Azure Data Lake Introduction

The Azure Data Lake connector will crawl content from the Azure Data Lake Store cloud at either root or specified paths.

An Azure Data Lake makes it easy for developers, data scientists, and analysts to store data of any size, shape, and speed, for all types of processing and analytics across platforms. It removes the complexities of storing data while making it faster to get up and running with batch, streaming, and interactive analytics. Azure Data Lake works with existing. It integrates seamlessly with operational stores and data warehouses so you can extend current data applications.

For more information about the Azure Data Lake Store, see the official Microsoft Overview of Azure Data Lake Store documentation.

Features

Some features of the Azure Data Lake connector include:

- Performs incremental crawling (so that only new/updated documents are indexed)
- Fetches Object ACLs (Access Control Lists) for Azure document-level security
- Runs from any machine with access to the given Data Lake source
- Service-to-Service Authentication via OAuth 2.0 token

Content Retrieved

The Azure Data Lake connector retrieves any type of document and its attributes. Listed below are the included documents.

Include

- Documents stored in folders
- Documents stored in subfolders

Future Development Plan

No new features have been identified to add to the development plan.

Anything we should add? Please let us know.