Delete Duplicates:** Removes the detected duplicate files according to the selected retention rules.
- You can also monitor real-time updates during the scan and use the **Stop** button to cancel the process if required.



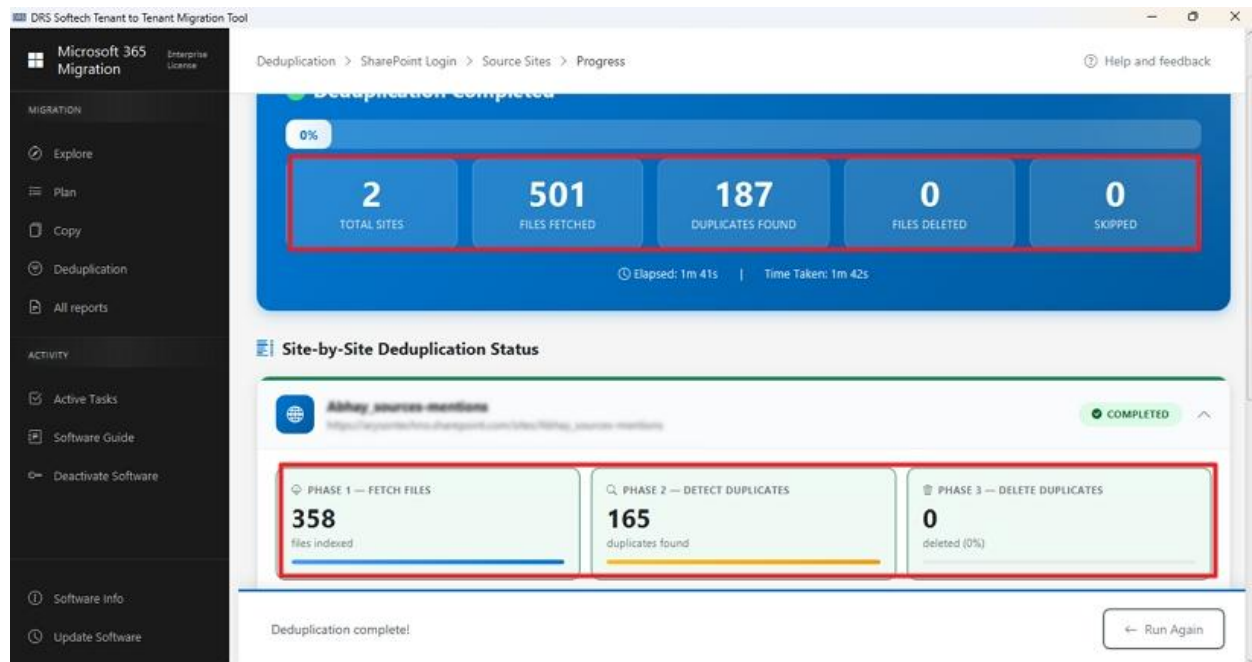
Step 14: After Scanning Options

Once the duplicate scan is complete, choose **Skip Deletion** (Phase 3) to keep the detected duplicates, or select Proceed to **Delete** to remove the duplicate files from the selected SharePoint sites.



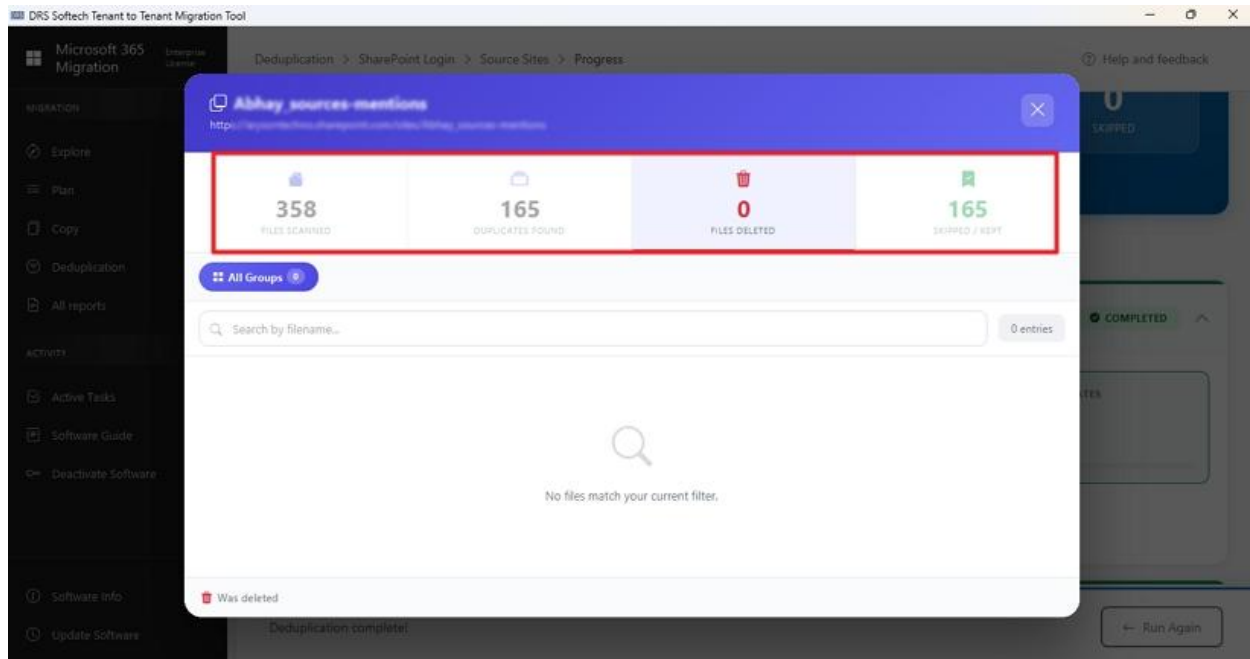
Step 15: View Duplicate Records

To review the detected duplicate files before deletion, check the site-wise deduplication status and click **View Duplicate Files** for the selected SharePoint site. A pop-up window displays details such as **Total Files Scanned**, **Duplicate Groups**, **Files to Delete**, and **Retained Files**. You can also view duplicate records by Groups, Images, Documents, or Others.



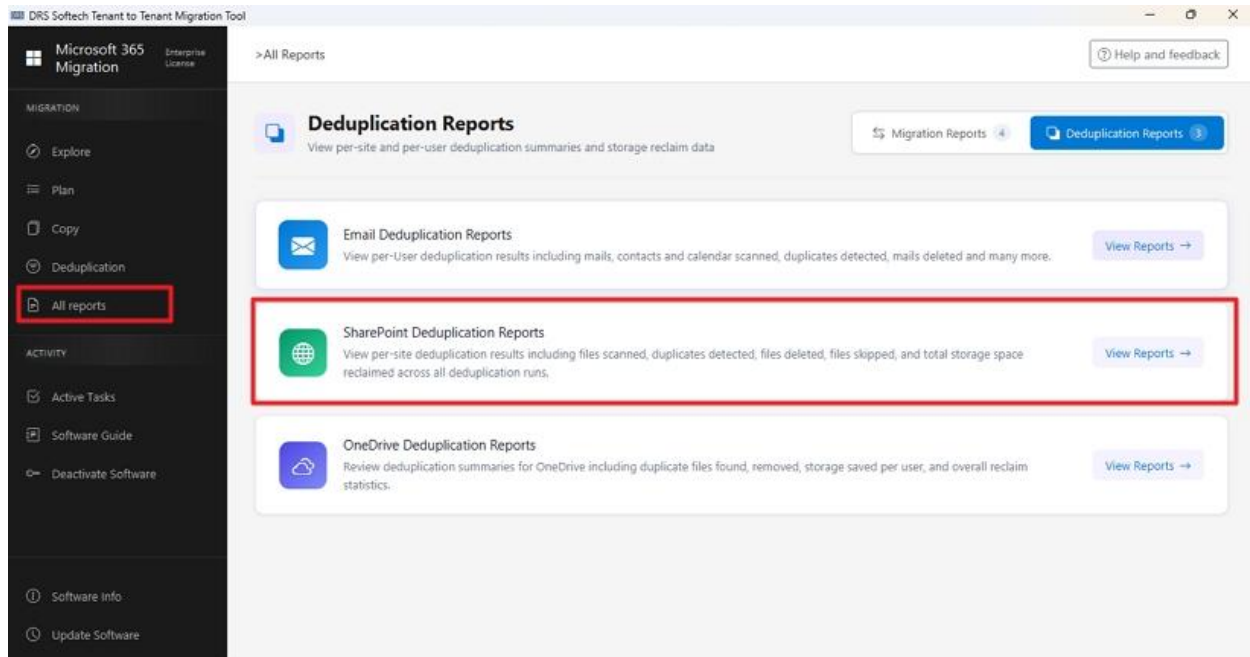
Step 16: Deduplication Completion

After the process finishes, the SharePoint Deduplication Completed status appears. Click **View Deletion Summary** to review the detailed deletion results.



Step 17: Deduplication Report

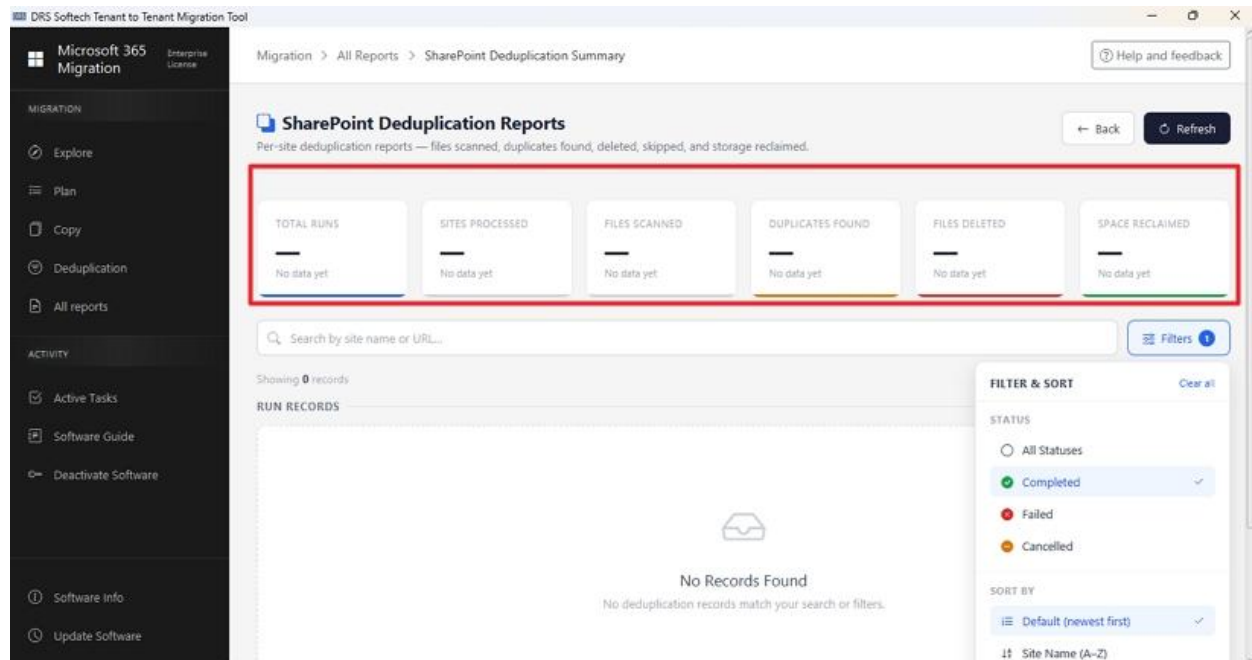
Go to the **All Reports** tab and select Deduplication Reports. Under the SharePoint Deduplication Report, click **View Reports** to see the report details.



Step 18: Verify Reports

The SharePoint Deduplication Reports section provides a complete summary of the deduplication process. It helps administrators verify duplicate removal, review the

results, and check the reclaimed storage space. You can also click the Delete icon to remove a report.



Step 19: Search Reports

If multiple SharePoint Deduplication Reports are available, use the Search pane to find a specific report by file name or URL. You can also apply the available filters to locate the required report quickly and verify the deduplication results.

