



Secure Connectivity to Enterprise Systems through Search

Technical Overview of BA Insight's Connectivity Software

| | |
|--|-----------|
| Introduction | 3 |
| Enterprise Content is Exploding | 3 |
| Agile Information Integration via Search | 3 |
| The BA Insight Software Portfolio | 4 |
| Content Connectivity | 4 |
| The Fundamentals of Indexing Connectors | 5 |
| What Do Connectors Do?..... | 5 |
| What Makes Connectors for Search Hard?..... | 6 |
| The Connector Framework | 7 |
| Secure Unified View | 8 |
| Unparalleled Security | 8 |
| Administration and Configuration | 9 |
| Smart Mapping | 10 |
| Targets..... | 11 |
| Indexing Connectors | 12 |
| Supported Source Systems..... | 13 |
| Creating New Connectors | 14 |
| Scalability and Performance | 14 |
| Advanced Scenarios | 15 |
| Summary | 16 |
| About BA Insight | 16 |

Introduction

BA Insight provides software that enables organizations to rapidly deploy powerful search-driven applications at a fraction of the cost, time, and risk of alternatives. This includes the Connectivity software, which provides secure connectivity to a wide variety of content systems, enabling unified views of all knowledge assets.

This whitepaper describes the capabilities and architecture of the Content Connectivity software, in particular the Connector Framework and Indexing Connectors, which provide high performance, secure crawling of content from many different enterprise systems. These securely index both full text and metadata from source systems into SharePoint's search engine, enabling a single searchable result set across content from multiple repositories. They also support a variety of important scenarios beyond traditional search.

Enterprise Content is Exploding

Corporations are experiencing exponential growth in the volume of information they have under management, and they are also adopting new business systems at a rapid rate. The result is that modern organizations now manage an increasing number of burgeoning and unwieldy data repositories, both on-premise and in the cloud.

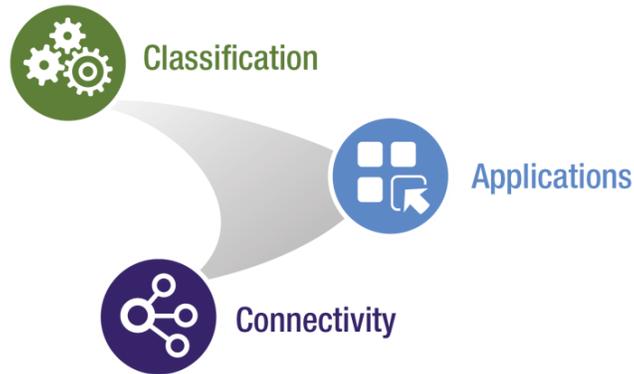
In an ideal world, businesses would know more about the information they have and where to find it. They would be able to manage it more closely with their needs and user responsibilities -viewing it in its entirety to make decisions or re-engineer key processes. Unfortunately, we do not live in an ideal world. The vast majority of corporate data is held in a manner that precludes both accessibility and ongoing management. Much of it is unstructured and cannot be found, or it is resident in application siloes with little relevance beyond its immediate application context. The wide range of proprietary data formats and the various security models that underpin such systems only serve to further exacerbate the problem. The result is that most large enterprise employees are fundamentally unable to access information that is important to them.

Agile Information Integration via Search

Due to the pervasive nature of information fragmentation and access problems, most large organizations have a need for a flexible, secure, and cost-effective platform for information integration and access.

The ability to crawl multiple systems, discover new and modified information, and map this into a consolidated view is remarkably powerful and effective. A single searchable result set across content from multiple repositories is a natural experience for end users, and the schema-less nature of a search index eliminates many of the headaches of traditional information integration.

The BA Insight Software Portfolio



BA Insight provides connectivity, classification, and application software that accelerates and future-proofs the creation of cloud-based and on-premise SharePoint portals for enterprises, transforming how users access information. Our software dramatically reduces implementation time, cost, effort, and risk while providing a better user experience.

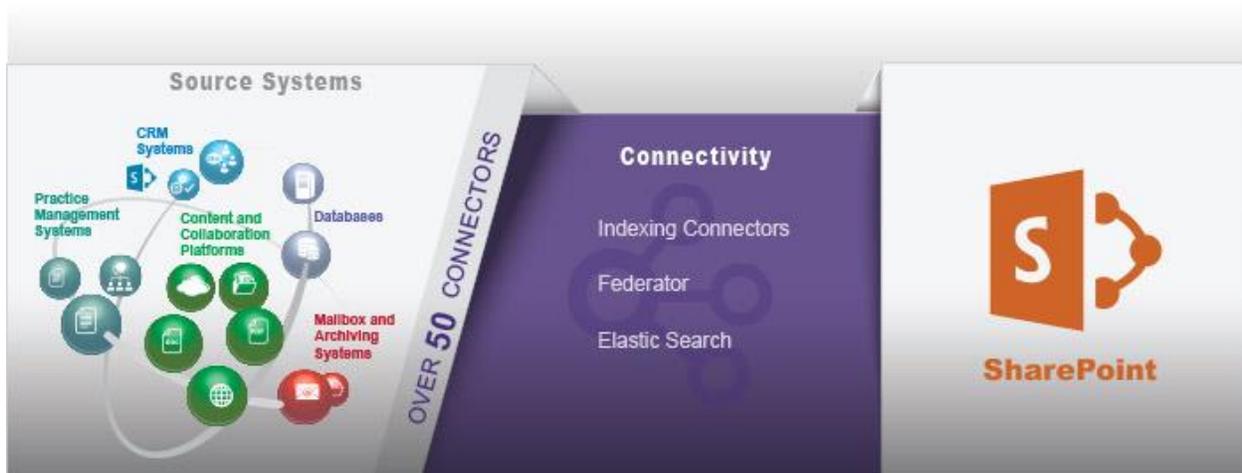
BA Insight's software portfolio provides a comprehensive set of capabilities:

- Applications - Smart Previews, Visual Refiners, User-Generated InfoSites, and Matter Comparison help users find relevant information faster for improved productivity.
- Classification - Auto-tagging, metadata generation, and text analytics make content findable.
- Connectivity - Secure connectors and federation to a wide variety of content systems, enable unified views of all knowledge assets.

BA Insight's Software Portfolio offers full flexibility in how it can be purchased, configured, and implemented. We support deployments of all types - from multi-component, enterprise-wide rollouts to departmental, component-based, phased approaches – and everything in between.

Content Connectivity

The Connectivity software within the BA Insight Software Portfolio provides secure connectivity and information integration with many of the most common data repositories available. It handles the persistent security mapping, data connectivity, context mapping, and enrichment that businesses need to leverage their information to make crucial business decisions.



The Fundamentals of Indexing Connectors

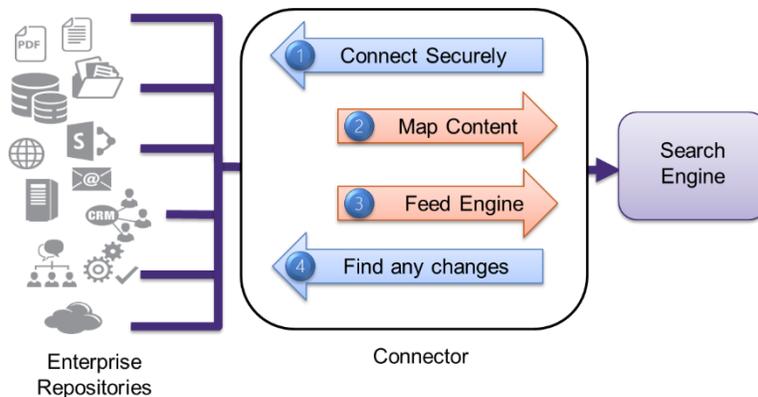
Capturing content is fundamental to search - if it's not crawled and indexed, you can't find it! Yet many organizations struggle to incorporate external content in search – it is much harder than it may seem. An understanding of the basics of indexing connectors sets the groundwork for the rest of this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

What Do Connectors Do?

Indexing connectors extract content from source systems and transmit it to a search engine for indexing. Each enterprise repository typically has a specific way to extract content (access method or API), a particular layout of content (schema), and specific security capabilities. Therefore, each type of system may need a connector developed specifically for it.

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 extracts content and feeds it to the search engine, in a process called crawling. There are two main types of crawls:

- Full crawls, which extract all desired content
- Incremental crawls, which extract only content which has changed since the last crawl



What Makes Connectors for Search Hard?

Many types of systems have connectors; for example most database systems have adapters to a variety of systems and ETL (extract, transform, and load) facilities. Some business process management (BPM) and workflow systems have connectivity to enterprise systems. SharePoint has a Business Connectivity Services (BCS) facility that allows different systems' content to be surfaced as an external list. However, none of these facilities will suffice for enterprise search.

Several key requirements on connectors used for enterprise search make them more difficult than they may seem:

Unstructured content: indexing connectors must work with unstructured content as well as structured data. Large documents, attachments, and complex systems with customer-configurable schema are typical. This is not what ETL systems are designed for, but it is essential for indexing connectors.

High throughput: in order to make content complete, a copy of every desired item must be indexed. This requires very high throughput. Many installations have millions of documents or even hundreds of millions. Just 1 million documents indexed at 1 document/second would take over 11 days for a full crawl! Throughput of hundreds of documents/second can be required in practice. This is out of the range of any approach that does an item at a time, including BPM systems and external lists.

Light touch: the flip side of high throughput is the need to minimize impact on the source system, which is usually a business-critical system in production. Indexing connectors must ensure that they will not impact the performance of these systems.

Security: it is essential that users see only content that they are entitled to see, especially because search is often used by many more people than, for example, BI. But security for search is particularly tricky, as we will detail later. ETL systems often disregard security as they move content into a data warehouse.

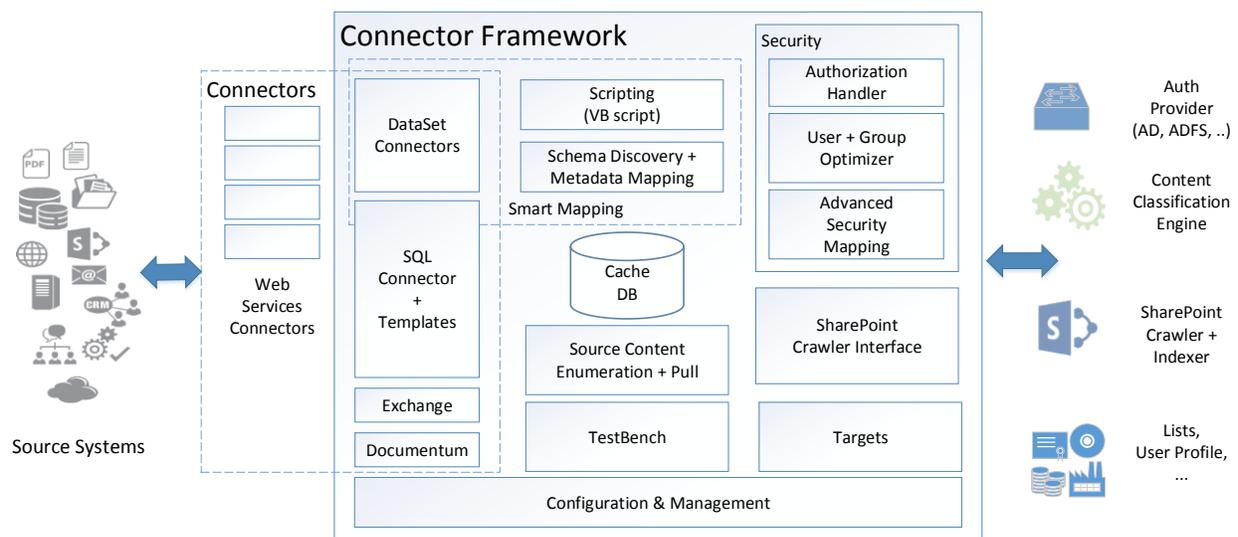
Click-through: with enterprise search, the source system retains the master information and the search index has only a representation (pointers). Users expect to click on a search result and be working on the original item or document. However, many enterprise systems require a specific method to bring up that

item. Indexing connectors must include the click-through method for each source system, which may also vary by the type of content referenced.

To compound this, applications of enterprise search tend to be heterogeneous. They have multiple sources, with different types of information, different schema, and different security models – all in one combined result set. It's no wonder that so many people underestimate connectors, only to find themselves in difficulty.

The Connector Framework

The heart of BA Insight's indexing connectors is the Connector Framework. It acts as a scalable hub for information integration, and also includes robust testing and administration features.



A set of components and capabilities plug into this hub, including:

- **Connectors** of various types (Web Services, SQL, embedded, and dataset connectors) which integrate securely with complex business systems without installing software on production systems.
- **Security Integration** across the heterogeneous security schemes used by different source systems. 'Early binding' security makes it possible to deliver secure, high-performance search solutions.
- **Smart Mapping** of content and metadata including scripting and schema management. This also provides powerful capabilities when combined with dataset connectors, such as associated crawls.
- **Targets** provide a mechanism to direct content into different places, in addition to search indexing. For example, content can be copied to a SharePoint list, along with its metadata and security settings.

Secure Unified View

BA Insight's Connector Framework provides full security and operates at high throughput to minimize crawl times – while maintaining a light touch on all source systems. It requires only read access and no client software needs to be installed on any source system server. It is scalable and incorporates redundancy for reliability as well as scale-out in content size and indexing throughput.

A library of pre-built content connectors, currently over 50, is available for a broad range of sources including both structured and unstructured content. Full support for attachments provides access to all the content in a source system. Flexible configuration allows you to index only the back-end system content you desire, presenting it to end-users in the manner they demand.

The result is seamless and simultaneous access to all content. A single consolidated search index, referencing content from many repositories, is surfaced as a single unified result set with appropriate relevancy ranking and faceted navigation. Common, consistent metadata can be created across all sources (using BA Insight's Content Classification software) to provide great findability and navigation. This approach maximizes the value of an organization's existing ERP, CRM, ECM, and messaging systems by securely unlocking and surfacing this information in a unified view.

Unparalleled Security

The Connector Framework provides powerful security integration across the heterogeneous security schemes used by different source systems. It identifies and maps security schemas from any system to support the early binding security needed for responsive and accurate search results. AD-based systems benefit from automatic AD group binding; non-AD 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 EMC Documentum or dynamic authentication providers such as CA SiteMinder. The more source systems 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.

Administration and Configuration

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

The figure below shows an administrative screen focused on content connections. From each of the tabs along the top (content, connections, targets, datasets, tasks, and tools), administrators have contextual information and actions.

The screenshot shows the 'Content Management' interface with a navigation bar at the top containing tabs for Content, Connections, Targets, Datasets, Tasks, and Tools. Below the navigation bar is a 'Content Management' header with an 'Add New' dropdown menu. On the right side, there are links for 'Manage Crawls' and 'Content Sources Count:39'. The main area contains a table with the following columns: Actions, Content, Type, Connection, Enable Indexing, and Alerts. The table lists various content sources such as Alfresco, Autonomy EAS, EBIZ Customer, and Oracle eBusiness Suite, along with their respective connection types and indexing status.

| Actions | Content | Type | Connection | Enable Indexing | Alerts |
|---------------------|--------------------------------------|--------------------------|------------------------------|-----------------|--------|
| Test Tasks Metadata | Alfresco | WebService Content | Alfresco | True | |
| Test Tasks Metadata | Autonomy EAS | WebService Content | Autonomy EAS | False | |
| Test Tasks Metadata | Autonomy WorkSite Documents | Website Document | Autonomy Worksite | False | |
| Test Tasks Metadata | EBIZ Customer | Secure SQL Based Content | Oracle e-biz | True | |
| Test Tasks Metadata | EBIZ Customer and Order Info -bad | Secure SQL Based Content | Oracle e-biz | False | |
| Test Tasks Metadata | EBIZ Customer Invoice | Secure SQL Based Content | Oracle e-biz | True | |
| Test Tasks Metadata | EBIZ Employee Detail | Secure SQL Based Content | Oracle e-biz | True | |
| Test Tasks Metadata | EBIZ Item Lot on Hand with Inventory | Secure SQL Based Content | Oracle e-biz | True | |
| Test Tasks Metadata | EMC Documentum Documents | Documentum Document | EMC Documentum Server | True | |
| Test Tasks Metadata | eRoom | WebService Content | eRoom | False | |
| Test Tasks Metadata | eRoom Direct | WebService Content | eRoom Direct | True | |
| Test Tasks Metadata | File System Documents | Secure SQL Based Content | File System | True | |
| Test Tasks Metadata | FileNet P8 | WebService Content | FileNet P8 | False | |
| Test Tasks Metadata | FileNet P8 Direct | WebService Content | FileNet P8 Direct | False | |
| Test Tasks Metadata | HP Trim - 65 | WebService Content | HP Trim - 65 | True | |
| Test Tasks Metadata | HP Trim - 70 | WebService Content | HP Trim - 70 | True | |
| Test Tasks Metadata | Hummingbird DMS Documents | Secure SQL Based Content | Hummingbird DMS | False | |
| Test Tasks Metadata | IBM Content Manager | WebService Content | IBM Content Manager | False | |
| Test Tasks Metadata | IBM Lotus Notes Marketing Content | WebService Content | IBM Lotus Notes DBs | True | |
| Test Tasks Metadata | Legal | Secure SQL Based Content | Legal | True | |
| Test Tasks Metadata | LiveLink | WebService Content | LiveLink | True | |
| Test Tasks Metadata | MS Dynamics CRM Accounts | WebService Content | MS Dynamics CRM | False | |
| Test Tasks | MS Exchange Private Mailboxes | Exchange Private Email | MS Exchange Server | True | |
| Test Tasks Metadata | New Vault Connector | WebService Content | New Vault Connector | False | |
| Test Tasks Metadata | Newsqator News Feeds | WebService Content | Newsqator.com | False | |
| Test Tasks Metadata | Oracle eBusiness Suite Departments | Secure SQL Based Content | Oracle eBusiness Suite | False | |
| Test Tasks Metadata | Oracle eBusiness Suite Employees | Secure SQL Based Content | Oracle eBusiness Suite | False | |
| Test Tasks Metadata | Oracle eBusiness Suite Products | Secure SQL Based Content | Oracle eBusiness Suite | True | |
| Test Tasks Metadata | Oracle eBusiness Suite Projects | Secure SQL Based Content | Oracle eBusiness Suite | False | |
| Test Tasks Metadata | Oracle eBusiness Suite Projects List | Secure SQL Based Content | Oracle eBusiness Suite | False | |
| Test Tasks Metadata | Oracle Portal Engineering Documents | Secure SQL Based Content | Oracle eBusiness Suite | True | |
| Test Tasks Metadata | Oracle Stellent UCM Documents | WebService Content | Oracle Stellent UCM | True | |
| Test Tasks Metadata | Salesforce.com CRM Accounts | WebService Content | SalesForce.Com | False | |
| Test Tasks Metadata | Salesforce.com Opportunities | WebService Content | SalesForce.Com Opportunities | True | |
| Test Tasks Metadata | SAP Customers | WebService Content | SAP | True | |
| Test Tasks Metadata | SAP Invoice Lines | WebService Content | SAP Invoice Lines | False | |
| Test Tasks Metadata | SAP Invoices | WebService Content | SAP Invoices | False | |
| Test Tasks Metadata | Siebel CRM Customers | Secure SQL Based Content | Siebel CRM | True | |
| Test Tasks Metadata | Symantec Evault Archived Mailboxes | WebService Content | Symantec Evault | True | |

Once deployed, crawling is monitored and managed transparently through the familiar SharePoint crawl management tools. Scheduled jobs for common tasks such as security sync, backups, and mailbox management enable straightforward administration.

An integrated Test Bench makes it possible to rapidly deploy, configure, and test connectivity. Administrators can test the output of any defined Content Source, display all of the properties returned for each Item, and provide visibility into performance, security, and metadata contents. It is not necessary to crawl content to use the Test Bench, so it is often used to check a deployment without populating the search index or to troubleshoot indexing without replacing or updating items.

Content Test Bench

FileNet_Content

Handler Class (ServiceSource) found for content type.

Maximum Displayed Results: 10 Use Original Paging Size Include Trace Leave Temp Files Show ACLs Test

Incremental 1 Days Do Real Time Security Trimming Select a user: Run

CrawlStarted:2:09:47 PM For Item:1
 Init Class:2:09:47 PM:SPWorks.Search.Service.ServiceSource
 After Init Class:2:09:49 PM:SPWorks.Search.Service.ServiceSource
 License Status:Pass Item Limit:10000000
 Is Fast:False
 Items Found:11 in 20.0216472 seconds
 Returning Results:
 More Exist:True
 Last Item Update Date:10/13/2011 9:12:23 AM:Zone:Utc
 Full Final Date:12/20/2012 11:10:07 AM:Zone:Utc
 Returning Results:

Item:3657bc36-49af-4f04-a93b-a1b52637f0af:: found in 12.5792096 total seconds
 Times (sec): Retrieve=2.4360455 Datasets=0.0009803 Metadata 0.931285 Security 0.0019606 Unstructured 0.039212 Group Tracking 0.0176454
 FilePath: C:\Program Files\Common Files\Microsoft Shared\Web Server Extensions\14\LOGS\ESC\1_0dfa904e-53c2-4ebf-a64e-c8c441c76e8a.log File Size=7630 File is Temp=True
 isPublic: True
 item extension: p8fn

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. Although SharePoint Search sets up a default mapping between crawled properties and managed properties, it only does so for content in stored in SharePoint. For all other content sources, administrators must do all mapping manually.

Smart Mapping makes this process much simpler, by auto-generating property names and the metadata mapping based on the SP source system schema. It also tracks any manual modifications or overrides and respects these whenever the mapping is refreshed.

Generate Metadata

Handler Class (SPWorks.Search.SharePoint.SharePointDataSet) found for content type.

Auto create managed Property Prefix: ESC_ Generate

Clean SQL:

| Property | Metadata Description | Content Type | Active | Managed | Alerts | Content |
|-----------------------------|----------------------|--------------------|--------|---------|--------|---------|
| delete ESC_ESC_FIRST_NAME | | Text Metadata | Yes | No | | 8 |
| delete ESC_ESC_LAST_NAME | | Text Metadata | Yes | No | | 8 |
| delete ESC_ESC_EMPLOYEE_ID | | Numeric Metadata | Yes | No | | 8 |
| delete ESC_ESC_EMAIL | | Text Metadata | Yes | No | | 8 |
| delete ESC_ESC_PHONE_NUMBER | | Text Metadata | Yes | No | | 8 |
| delete ESC_ESC_HIRE_DATE | | Date/Time Metadata | Yes | No | | 8 |
| delete ESC_ESC_SALARY | | Numeric Metadata | Yes | No | | 8 |

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 Visual Basic syntax, can be applied to security, crawling, and metadata to handle even the most demanding applications.

Text Value* Reference OS Columns with brackets - [COLNAME]

Advanced Scripting Help

```
if HOST.GetStringValue("name").startswith("M") then
    return "Mark"
else
    return HOST.GetStringValue("name")
end if
```

Single Value Append all rows into a single value

Seperator If dataset contains multiple rows they are appended with this separator.

Max Rows* if multiple rows returned for dataset. 0 = no limit

Active Include In Results

Managed Auto Create Managed Property

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 metadata content systems, and it can be filtered to allow for fine granularity when matching the metadata to the data content. This ultimately provides a richer and more accurate search result to the end user.

Dataset connectors essentially look up and join information across systems during the crawling process. For example, a dataset connector can combine customer data across an ERP system and a CRM system. Imagine crawling 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 can use it for refinement and navigation. When exploring market information, the user sees also customers in that specific market segment.

Targets

Targets provide a mechanism to direct content into different places, in addition to search indexing. For example, content can be copied to a SharePoint list, along with its metadata and security settings. Targets support a wide variety of advanced scenarios and are a powerful tool to the solution developer.

With Targets, a synchronization schedule is used to move new content to specific locations. This is analogous to the full crawl and incremental crawls used by search. The Connector Framework pushes content to a target, unlike a connection to SharePoint Search. (SharePoint Search uses a pull mechanism for connectors when crawling content). Pre-built targets include SharePoint Lists and Libraries, the SharePoint User Profile Store, and various non-SharePoint search engines that support a push interface

(including FAST ESP). Custom Targets, fileshare Targets, and database Targets can also be created on demand.

| | |
|----------------------------------|--|
| Target Type | List or Document Library |
| List* | List name to add <input type="text"/> <input type="button" value="Create List"/> <input type="button" value="Create Doc Lib"/> Select list to map <input type="text" value="AList"/> Configuration Options: <input type="checkbox"/> Link Column is Text (use for improper URLs: notes:// mail:// etc) <input type="button" value="configure"/> <input type="button" value="Open"/> <input type="checkbox"/> Create base properties (From, To, Folder) |
| Folder Mapping | No Folder Mapping Pick an active metadata field from the content source to map folders to. The content must be in the form of \fold1\fold2 or /fold1/fold2. We support a maximum of 10 folder levels. |
| File Types to Include | <input type="text" value="pdf,doc,docx,xls,xlsx,vsd,vdx,xml,rtf,txt,zip,wpd,ppt,pptx,rar,gzip,eml,msg,tif,tiff,htm,html"/> Comma separated list of extensions to include. Leave blank to exclude all files. |
| Zipfile Expansion | <input type="checkbox"/> Expand ZIP Contents Only valid for list and not document library. This will expand out the contents of the zip file into separate attachments. |
| Include esbase properties | <input type="checkbox"/> |
| Security Model | No Security |
| Grant Read Role | Full Control Site Collection Role used to apply read rights to the individual items |
| Owner Role | Full Control Site Collection Role used to apply full rights for the owner account |

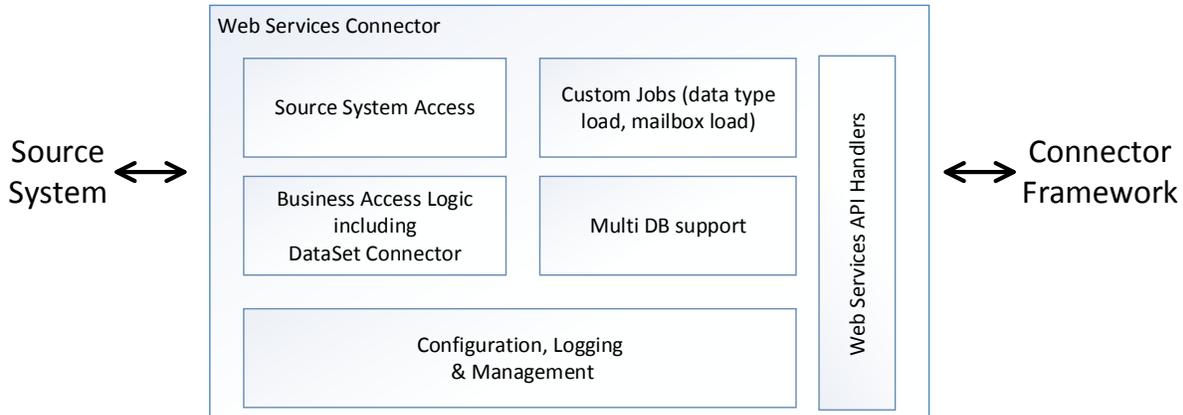
Indexing Connectors

Along with the Connector Framework, BA Insight provides a wide range of Indexing Connectors. Each connector is developed and maintained for a particular source system. There are several types of connectors, including:

SQL-based Connectors: for source systems that expose content via an underlying database. These connectors use a common framework with template-based administrative screens. The SQL calls are available for tailoring, either for performance optimization or to support advanced scenarios.

Embedded Connectors: for specific source systems requiring particularly complex and high performance operation, such as EMC Documentum and Microsoft Exchange.

Web Service Connectors: for source systems that publish APIs for content access. Web services connectors include a number of functions and communicate to the Connector Framework through a published Web Services-based API. The structure of these connectors is shown below.



All connectors share a high throughput, light touch approach to selecting and extracting content. They are all agentless – i.e. they do not require any software to be installed on the source system, and can communicate over a network to remote systems. They need only read access, so there is no risk of compromising source systems.

Many of BA Insight’s Indexing Connectors also can 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, both the file and the metadata are indexed as a single item using an associated crawl.

Supported Source Systems

Connectors are available to over 50 systems of a variety of different types:



CONTENT AND COLLABORATION PLATFORMS

EMC Documentum, EMC eRoom, HP TRIM, IBM FileNet P8, IBM Content Manager, IBM Connections, Objective, OpenText LiveLink/RM, OpenText Hummingbird, Oracle UCM/Stellent, Lotus Notes Databases, Xerox DocuShare, Confluence, Alfresco, Jive, CuadraSTAR



MAILBOX AND ARCHIVING SYSTEMS

Microsoft Exchange, IBM Lotus Notes, Symantec Enterprise Vault (eVault), Symantec eVault File Archive, Microsoft Exchange Public Folders, HP Autonomy Enterprise Archive Solution (EAS, aka Zantaz)



ENTERPRISE PORTALS

IBM Websphere, Oracle WebCenter Interaction (PlumTree)



PRACTICE MANAGEMENT SYSTEMS

Aderant, Autonomy Worksite/iManage, Elite / 3E, KnowledgeMill OnePlace, LegalKEY
NetDocuments, PLC/PracticalLaw, ProLaw, RealPractice



CRM SYSTEMS

Salesforce.com, Microsoft Dynamics CRM, Microsoft Dynamics CRM online, Lexis/Nexis
Interaction CRM, Any SQL based CRM system



DATABASES

Microsoft SQL Server, MySQL, IBM DB2, Oracle



SEARCH AND CLOUD SYSTEMS

SharePoint Online, Google Drive, SharePoint 2013, SharePoint 2010, SharePoint 2007,
FAST Search for SharePoint, Microsoft Search Server, Box, Scopus, PharmaCircle

Creating New Connectors

BA Insight has extensive experience in creating and maintaining indexing connectors, and a proven process for approaching new systems. The Connector Framework provides facilities for testing, troubleshooting, and optimizing content extraction, which makes creating new connectors faster and simpler. Connectors built on this framework also inherit many powerful features such as Targets and Smart Mapping, and present a consistent and effective interface to administrators.

There are two main facilities for creating new connectors. A Universal SQL connector toolkit supports secure indexing for any SQL-based source system with the flexibility to tailor the way database content is composed and transformed into indexed items. Developers can also use the Web Services API to integrate crawling into their system or to create new connectors themselves. BA Insight also provides services to create custom connectors and/or mentor developers who wish to create connectors.

Scalability and Performance

The Connector Framework allows users to process search data from different source content systems and add metadata information at high throughput. Essentially, the retrieved content is passed through with barely any performance impact. It is multi-threaded and scales out so that higher throughput can be

gained with more hardware resources. A throughput of hundreds of documents per second (DPS) can be achieved on relatively modest hardware.

Typically, the bottleneck in enterprise search deployments with BA Insight connectors is not the connectors – it is the source systems themselves. For this reason, there are several mechanisms built into the Connector Framework to optimize the performance of access to sources systems.

For incremental crawls, a powerful facility is used to find new items efficiently. This source system enumeration works even in the absence of efficient change logs on the source system; it speeds incremental crawls dramatically. Another significant source of stress and potential performance issues to the source system is the lookup and translation of security accounts for each indexed item. The Connector Framework successfully eliminates this bottleneck by implementing a user group synch job offline, which performs such user group loading and mapping prior to index time.

The Connector Framework is installed as a SharePoint Service Application, which provides seamless scale-out to additional server nodes within a farm as well as fault-tolerance and backup/restore facilities. Administrators work with a familiar paradigm, using familiar user interfaces and scripts.

Advanced Scenarios

The Connector Framework, Indexing Connectors, and DataSet Connectors are flexible and extensible. They can be configured in a wide variety of ways to meet challenging requirements and handle advanced scenarios. In this paper we can barely scratch the surface, but a few examples should help in understanding the possibilities.

By combining DataSet Connectors with a fuzzy lookup, metadata can be normalized – resulting in cleaner content in the search index and better search results. For example, imagine a project management system hosting information about company projects. Even if the name of the project is misspelled or incomplete, a DataSet Connector can retrieve matching projects so each document indexed from any source includes a managed property containing the official name of the project.

Using this facility, along with a target for the SharePoint User Profile Store, supports a consolidated user profile based on content from a variety of different sources. This improves mysites, people search, and other applications driven from the user profile.

Targets provide a simple content selection and relocation facility, as well as a content migration capability. Combining a SharePoint Library target with workflows and/or the SharePoint Content Organizer supports very flexible content selection, migration, and replication across systems and domains.

Scripting can be used to call remote facilities. Examples include translating information into other languages, calling other systems to provide related information, or finding the top news stories related to an item and populating links on the indexed item so that users can click through to the story.

Summary

Indexing Connectors are harder than they may seem. They require a balance of high performance and light touch, 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.

The BA Insight Connector Framework 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. The Connector Framework is tightly integrated with SharePoint and operates seamlessly with SharePoint Search. It also provides targets to a variety of other systems that range beyond search applications.

About BA Insight

BA Insight's Software Portfolio accelerates and future-proofs the creation of cloud-based and on-premise SharePoint portals for enterprises, transforming how users find information. Our software dramatically reduces the time, cost, effort, and risk of implementing SharePoint portals while providing a greatly improved user experience. Customers use our Applications to improve productivity by finding the right information faster using Visual Refiners, Smart Previews, User-Generated InfoSites, and Matter Comparison; our Classification software to increase findability using auto-tagging, metadata generation, and text analytics; and our pre-built connectors and federation to provide secure connectivity to a wide variety of content systems.

We serve organizations of all sizes around the globe including the Australian Government Department of Defence, CA Technologies, Chevron, Deloitte, Ford Motor Company, Keurig Green Mountain, Pfizer, and Travers Smith. BA Insight is a Microsoft Gold Certified Partner and member of the Microsoft Business-Critical SharePoint Program.

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