Box Crawler Setup

Introduction

This is a short guide aimed at showing how to configure Box to enable the SimSage Box crawler.
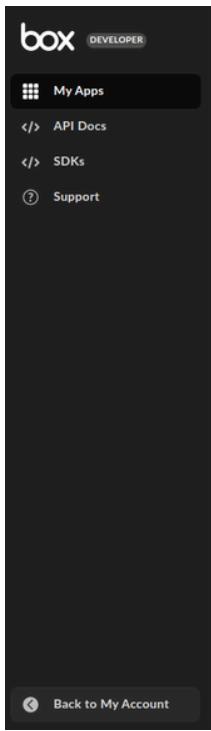
Prerequisites

In order to create a SimSage Box crawler, we must first have the following credentials at hand:
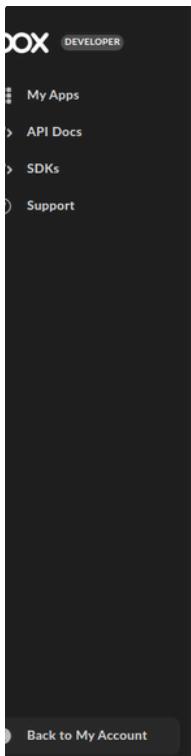
- Client ID
- Client Secret
- Enterprise ID
- Folder List

Setting Up
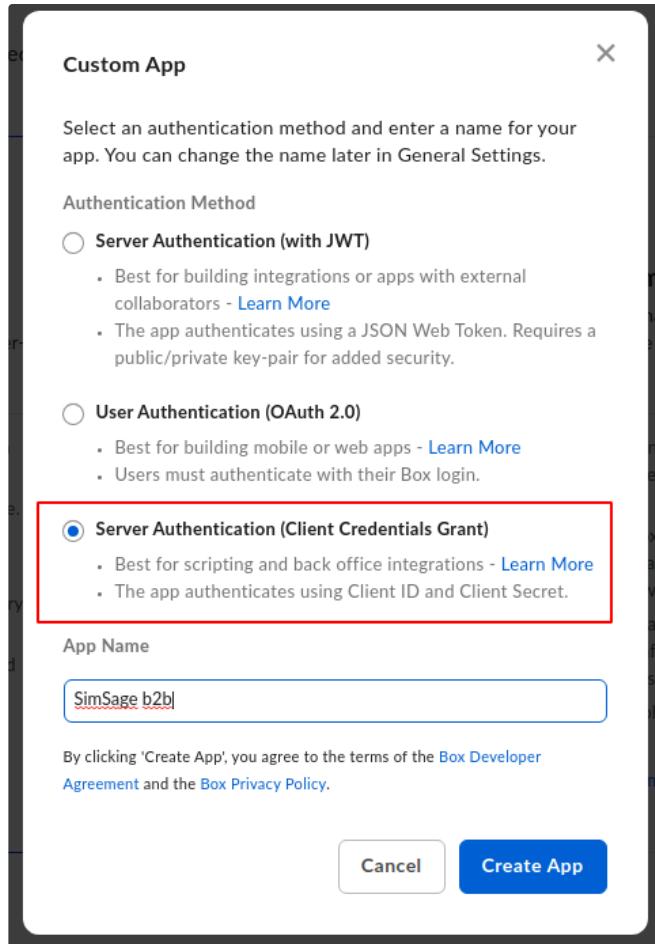
Sign-in to your Box account and visit https://app.box.com/developers/console:

Click "Create New App" in the top right-hand corner

Select "Custom App"

Select "Server Authentication (Client Credentials Grant)"

Give the App a name, the example name given here is "SimSage b2b".

Click "Create app"

Copy the "Enterprise ID" shown on the screen, this number is required by the SimSage Box crawler (the example show "1341231" as the Enterprise ID").

Click the "Configuration" tab

Copy the Client ID visible in the Oauth 2.0 Credentials section. this ID is required by the SimSage Box crawler. Scroll down the page until you see the App Access Level and switch to "App + Enterprise Access".

Scroll down to the bottom of the page and find the "Advanced Features". Check the "Make API calls using the as-user header" and "Generate user access tokens" boxes.

Click "Save Changes", you should see "Successfully update the app." as shown below.

You now need to "enable" this application. The Box administrator must approve these changes. If you are the Box administrator, you will receive an email requesting approval. Once approved, the "Authorization" tab of the this section will show the screen shown below.

Once the application is "Enabled" you can start crawling Box using SimSage.
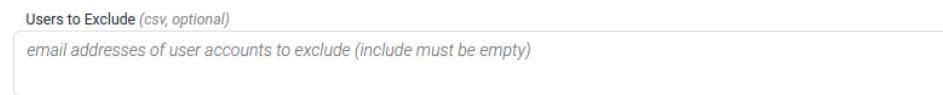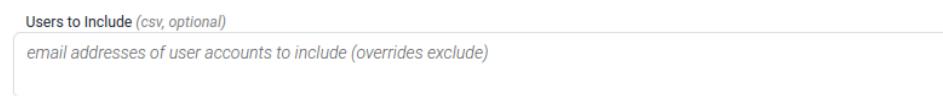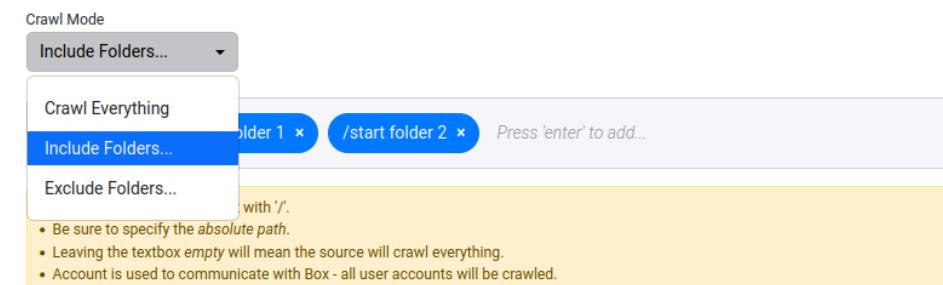
Folder Filtering

Once you have successfully set-up your box account we can begin crawling through specific folders, you have *three* options available to you:

- Include Folders
- Exclude Folders (beginning from root)
- Crawl Everything

> 🟪 Folders need to be reference by their *absolute path*

To select on of these three options, select from the drop down shown below:

if the option selected is to include or exclude folders, a text box will appear prompting you to specify the *paths* of these folders.

> 🟪 By default - all *sub-directories* are also resolved when including/excluding folders

> 🟪 Excluding folders or users can only be done once at source creation time.