BA INSIGHT

Enterprise Integration

Technical Overview of BA Insight’s ConnectivityHub and Connectors

January 2020

Contents

Introduction

This paper describes the capabilities of the BA Insight ConnectivityHub. Based on our experience, connectivity is the cornerstone to any internal search/enterprise search implementation. ConnectivityHub is a platform that "connects the search dots" by securely indexing both full text and metadata from source systems into a single index, thus enabling a single, searchable result set across content from all sources. This paper drills into the ConnectivityHub architecture, explains what BA Insight's connectors are, highlights the challenges behind connectors, and defines which search indices ConnectivityHub can be used.

Here are some example use cases for ConnectivityHub:

- Connectors *without* security
  - Indexing content from public websites or documentation portals where everyone in the organization requires access to the content.
- Connectors *with* security
  - Indexing content from systems that only provide access to certain users or groups. Examples include, only enabling HR to search for personnel records and finance to see salary information; or replicating complex multi-level security models in systems where security is applied at the folder and item level.
- Datasets
  - Enhancing the metadata associated with your content by adding additional supplementary metadata for the content from another content source or application. For example, your document management system (DMS) holds documents that has metadata such as title, author, date modified, project name and project number. The finance system holds all the metadata regarding project budget and spend. There is no way to search on both sets of metadata (DMS and the finance system). Using a dataset, it is possible to index the content with metadata from both content sources. In turn, users (with the correct access) can search based on any of the metadata attached to the content.
- No-Search Target
  - Search engines are not the only destination that ConnectivityHub can push content to. A "No-search" Target enables you to push content to FTP sites (or other destinations), meaning users can share files between files and content.
- Cloud-based Search Solutions
  - ConnectivityHub not only provides the ability to securely add content to stand alone on premises search engines, but it also enables you to add content to Cloud-based search solutions such as Elastic Cloud and Microsoft Azure Cognitive Search.
- In-App Search
  - As part of our cloud search strategy, we have incorporated access points into multiple cloud based applications such as Salesforce, Dynamics, Outlook, Microsoft Teams and even ServiceNow. Content is added to a search index using the BA Insight ConnectivityHub and connectors. This content can then be accessed via the access points from directly within the application that users are working thus creating a single point of entry for organizational content.

ConnectivityHub – Architectural Overview

ConnectivityHub is our highly scalable, purpose-built platform upon which we build connectors for business and enterprise systems to search indices. It is also used by our customers to develop their own connectors.

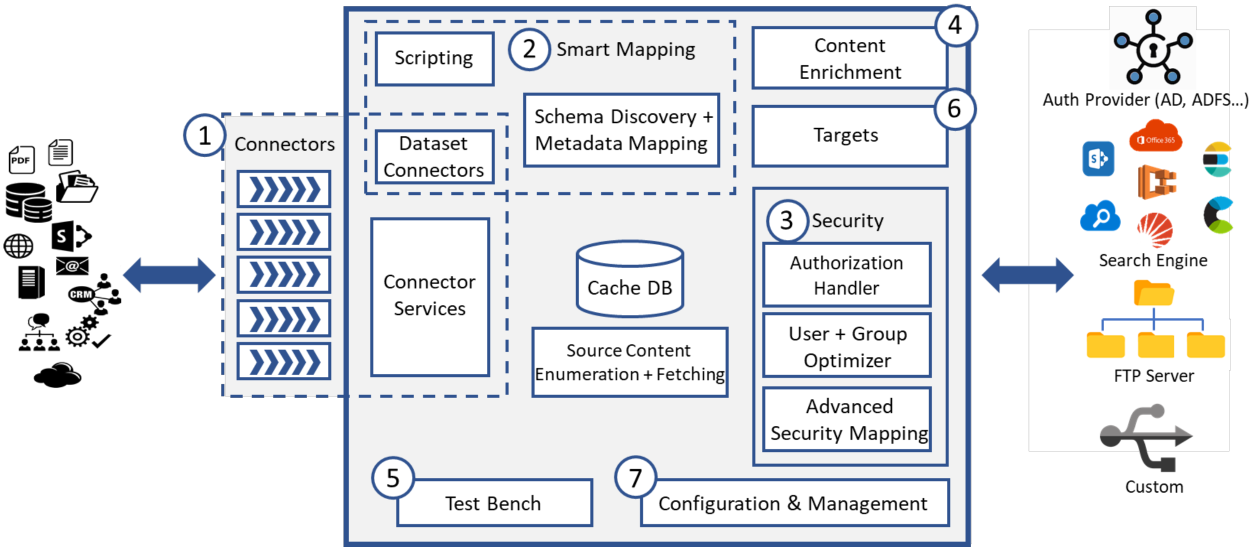


Figure 1 ConnectivityHub Architecture

The above diagram shows the components and seven stages that enable ConnectivityHub to securely index content into a search engine. The order of the stages and the stages required depend on the exact use case:

Stage 1 - Connectors enable you to securely access complex business systems without installing software on production systems.

Stage 2 – Smart Mapping of metadata includes scripting and schema management to index and augment metadata from source systems. When combined with dataset connectors, this also combines results from multiple sources into a single set of logical items.

Stage 3 – ConnectivityHub provides mappings across the heterogeneous security schemes used by different source systems. 'Early binding' security makes it possible to deliver secure, high-performance search solutions.

Stage 4 – Content Enrichment allows smart processing of content pushed to the index via Classification, Machine Learning, Image or Video processing (OCR, text to speech, etc.) and Natural Language Processing.

Stage 5 - Test Bench enables the testing of results, security, metadata mapping and content enrichments before the content is pushed to the search index.

Stage 6 – Targets provide a mechanism to direct content into a search engine, or even another location, such as an FTP site.

Stage 7 - Configuration and Management is not specifically a stage within itself. This is something that is used throughout the process. ConnectivityHub components enable administrators to configure stages 1 to 6 via a graphical user interface. In addition to the above functionality, administrators can configure indexing schedules, use the ConnectivityHub Wizard to set up new connections, and manage alerts and logging.

The next sections provide a more comprehensive deep dive into all the BA insight ConnectivityHub features and stages.

1.Connectors

As mentioned above, ConnectivityHub is the platform that we use to build our connectors. As of today, we have over 90 different connectors. These connectors enable you to push content into the following search engines: SharePoint, O365, Elasticsearch, Elastic Cloud, Elasticsearch for AWS, Azure Cognitive Search and Solr.

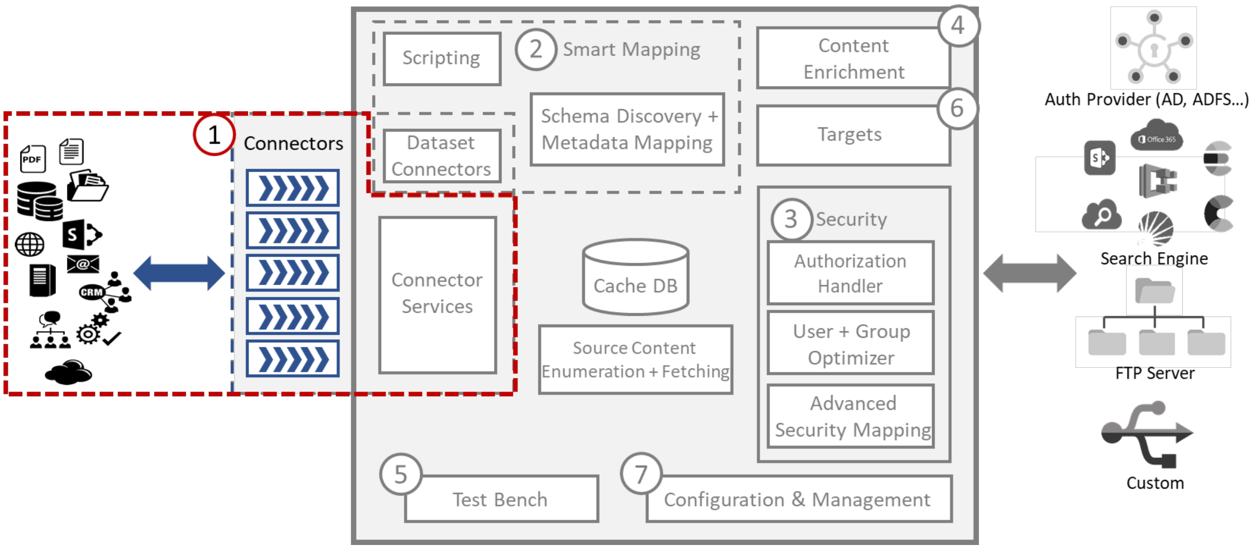


Figure 2 ConnectivityHub Architecture: Connectors

Each connector is developed and maintained for a particular source system. Connectors include several functions and communicate to ConnectivityHub through a published Web Services-based API. The structure of these connectors is shown below:

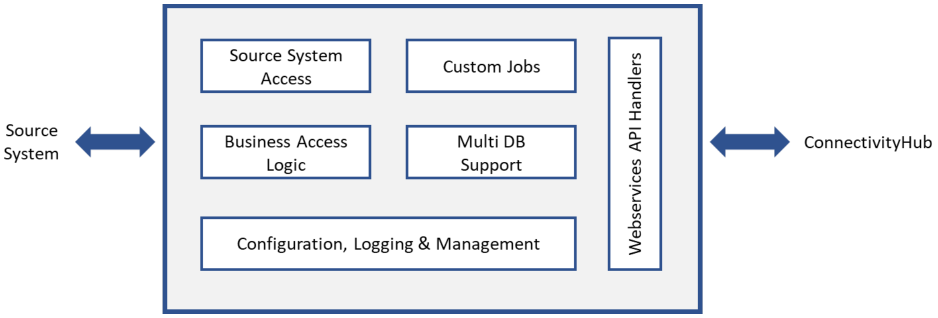


Figure 3 Connector Services

Connectors extract content from source systems and transmit it to a search engine for indexing. Each enterprise repository typically has a specific mechanism through which content can be requested (access method or API), a particular layout of content (schema), and specific security capabilities. Therefore, each system needs to have a connector developed specifically for that content source. Part of the process is to protect the security defined at the source system. This means that users never have access to content, through search results that they cannot access at the source system. Equally, we provide functionality for the administrator to override source system security in scenarios where this is required; e.g. HR need to perform an audit around sensitive data.

A connector establishes a secure connection to the source system and maps the content - including metadata and attachments - from the source system schema to the search engine schema. It then requests content and feeds it to the search engine. There are two main types of indexing operations:

- Full, which extract all desired content.
- Incremental, which extract only content that has changed since the last time the index was updated.

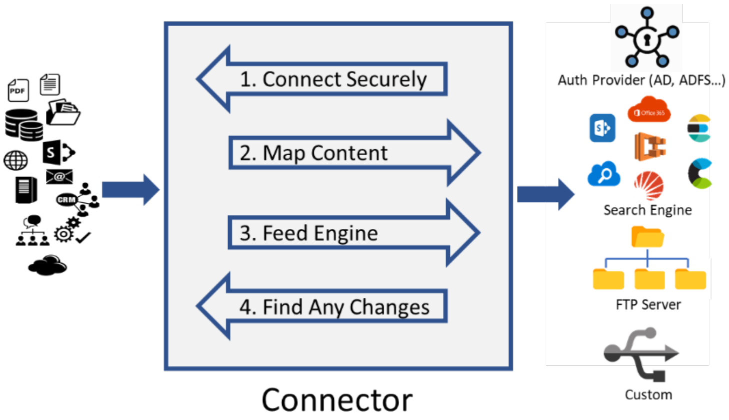


Figure 4 Connector indexing Process

Connectors are designed for high throughput and light touch when selecting and extracting content. They are agentless – i.e. no software needs to be installed on the source system and can communicate over a network to remote systems. They only require read access, so there is no risk of compromising the integrity of content in source systems.

Many of BA Insight’s connectors can also act as dataset connectors. For example, a SQL system may have an associated file system for raw storage, or a file-based system may have an associated database holding metadata. In these cases, the file and the metadata are “joined” and indexed as a single item.

Capturing content is fundamental to search - if it's not indexed, then you can't find it! Yet many organizations struggle to incorporate all needed content into search – it is much harder than it may seem. An understanding of the basics of connectors sets the groundwork for the rest of this paper.

Creating New Connectors

BA Insight has extensive experience in creating and maintaining connectors, and a proven process for approaching new systems. We have many connectors on our roadmap, which are prioritized based on market demand. When a customer wants a roadmap connector, we meet with them, understand their requirements (for example, complex security requirements) create a connector specification document, develop it, test it and deliver it.

For connectors that are not on our roadmap, we perform market research, evaluate them, decide whether there is a market for them (which in most cases, there is) and then follow the same process as described above.

As our ConnectivityHub provides facilities for testing, troubleshooting, and optimizing content extraction, it means that creating new connectors is faster and simpler.

Connectors built on this framework inherit many powerful features such as Smart Mapping and content enrichment. They also become part of our product suite meaning that they are fully supported.

2. Smart Mapping

Mapping metadata schemas between source systems and a common search index is one of the most important tasks in setting up a search deployment. It is also one of the most laborious, but Smart Mapping makes this process much simpler by auto-generating property names and the metadata mapping based on the source system schema.

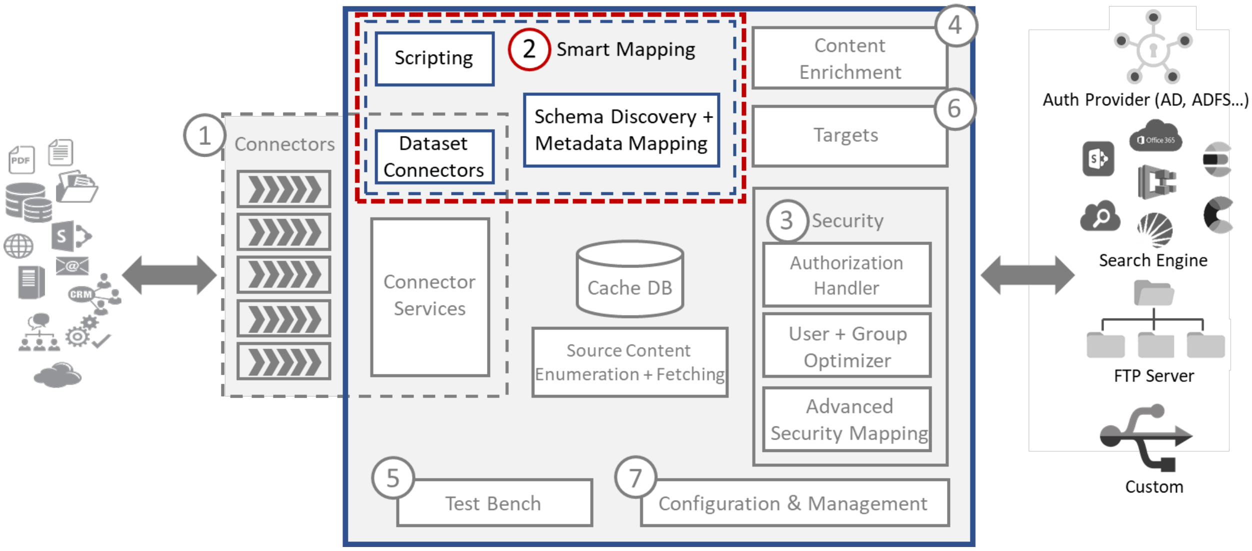
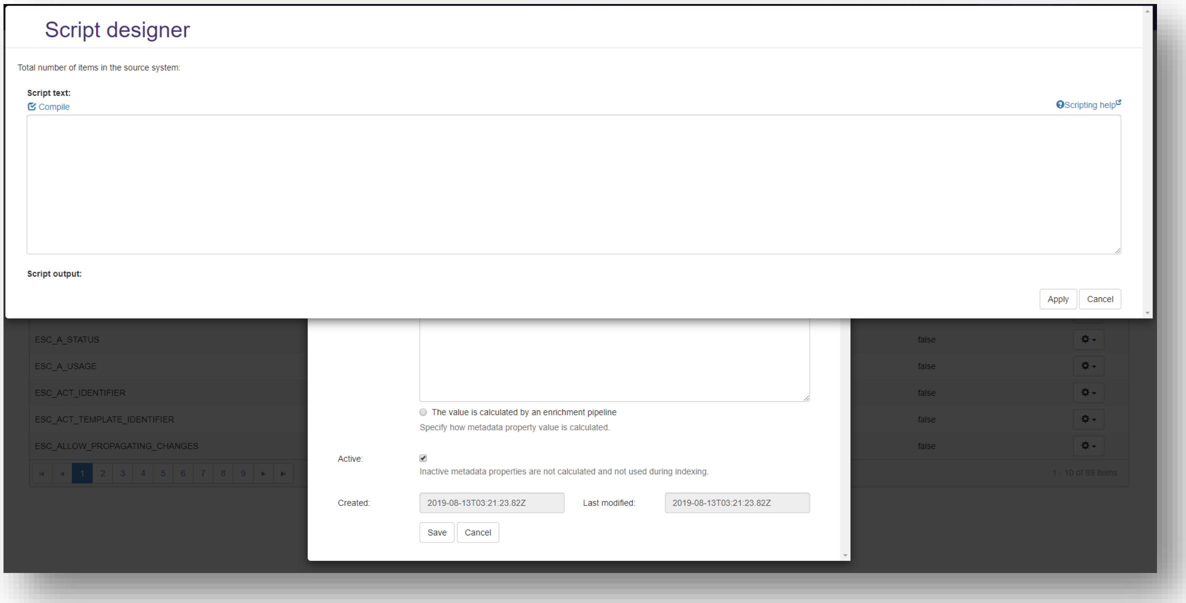


Figure 5 ConnectivityHub Architecture: Smart Mapping

Connections, content sources, items, and metadata can be configured and extended to customize the content and tailor how search fields are populated. Custom scripting, using familiar VBScript syntax, can be applied to connections, security, indexing, and metadata to handle even the most demanding applications. A Script Designer allows you to tweak and test scripts right from the Test Bench, and a library of sample scripts gets you going quickly.



Script designer

Total number of items in the source system:

Script text:

☑ Compile

Scripting help

Script output:

Apply Cancel

ESC_A_STATUS	false
ESC_A_USAGE	false
ESC_ACT_IDENTIFIER	false
ESC_ACT_TEMPLATE_IDENTIFIER	false
ESC_ALLOW_PROPAGATING_CHANGES	false

1 2 3 4 5 6 7 8 9

1 - 10 of 89 items

○ The value is calculated by an enrichment pipeline

Specify how metadata property value is calculated.

Active: ☑

Inactive metadata properties are not calculated and not used during indexing.

Created: 2019-08-13T03:21:23.82Z Last modified: 2019-08-13T03:21:23.82Z

Save Cancel

Figure 6 Metadata Script Designer

Dataset Connectors are another element of Smart Mapping. These provide a way to enrich indexed content with metadata from an associated content source. Metadata can be retrieved from multiple content systems, and it can be filtered to allow for fine granularity when matching the metadata to the content. This ultimately provides a richer and more accurate search result to the end user.

Dataset connectors essentially look up and "join" information across various systems during the indexing process. For example, a dataset connector can combine customer data across an ERP system and a CRM system. Imagine indexing customer billing records from the ERP and retrieving the market segment designation and sales territory of the customer from the CRM system. The user performing a search can then see this information associated with each record and use it for faceting and navigation. When exploring market information, the user also sees customers in that specific market segment.

3. Security

ConnectivityHub provides powerful security mapping across the heterogeneous security schemes used by different source systems.

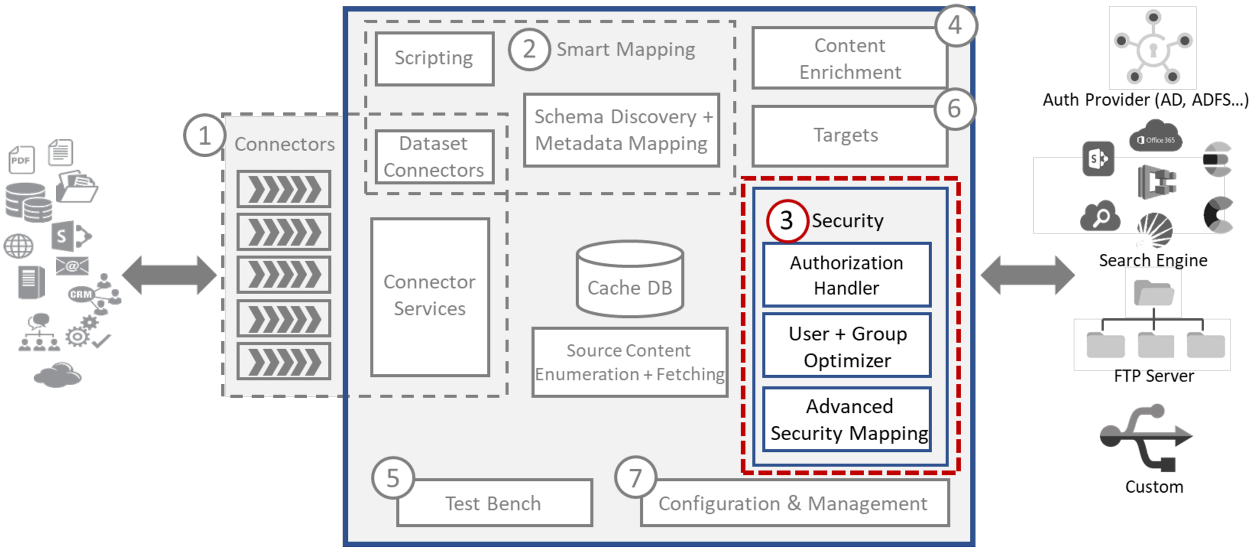


Figure 7 ConnectivityHub Architecture: Security

It identifies and maps security schemas from any system to support the early binding security needed for responsive and accurate search results. Active Directory-based systems benefit from automatic Active Directory group binding; non-Active Directory systems benefit from advanced security mapping that goes beyond the claims-based security of native search platforms. This means you can handle the toughest and most sophisticated security challenges across heterogeneous systems and ensure rigorous adherence to all permission and access protocols.

Advanced Security, including role-based and attribute-based security, handles the complex security scenarios that arise with sophisticated source systems such as ServiceNow, FileNet, HP Trim, Confluence, or Objective. As more source systems are included in a search application, the more complex the security tends to be. For example, deployments with connectors to multiple different cloud systems pose daunting security issues even if each system is relatively straightforward by itself. BA Insight's capabilities for advanced security are specifically designed for heterogeneous, complex search security scenarios.

4. Content Enrichment

ConnectivityHub provides content enrichment. It is possible to connect to external systems or processes and enrich content by adding relevant metadata and/or normalizing terms. This is done by making calls, via our connectors, to the supplementary services during the indexing of content. This process is called "Metadata Expansion". It is a powerful mechanism to provide superior findability and relevancy that wouldn't be possible with just the metadata from the original content source.

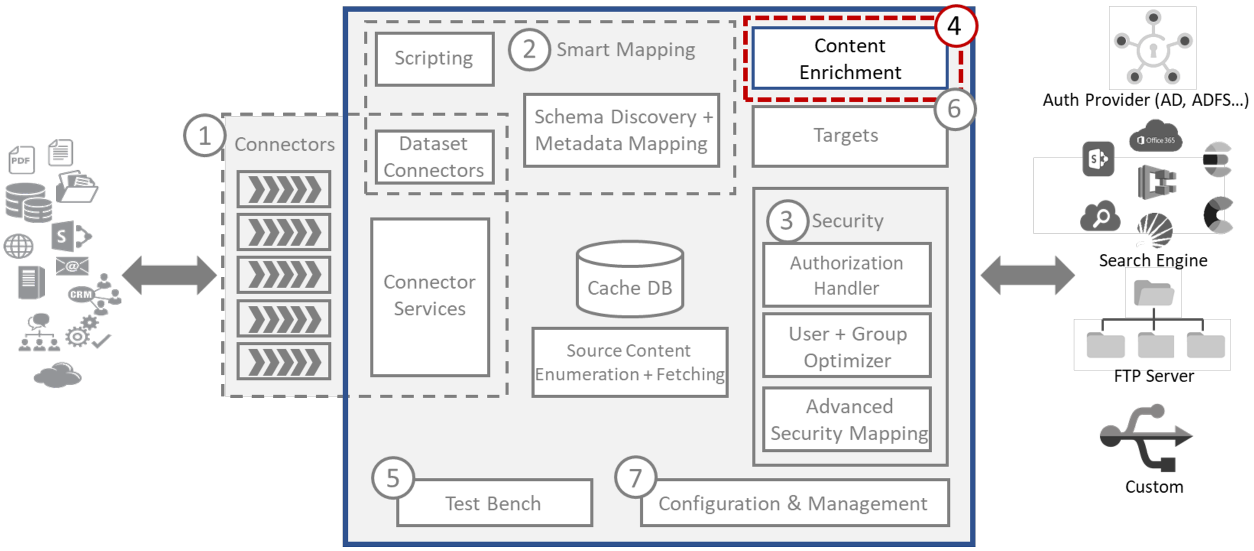


Figure 8 ConnectivityHub Architecture: Content Enrichment

For more complex data processing scenarios such as classification, multi-media processing, translation or concept extraction, ConnectivityHub can be seamlessly integrated with BA Insight's AutoClassifier product or other classification technologies to apply more advanced processing logic. Such logic includes machine learning-based models to classify content, OCR and object recognition (i.e. extracting text from images and images stored in documents or video transcript creation), automatic summarization, language detection and key concept extraction, sentiment analysis, etc. This information can be subsequently used to feed the search engine or decide how the content should be processed. For instance, content identified as sensitive or containing personally identifiable information can be filtered out at indexing time regardless of where the content originated.

5. Test Bench

An integrated Test Bench makes it possible to test connectivity to confirm correct configuration of the connectors before it is pushed to a search index.

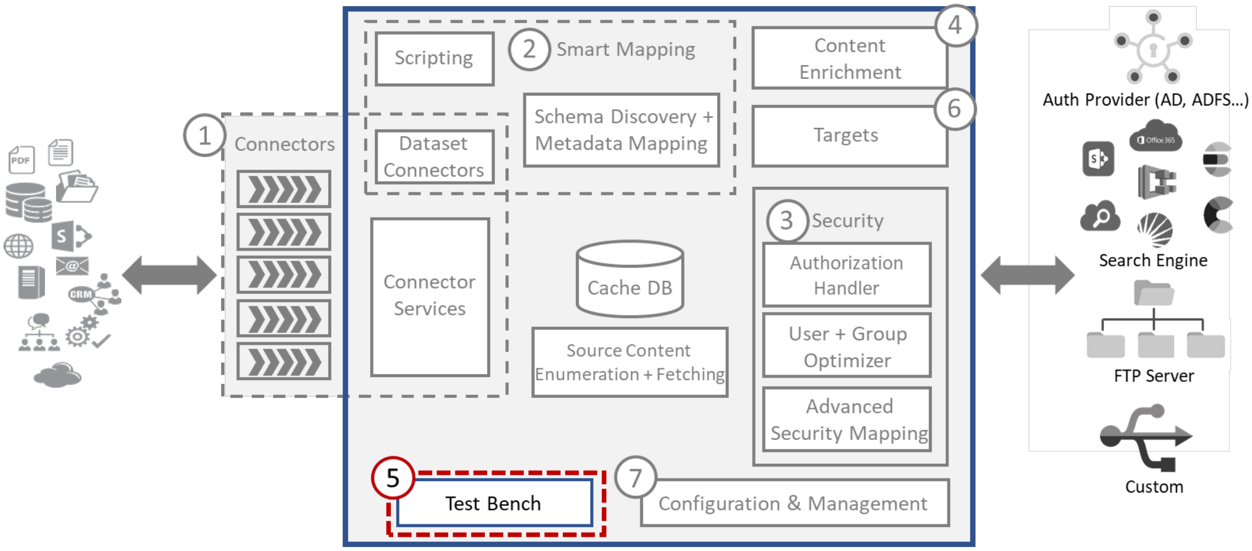


Figure 9 ConnectivityHub Architecture: Test Bench

Administrators can test the output of any defined content source, display all the properties returned for each item, and provide visibility into performance, security, and metadata contents. Since no content is indexed by Test Bench, it is ideal for verifying or troubleshooting connector configurations.

6. Targets

Targets allow you to specify which search index to populate. They enable you to define the data that will be pushed into the search index e.g. the document types to include or exclude, how the data should be mapped i.e. via a search schema and even any additional security that the search schema may use.

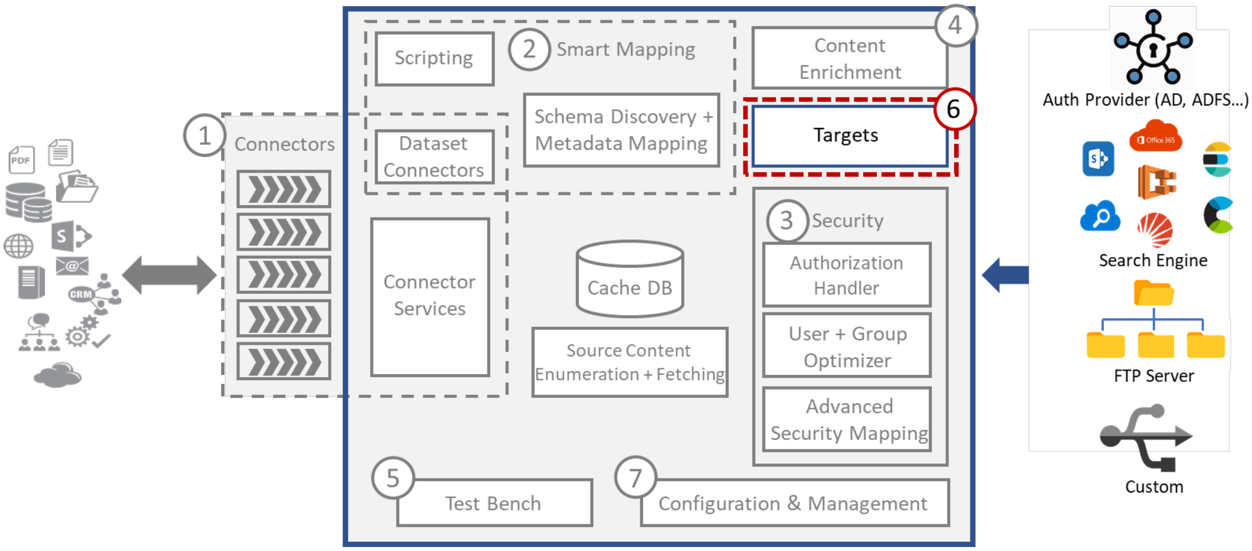


Figure 20 ConnectivityHub Architecture: Targets

Updating the index is handled via scheduled jobs with access to historical data for straightforward administration. It is possible to run full Target updates along with incremental Target updates too. It is even possible to clear the index, via a "Content Reset" and start again if needed.

No-search Targets

Besides the standard search index scenario, ConnectivityHub can also be used to push content to other destinations such as an FTP site. The mechanism can be, for instance, used to maintain two systems in sync or to notify a remote system when documents matching specific criteria exists, are created or updated.

7. Configuration and Management

ConnectivityHub makes it easy to administer and configure connectors, metadata mapping, and content targeting for all connections. It provides facilities that simplifies configuration, operation, and troubleshooting of the overall system - reducing administrative effort and speeding problem resolution.

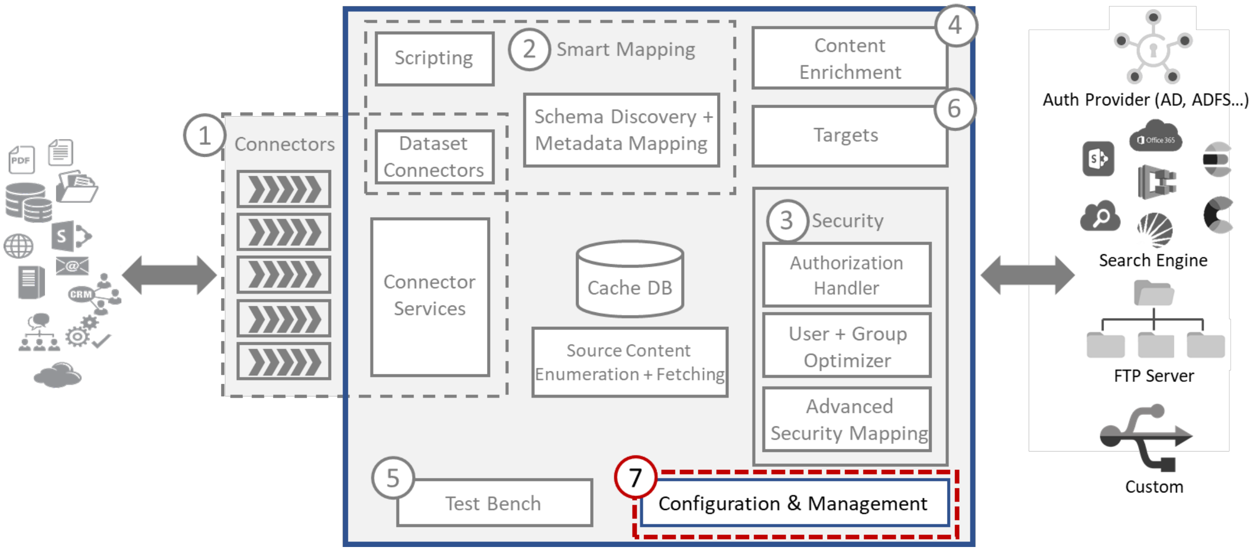


Figure 31 ConnectivityHub Architecture: Targets

Using ConnectivityHub, there are two options when indexing content:

1. Use the ConnectivityHub wizard.
2. Manually create a new connection.

Use the ConnectivityHub Wizard

The ConnectivityHub wizard enables administrators to quickly set up new connections. It is accessed through the "Quick Mode" option in the ConnectivityHub menu and intuitively guides the administrator through the complete set-up process including creating a new connection to the application or database and automatically running all tasks required to set up metadata and security.

Manually Creating a Connection

Manually creating a new connection is typically done when there is a need to use some of the more advanced features of ConnectivityHub.

Administrators can set up a new connection to the application i.e. Documentum, Exchange, OneDrive etc. and configure security. They can configure which content is to be added to the search engine e.g. content types, file types, which repositories should be indexed and even use scripting for those more particular or specific requirements. Administrators can also utilize some of the ConnectivityHub advanced functionality such as Content Enrichment and Dataset Mappings.

Once all of the configuration has been set up and checked, the content can be pushed to the search index using a Target via a "Full Target" update or "Incremental Target" update. Administrators can also use the handy schema mapping feature that permits the mapping of metadata properties to the specific requirements of particular search indices.

ConnectivityHub for Microsoft Search/O365, Elastic and Azure

BA Insight's ConnectivityHub can be used with many search engines including Microsoft Search/Office 365, Elasticsearch and Azure Search. This section of the whitepaper explores the integration between ConnectivityHub and the supported search engine(s).

Microsoft Search/O365

In the near term, you will need to use the BA Insight Connector Framework to index content into O365. This is currently being used by more than 100 customers. At present, there is a Microsoft requirement to use something referred to as the Microsoft Cloud Hybrid Search Service Application (Cloud SSA). It allows for an index in O365 to include content from SharePoint on-premises and other sources.

As Microsoft Search evolves and becomes available, ConnectivityHub can be used as an extension of Microsoft Search. For customers who are using the BA Insight Connector Framework with the Cloud SSA, they can migrate as they choose.

Elasticsearch

ConnectivityHub supports Elasticsearch online and on premises, along with Amazon Elasticsearch. Unlike some search engines, for example SharePoint, where there is a single index, when using ConnectivityHub with Elasticsearch, one search index is created per content source. This provides a lot of flexibility of management.

Normally, content is added to an Elasticsearch index as plain text. This provides limitations around what can be searched (i.e. documents are not searchable). In order to be able to perform full text searching against ingested documents, the Ingest Attachment Processor Plugin must be installed in Elasticsearch.

Security is an incredibly important part of any search engine. OpenSource Elasticsearch out of the box does not provide item level security. It is only available with Elasticsearch when using their X-Pack product.

ConnectivityHub and BA Insight ingestion connectors handle security using Advanced Security Mapping, meaning that organizations do not require X-Pack.

Azure Cognitive Search

ConnectivityHub works with Azure Cognitive Search in a similar manner as it works with Elasticsearch. One search index is created per content source as again, this provides a lot of flexibility of management. ConnectivityHub with Azure Cognitive Search uses the Tika Text Extraction Service to extract text from documents/content so that users can search both inside documents/content and its metadata.

Azure Cognitive Search has no built-in security, meaning that ConnectivityHub and BA Insight ingestion connectors utilize the Advanced Security Mapping feature of BA Insight SmartHub to provide item level security.

Apache Solr

ConnectivityHub works with Apache Solr in a similar manner as to how it works with Elasticsearch. One search index is created per content source as again, this provides a lot of flexibility of management. ConnectivityHub with Apache Solr can either pre-extract the content of binary files using the Tika Text Extraction Service or rely on Apache Solr directly to perform the same task.

Naturally, ConnectivityHub also support both HTTPS and password-based authentication when connecting to Apache Solr.

Summary

Search connectors are more challenging than they may seem. They require a balance of high performance and light touch, along with rigorous security and easy administration across a wide range of sophisticated source systems. BA Insight has a proven architecture, a wide range of supported connectors, and extensive experience to ensure secure successful search deployments.

BA Insight's ConnectivityHub provides a robust, flexible hub for secure, high throughput content integration. Powerful security integration across heterogeneous and complex security schemes is built-in. Smart Mapping provides automatic mapping of metadata properties; Dataset Connectors that support lookup and content normalization; and flexible content processing. Finally, using BA Insight connectors, you can add content into multiple search engines including O365, Elasticsearch, Azure Search and Solr.

About BA Insight

As an innovator in AI-driven search, BA Insight’s best of breed approach helps companies make search intelligent by providing technology that connects machine learning, cognitive computing, and enterprise systems, powering a new generation of intranets and cognitive search solutions. Our customers have the freedom to leverage the best search engines and cognitive computing capabilities available, providing users with an internet-like search experience while saving them precious time looking for needed information. We support multiple search platforms including Azure Search; Elasticsearch, Elastic Cloud; and Elastic Cloud Enterprise; and SharePoint search (online, on-prem, and hybrid).

Our modular software product portfolio features SmartHub, delivering a personalized, internet-like user experience; connectors, providing secure connectivity to a wide variety of systems; classification, increasing findability using auto-tagging, text analytics, and metadata generation; and analytics, providing valuable data to make intelligent decisions about your intranet.

Hundreds of organizations and over 3.5 million users benefit from BA Insight’s software on a daily basis to provide compelling intranets that people love to use. This includes respected organizations such as the Australian Government Department of Defence, CA Technologies, Chevron, DLA Piper, Keurig Green Mountain, Mars, Pepsi, Pfizer, and Travers Smith. BA Insight is a Microsoft Gold Certified Partner, a member of the Microsoft Enterprise Cloud Alliance, and an Elastic Partner.

Visit www.BAinsight.com for more information and follow us at @BAInsight.