- **SharePoint site URL** and **library path**
- **Client ID and Client Secret** from Azure AD (for app registration)
- **Tenant ID**
- Proper **permissions** on SharePoint (usually delegated or application permissions for reading files)

## Register an App in Azure AD (to get Client ID, Client Secret, Tenant ID)

**Step 1: Go to Azure Portal**

- Visit https://portal.azure.com
- Search for **"App registrations"**
- Click **"New registration"**

**Step 2: Register the App**

- **Name:** Something like DatabricksSharePointAccess
- **Supported account types:** "Accounts in this organizational directory only"
- **Redirect URI:** Leave it blank or put http://localhost (not used for this task)
  Click **Register**

**Step 3: Copy These Values**

After registration, we'll get:

- **Application (client) ID**
- **Directory (tenant) ID**

**Step 4: Create a Client Secret**

- Go to **Certificates & secrets** → **New client secret**
- Name it something like databricks-secret
- Set an expiry (e.g. 6 or 12 months)
- Copy the **value** immediately (this is your **Client Secret**)

⚠️ Note: You won't be able to see the secret again later

## Give Permissions to Access SharePoint

**Step 1: Go to API Permissions tab in the app**

- Click **+ Add a permission**

- Choose **Microsoft Graph**
- Choose **Application permissions**

Now add the following permissions:

- Sites.Read.All (to read all site contents)
- Files.Read.All (if you want to read all user files)

🛑 **Important:** These are **application** permissions (not delegated), so you must **grant admin consent**.

**Step 2: Click Grant admin consent for [Tenant Name]**

That's it. Now your app has permission to access SharePoint content.

✅ **Key Points:**

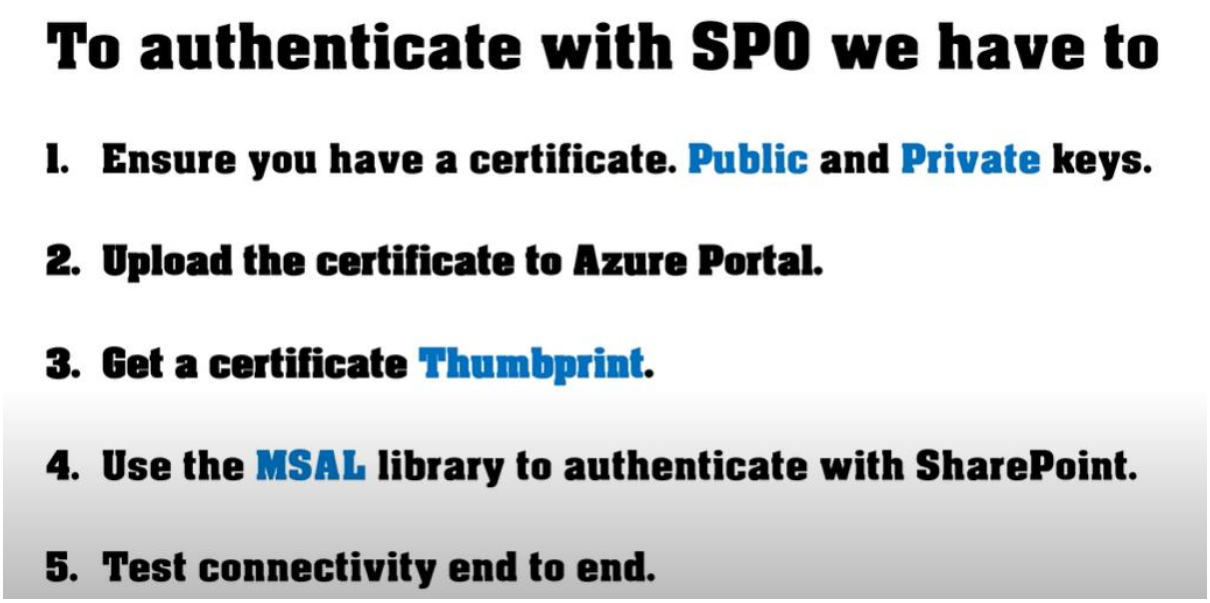
- **Microsoft Graph API** requires **Azure AD** authentication (Client ID, Client Secret, and Tenant ID).
- You still need an **Azure Active Directory** account or service principal **(App Registration)** to authenticate and interact with **SharePoint or OneDrive**.

**3. Do I Need Azure AD for All Microsoft 365 APIs?**

Yes, for most API endpoints of **Microsoft 365** (like **OneDrive**, **SharePoint**, **Outlook**, etc.), you will need **Azure AD** for authentication and authorization.

**Short Summary:**

- You **do** need Azure AD credentials (Client ID, Secret, Tenant ID) to access **SharePoint** or **OneDrive** via **Microsoft Graph API**.
- Even though you're using Databricks (standalone or not), **authentication** to Microsoft services still happens through **Azure AD**.




Another method:

**What's involved:**

1. **Extract**: Read metadata + SharePoint file URLs (likely from Excel or SharePoint List).
2. **Download**: Get files from SharePoint.
3. **Transform**: Format metadata to match Veeva Vault / Veeva CRM document object structure.
4. **Load**: Upload documents to Veeva Vault using its **API (REST or SOAP)** with proper metadata.

Workflow:

**Step 1: Read Metadata from Excel**

**Step 2: Download Files from SharePoint**

⚠️ If your SharePoint requires authentication, you'll need to use Microsoft Graph API or Office365-REST-Python-Client with OAuth tokens.

**Step 3: Format Data for Veeva**

You'll need to map your SharePoint metadata to Veeva's **document object structure**. This might include:

- document_name

- product__c
- document_type__v
- version__v
- etc.