BA INSIGHT

Modernizing Enterprise Search Using Hybrid Index and Search Time Merge

How BA Insight SmartHub Brings Web-like Search to the Enterprise

Federation: What is it?

Wikipedia defines federated search as:

“**Federated search** is an information retrieval technology that allows the simultaneous search of multiple searchable resources. A user makes a single query request which is distributed to the search engines, databases or other query engines participating in the federation. The federated search then aggregates the results that are received from the search engines for presentation to the user. Federated search can be used to integrate disparate information resources within a single large organization ("enterprise") or for the entire web.

From a technology perspective, there are two approaches to search labeled as ***index-time merging*** and ***query-time merging***. In general, vendors and analysts in the enterprise search/insight engines market have advocated index-time search and have positioned query time merge as something that can be used but is not recommended. So, the approach has been an “either/or” paradigm where they define use cases in which one is the right solution. They are effectively heavily biased to index-time merge, which in reality is what the vendors provide or at least promote heavily. It creates the view that one approach versus the other needs to be chosen. This school of thinking removes a valuable approach from potential implementation strategies that has been proven in many scenarios to be the correct approach. Single index search is a solid strategy; however, it is not the only viable strategy. The reality is that with the continued growth of data and availability of new, open and flexible search engines such as Azure Cognitive Search, Amazon OpenSearch, and Elasticsearch, along with modern applications with robust APIs, it is possible to have a hybrid approach as opposed to an either/or approach.

Today, the “one-size-fits-all” approach that has been advocated and has worked for the last decade won’t work in many situations. There are times that business requirements are best served using a single-index approach and others that will be best served using a hybrid indexing approach, where the solution combines the index time and query time merging. Our view is that hybrid indexing is becoming the de facto approach for deploying enterprise search.

An interesting paper to read is Google’s research paper on the journey around indexing the world wide web. Abhishek Das and Ankit Jain of Google summarize:

> *Why can we not have multiple indices -- bucketed by rate of refreshing? We can and that is what is standard industry practice.*

What Google establishes as standard industry practice for the web is the concept that search indices should be based on the business drivers and needs of the content that will go within them. Each index serves a purpose and works in tandem with all other indices, also serving specific purposes. Purposes range from speed of refresh, necessary metadata enrichment, geographical distribution, performance spread, etc. Within enterprises, the same reality exists. Each application has its own index for its own purposes, and where possible we should make them work in tandem with each other without requiring the crawling and indexing with the infrastructure and support costs it brings.

Query-Time Federation: Why Such a Bad Reputation?

Historically, organizations that attempted this type of federated search via query-time merge could run into challenges that resulted in poor results. As an example, performance of query results is often raised as an issue of federation. A decade ago, query-time indexing of multiple search indices could and did cause performance and user satisfaction issues. And this still holds true for older applications, but not so for many modern applications, especially search engines like Azure Cognitive Search, Amazon Kendra, Amazon OpenSearch Service, Elasticsearch, Microsoft Search, SharePoint Online SharePoint On-Premise, and Solr.

Outdated federated approaches waited for the “slowest” index and then returned all combined results. Modern federated approaches now return results in managed blocks that are incorporated as they are presented. The “slow” index causing these difficulties can be tuned or its computing environment addressed (or both) to raise performance. Alternatively, asynchronous web technology can interject results as they become available, almost like a real-time dashboard of information. Modern technological approaches have turned a weakness into a strength.

Another challenge that has been overcome is relevancy quality. True relevancy quality can only be achieved at the layer where the user interacts with the search interface - identifying user properties, preferences, and query intent - and then translating that to the relevant search indices that are best suited to answer that specific question. No one relevancy algorithm within a single search index could address those needs. By using a hybrid federated search approach, the search results are always fresh, resulting in improved end user satisfaction and productivity.

Traditional solutions can still be applicable and may be the best solution where simplicity is required. Tabbed, categorized, or bucketed results can still be used to provide users easy to navigate and recognizable interfaces.

Different search engines can, and often do, provide varying levels of query sophistication. In the past, federation technologies at query time were limited by a "least common denominator" approach to align all queries to all search engines with the capabilities of the least capable search engine driving those queries, meaning results returned were controlled by the least capable index. New, sophisticated query parsers, like SmartHub, pass bespoke queries to each search index that are specifically crafted at runtime to optimize and align for each individual index.

The table below provides a list of historical limitations to query-time federated search and describes how SmartHub addresses them:

Historic Objections	Modern Technical Realities
Query-Time Merge introduces complexity, redundancy, and administrative burden.	• Query-Time Merge simplifies the process as you no longer have to crawl systems to keep the central index fresh and up to date. The search results always contain the latest information since the query is searching the original source. • Metadata changes made in the federated system are immediately available in the search solution without having to re-index the new information. Previously, re-indexing may have taken days or weeks. • Query-Time Merge allows you to pinpoint similar documents across sources. This is because the same document can be surfaced from different sources with different metadata without the index thinking they are the same and automatically trimming the duplicates.
Federation prevents correlation of signals.	• These use cases are driven by capabilities such as analytics, learn to rank, document similarity and natural language query, where you can curate information from different sources to get suggestions for relevant documents, all regardless of the source index containing the information.
Query-Time Merge can be brought down by the weakest link.	• Administrators can decide which sources get interleaved versus sources that are surfaced in their own containers. This negates the issue of performance. They have the flexibility to separate sources that are less performant into “side-sources” that are displayed in a separate container and respond in their own timeframe.

Query-Time Merge prevents cross-source relevance calculation.	• Based on previously discussed cross-source correlation of signals (natural language query, learn to rank, etc.), query-time boosting of relevant information does not rely on cross-source relevance. Federated results can identify the best source of information and then rely on the individual relevance returned instead of deciding which document(s) across all sources are the most relevant. • Additionally, the ability to normalize relevancy values returned by the sources allows interleaving the top documents from each source in a consistent and applicable manner. The best results are interleaved in the correct order due to the normalization.

How BAI Makes Query-Time and Index-Time Merge Intelligent with SmartHub

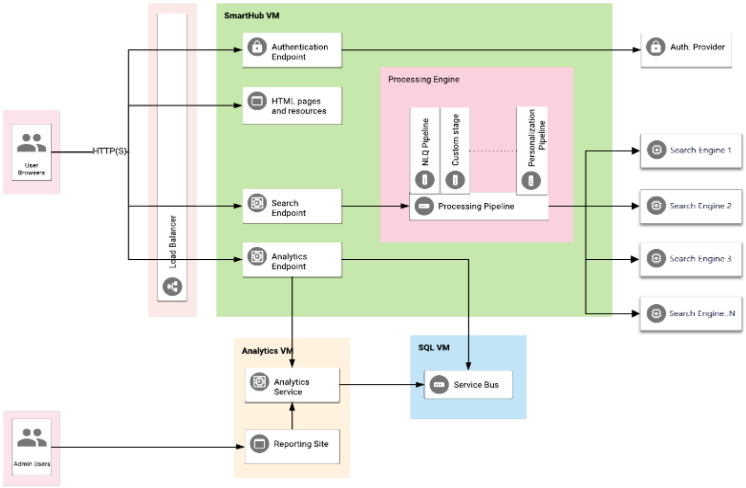
The BA Insight SmartHub Query Engine supports Federated, Single Index Search and Hybrid Search, where the solution is a combination of the two. It transforms a user's query and broadcasts it to a group of disparate systems, each with its own index with the appropriate syntax. It merges the results collected from the indexes, re-sorting them in a succinct and unified format, allowing the user to work within the returned merged result set using various tools. An example of this is sorting or using facets to help the user quickly find information. This powers organizations to go beyond the traditional single index approach, enabling the implementation of powerful enterprise search solutions that can integrate results from multiple search engines into a single, interleaved result. Users can effectively use any technology within the organization to obtain the best search results.

Orchestrating Results Across Search Engines

The BA Insight SmartHub Query Engine provides intelligent brokering capabilities between multiple indices. The supported list includes Amazon Kendra, Amazon OpenSearch Service, Azure Cognitive Search, Elasticsearch, Microsoft Search, SharePoint Online, SharePoint On-Premise, Solr, NetDocuments, and Egnyte. SmartHub can be configured to retrieve results from any combination of the supported search indices via a simple configuration screen.

When the SmartHub Query Engine receives a search query, it transforms and dispatches the query to the appropriate back-ends to obtain the best responses. It recombines the responses collected from various systems into a unified set of results and makes them available to the user via the intuitive search user interface. The high level SmartHub architecture is shown in the diagram below:

SMARTHUB ARCHITECTURE



SmartHub has been designed with flexibility and adaptability in mind. As part of deployment, an enterprise can add its own logic and set of transformations by adding stages to the query and result pipelines, effectively enabling the creation of “smart answers”, similar to internet search engines. Each stage allows a distinct transformation to the query or results to be applied. For example, synonym expansion, stemming, rank boosting, query rewrite (i.e. *white paper federation* could be rewritten as: *“White Paper”* as a phrase and the *Product* metadata must be equal to *SmartHub*, resulting in much higher accuracy), translating results to English, injecting real-time inventory status for a product or part or check-in/check-out status for a document.

SmartHub takes care of interfacing with each index, dispatching appropriately formed queries to each, merging results back together and applying things such as paginating to results. This enables enterprises to focus on understanding end user needs and optimizing their search experiences rather than being concerned about how each search index works.

A key benefit of SmartHub is that it is implemented as an independent search service. Any search optimization done in SmartHub is independent of the presentation layer and can be immediately reused from anywhere- in a web UI, as a Chrome Plug-in, a Teams App, Salesforce App, ServiceNow App, etc.

SmartHub Query Engine in the Real World

There are many use cases in which federated enterprise search can address what a single index approach cannot - at least not as easily or effectively as a federated approach. Of course, there are cases in which a single index is the best approach, but in this paper, we are focusing on federated use cases. Following are some examples:

Purpose-Driven Indices

A pharmaceutical company had purpose-driven separate indices and was seeking a way to deliver a single interface for all use cases. One index was scientific, providing answers to scientists, utilizing things like chemical compound search, related information, etc. It had metadata and relationships that only made sense to the research and development community. Another index was corporate-related, containing policies, HR data, and other intranet-driven content. With SmartHub, this organization was able to bring together both indices, provide their users a single source of truth, leverage personalization to understand users, use AI integrations to determine the intent of their query, and then deliver the correct set of data, regardless of which index it was stored in.

Geographic Data Restrictions

A global organization faced data export restrictions, where data generated in specific countries was not allowed to be stored outside of those countries. However, the expectation was that end users should not have to go to a search site for each individual country to find what they were looking for. They needed a single interface that would search through all organizational data, regardless of which country it was stored. SmartHub provided a single point of entry to each county-specific index. They even went one step further. Through language detection capabilities in SmartHub, they were able to support their multi-lingual end users and multi-lingual documents. SmartHub analyzed the language used in the user’s query and returned the best result in the correct language for their query. This removed the need for separate language-driven interfaces or indices and saved the company even more time.

Supplementing Cloud/Multi-Tenant Applications

The proliferation of multi-tenant cloud applications has created a specific problem for many organizations. These applications become an integral part of their users' daily activities, but because the data is stored within the applications and is managed and controlled by the cloud provider, it limits the ability to relate it to other organizational data. As an example, a law firm that utilizes iManage stores all of their matter-related work product within iManage, but their Matter and Client information is in another system. They want to be able to leverage this Client and Matter data while searching for other related documents. They may want to find example lease agreements for clients over $100m in revenue with offices in the US and EU. With SmartHub, this law firm was able to integrate the iManage cloud index with their own supplemental Client and Matter index to achieve exactly that…and much more.

Personal Use Indices

One large problem with a single index approach often revolves around determining the data to add to the index. Very often, organizations choose not to include "personal" data stores like OneDrive, Box, or Personal Folders. The thinking is that this data is typically only available to one user, so why invest in indexing something with an audience of one? That makes sense, but it also instantly creates a problem for the end users. They now have to go to two places to search - one for their own data and another for corporate data. With SmartHub, the personal use index can be queried at the same time as the corporate index, and a single result set is returned to the user without infrastructure or indexing cost.

Very Large Data Sets

There are several use cases around extremely large data sets that pose a problem for a single index approach:

1. Many organizations have hundreds of terabytes of data archives spread across systems like Box, SharePoint, etc. These data sets are often unnecessary for primary searches, but they do need to be searched for regulatory queries, deep data analysis, or for archival and retrieval purposes. Data that is not queried on a regular basis and is very large in size or document count often places a strain on the central index infrastructure and cost requirements. With a federated approach, the archival systems can be queried directly, when needed, and have results returned along with results from the primary index. This allows organizations to maintain their access and regulatory requirements while removing the complexities and costs associated with indexing this data.

2. Purpose-built systems like NetDocuments in the legal market, where 100+ million documents are stored and actively used by legal professionals. The rate of change is in millions per week. Trying to even incrementally crawl and create a single index will always be behind the information in NetDocuments, resulting in bad search results. Our partnership with NetDocuments addresses this issue. A hybrid approach where we crawl and create a single index for SharePoint or any other search platform and perform query-time merging between that index and NetDocuments index always provides the freshest results.

Summary

Federated search at query time is now as good as federated search at index time. In many cases a hybrid approach is a better approach than either of the two options and should be considered as a way of modernizing enterprise search or as part of digital transformation projects.

About BA Insight

BA Insight provides a connector-based software portfolio that solves internal enterprise search problems by deploying internet-like search within your existing infrastructure. Our technology delivers unparalleled enterprise search experiences that focus on users – delivering highly personalized and relevant results. By bringing the internet to the enterprise, we help maximize the value of intranet and digital transformation projects. Our software connects millions of knowledge seekers to critical information worldwide at customers such as Amgen, Chevron, DLA Piper, Hatch, Duke Energy, Gilead, Nvidia, and Starbucks.

BA Insight is a Microsoft Partner with Gold and Silver certifications, as well as an Elastic Partner. For more information, please visit our website at www.BAinsight.com, email us at sales@BAinsight.com, call us at +1.339.368.7234, or follow us at @BAinsight.