Confluence Introduction

The Confluence connector will crawl content from any Confluence content repository. The connector will retrieve spaces, pages, blogs, attachments and comments.

The connector uses the Confluence REST API to crawl Confluence content and we support Confluence On-premise installation.

Features

Some of the features of the Confluence connector include:

- Performs incremental crawling (so that only new/updated documents are indexed)
- Fetches access control lists (ACLs) for document level security
- Is search engine independent
- Runs from any machine with access to the given Confluence URLs
- Suppports HTTP and HTTPs.
- Designed for supporting early binding mechanisms

For a Complete tutorial on Confluence see here

Summary of Confluence organization

This is the hierarchy of spaces/pages/blogs/attachments/comments for Confluence versions:

- **Dashboard:** Is the first page you see when login to Confluence, it provides quick access to the top level features of Confluence.
  - **Spaces:** Spaces are containers that group content related to a specific theme or topic. Spaces contain pages and blogs.
    - **Pages:** Like a web page or a page in a book, pages are places where you write content related to a specific theme or topic. Pages can contain attachments and comments
      - **Attachments:** Documents (images, files, videos, etc) that are embedded in a page or blog and contain relevant information about the topic or theme the page/blog is talking about.
      - **Comments:** Remarks users leave on a page or blog to share information with other users.
    - **Blogs:** A blog is a discussion or informational site published on the World Wide Web and consisting of discrete entries ("posts") typically displayed in reverse chronological order. Blog. Confluence blogs can contain attachments and comments
      - **Attachments**
      - **Comments**

Content Retrieved

The Confluence connector retrieves several types of documents, listed below are the inclusions and exclusions of these documents.

Include

- Spaces
- Pages
- Blogs
- Attachments
- Comments