

HBase Connector Introduction

The HBase connector will crawl content from an HBase Server.

Features

Some of the features of the HBase connector include:

- Extracts data from HBase
- Configurable namespace and prefix
- Kerberos authentication

Content Retrieved

The HBase connector retrieves content as stored in the objectData field of the table in the HBase server.

Limitations

Due to API limitations, the HBase connector has the following limitations:

- It expects an established structure on the table to crawl:

id: MD5 id of the document

humanName: The document id in a human readable form

createdTimestamp: The timestamp of when the document was created

updatedTimestamp: The timestamp of when the document was last updated

crawlTimestamp: The timestamp of when the document was crawled

objectData: The Aspire object in json format that has the content of the document

binaryFilepath: The path of the document binary file

- Incremental crawls are not supported at the moment.