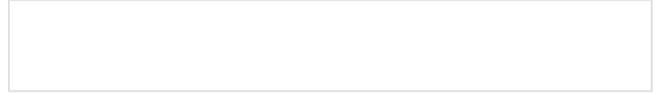


# SharePoint Online - Features



## Introduction

---

The SharePoint Online connector will crawl content from any SharePoint Online site collection URL. The connector will retrieve Sites, Lists, Folders, List Items and Attachments, as well as other pages (in .aspx format). This connector supports SharePoint running in the Microsoft 365 offering.

**This is not a Microsoft 365 connector**, the individual repository offerings within Microsoft 365, such as OneDrive, Calendar, Tasks, Yammer will have their own connectors.

## Environment and Access Requirements

---

### Repository Support

The SharePoint Online connector supports crawling the following the repositories:

Repository	Version	Connector Version
SharePoint	Microsoft 365	5.0

## Account Privileges

The connector offers two authentication options to access the SharePoint REST API: user account or Azure AD application.

### User Account

To configure a user crawl account use the following, see [SharePoint Online - Crawl Account Access](#).

 To use a user crawl account on multiple site collections, you'll have to follow the steps on each site collection the access is needed.

### Azure AD Application

To configure an Azure AD application for crawling, see [SharePoint Online - Azure AD Access](#).

 Using an Azure AD Application will grant access to all site collections under the tenant.

### Azure AD Application (with Delegated Permissions)

Azure AD Applications can be used with Delegated Permissions from a service account. The Application does not need to have **Admin** "Configured Permissions", it only needs to be authorized by an Admin account. See [SharePoint Online - Azure AD with Delegated Permissions](#).

 Using an Azure AD Application can be used to use Azure Authentication, restricted to the sites a service account can see.

## Environment Requirements

The connector uses SharePoint's REST API, so the Aspire Worker nodes must have internet access to connect to the Microsoft 365 environment. Optionally, you can configure a proxy on the connector to enable internet access.

## Framework and Connector Features

---

## Framework Features

Name	Supported
Content Crawling	yes
Identity Crawling	yes
Snapshot-based Incrementals	yes
Non-snapshot-based Incrementals	yes
Document Hierarchy	yes

## Connector Features

The SharePoint Online connector has the following features:

- Item filtering using include and exclude regex patterns. This is based on the item's display URL.
- Access Control Lists (ACLs) fetching, for document level security.
- Support for BCS external lists.
- Performs non-snapshot-based incremental crawling (so that only new/updated documents are indexed) using SharePoint's change log tokens.

## Content Crawled

The SharePoint Online connector can crawl the following objects:

Name	Type	Relevant Metadata	Content Fetch & Extraction	Description
Sites	container	<ul style="list-style-type: none"><li>• Last Modified Date</li></ul>	N/A	Any site or subsite underneath a seed. Not the same as the .aspx page for a SharePoint Site
Lists	container	<ul style="list-style-type: none"><li>• Last Modified Date</li><li>• Data size</li></ul>	N/A	Any type of SharePoint list including (but not limited to): Document Libraries, External Lists, Calendars, Task Lists, etc.
Folders	container		N/A	List Item Folders found on lists like Document Libraries or Link Lists.
ListItems	document		Yes	ListItems can take a number of different formats. For example, documents (PDF, doc, ppt, etc.), calendar events or announcements. For more info on how ListItems content types work, go to the MSDN <a href="#">article</a> .
Attachments	document		Yes	A document attached to a SharePoint List Item.

## Limitations

Due to API limitations, the SharePoint Online connector has the following limitations:

- The connector uses the REST API to access SharePoint databases directly; it doesn't use web crawling
- Crawling is only supported using a Site or a List as a root URL.