

# Microsoft OneDrive

**Microsoft OneDrive** (formerly `SkyDrive`) is a file hosting service operated by Microsoft.

This notebook covers how to load documents from `OneDrive`. Currently, only docx, doc, and pdf files are supported.

## Prerequisites

---

1. Register an application with the [Microsoft identity platform](#) instructions.
2. When registration finishes, the Azure portal displays the app registration's Overview pane. You see the Application (client) ID. Also called the `client_ID`, this value uniquely identifies your application in the Microsoft identity platform.
3. During the steps you will be following at **item 1**, you can set the redirect URI as `http://localhost:8000/callback`
4. During the steps you will be following at **item 1**, generate a new password (`client_secret`) under Application Secrets section.
5. Follow the instructions at this [document](#) to add the following `SCOPES` (`offline_access` and `Files.Read.All`) to your application.
6. Visit the [Graph Explorer Playground](#) to obtain your `OneDrive ID`. The first step is to ensure you are logged in with the account associated your OneDrive account. Then you need to make a request to `https://graph.microsoft.com/v1.0/me/drive` and the response will return a payload with a field `id` that holds the ID of your OneDrive account.
7. You need to install the o365 package using the command `pip install o365`.
8. At the end of the steps you must have the following values:

- `CLIENT_ID`

- `CLIENT_SECRET`
- `DRIVE_ID`

# Instructions for ingesting your documents from OneDrive

## Authentication

By default, the `OneDriveLoader` expects that the values of `CLIENT_ID` and `CLIENT_SECRET` must be stored as environment variables named `O365_CLIENT_ID` and `O365_CLIENT_SECRET` respectively. You could pass those environment variables through a `.env` file at the root of your application or using the following command in your script.

```
os.environ['O365_CLIENT_ID'] = "YOUR CLIENT ID"
os.environ['O365_CLIENT_SECRET'] = "YOUR CLIENT SECRET"
```

This loader uses an authentication called *on behalf of a user*. It is a 2 step authentication with user consent. When you instantiate the loader, it will call will print a url that the user must visit to give consent to the app on the required permissions. The user must then visit this url and give consent to the application. Then the user must copy the resulting page url and paste it back on the console. The method will then return True if the login attempt was succesful.

```
from langchain.document_loaders.onedrive import OneDriveLoader

loader = OneDriveLoader(drive_id="YOUR DRIVE ID")
```

Once the authentication has been done, the loader will store a token (`o365_token.txt`) at `~/.credentials/` folder. This token could be used later to authenticate without the copy/paste steps explained earlier. To use this token for authentication, you need to change

the `auth_with_token` parameter to True in the instantiation of the loader.

```
from langchain.document_loaders.onedrive import OneDriveLoader

loader = OneDriveLoader(drive_id="YOUR DRIVE ID", auth_with_token=True)
```

## Documents loader

### Loading documents from a OneDrive Directory

`OneDriveLoader` can load documents from a specific folder within your OneDrive. For instance, you want to load all documents that are stored at `Documents/clients` folder within your OneDrive.

```
from langchain.document_loaders.onedrive import OneDriveLoader

loader = OneDriveLoader(drive_id="YOUR DRIVE ID", folder_path="Documents/clients", auth_with_token=True)
documents = loader.load()
```

### Loading documents from a list of Documents IDs

Another possibility is to provide a list of `object_id` for each document you want to load. For that, you will need to query the [Microsoft Graph API](#) to find all the documents ID that you are interested in. This [link](#) provides a list of endpoints that will be helpful to retrieve the documents ID.

For instance, to retrieve information about all objects that are stored at the root of the Documents folder, you need make a request to: `https://graph.microsoft.com/v1.0/drives/{YOUR DRIVE ID}/root/children`. Once you have the list of IDs that you are interested in, then you can instantiate the loader with the following parameters.

```
from langchain.document_loaders.onedrive import OneDriveLoader

loader = OneDriveLoader(drive_id="YOUR DRIVE ID", object_ids=["ID_1", "ID_2"], auth_with_token=True)
documents = loader.load()
```