Query

WHITEPAPER

Evaluating Federated Search for Security

What to expect from an Open Federated Search for Security Solution

Table of Contents

Introduction

Security relevant data across cloud, SaaS, and on-prem solutions has dramatically increased in recent years. According to a Panaseer survey of 200 CISOs[1], more than 50% of security teams use more than fifty tools within their operation, and 70% use manual data collection, likely because centralizing this data is increasingly cost-prohibitive. As a result, security teams are becoming overwhelmed with accessing, searching, and evaluating the data stored on many distributed systems.

Open federated search allows users to search data from its source without first centralizing the data, or the need to write complex search queries in many different program languages. This capability is driving considerable interest in open federated search for security in order to expand data visibility across security operations while controlling data related costs. However, because this is a newer concept, there is ambiguity regarding what open federated search is and is not. This white paper is intended to provide guidance to organizations evaluating federated search for security solutions and clarify the definitions of terms, features, and functionalities associated with open federated search for security solutions.

Audience

This document is intended for knowledgeable security professionals: CISOs, Security Analysts, Security Architects, Security Engineers, or others exploring federated search for security solutions with goals to simplify access to siloed security data, reduce costs associated with centralizing data, and/or reduce the time, costs, and errors associated with manually pivoting across multiple data sources for answers to security related questions.

Goal

This white paper is intended to provide guidance to organizations evaluating federated search for security solutions and clarify the definitions of terms, features, and functionalities associated with open federated search for security solutions.

Understanding the Terminology

In a nascent category of products, terminology can be poorly defined or overlapping. Here is our attempt to clarify the various terms used to describe security search, and provide insight into how these have evolved over time.

The Evolution of Search

As networks and systems have continued to grow more complex, cybersecurity professionals have been increasingly challenged with maintaining full visibility into their environment. **Quickly finding, accessing, and evaluating security relevant data is essential for the success of security teams**, underscoring the criticality of using the right type of search.

Originally, with most enterprise software running out of a single server (including the first SIEMs), Apache Lucene based indexing for keyword searches was widely adopted. Then data grew beyond the scalability of one server, leading to Solr and elasticsearch — built on top of Lucene — in the open-source world. This system of indexing internal content and searching it within the organization would be termed Enterprise Search.

As professionals realized the need to search a variety of content outside of their indexed internal content, **Unified Search** and **Universal Search** were born. Essentially add-on features to enterprise search, these could search a larger variety of media and content and bring results from outside the organization's boundaries in addition to internal text based search, i.e. from the Internet.

In the 2010s, enterprise search hit its limitations. Growing enterprise IT footprints caused the amount of data available to grow exponentially. Enterprise Search vendors started to have multiple instances of their indexers, often breaking by location, business department, data sources, etc. Distributed Search emerged as a way to search across those multiple installations of the vendor's homogeneous search-engine stack.

As cybersecurity tools proliferated, the need to search across multiple products became essential. This led to the creation of "federated search" across products, but more specifically **Closed Federated Search**, as it still only searched through the data across that vendor's technologies.

Enterprises of 2023 no longer have data in a single vendor's products, which exposes the limitations of closed federated search. Organizations have heterogeneous data in different databases, search indexes, files, directory systems, cloud storage blobs, CMDBs, 3rd-party SaaS (exposed via APIs), etc. This heterogeneous and "everywhere" nature of enterprise data requires **Open Federated Search**, which allows querying data where it lives, without indexing it into a single vendor's stack.

Open Federated Search as a solution is focused on integrating results from multiple sources into a single search interface. However, keyword or text based searching does not provide enough context to do an acceptable job of combining results from heterogeneous vendors' data.

The next step in search is for the solution to understand the data schema, so it can automatically look up relevant context, perform cross-platform joins, and visually present the data objects with relevant contextual activity. Data schemas are different in different industry verticals, but leads to **Open Federated Search for Security** for the purposes of this document.

The following terms are defined in the order of its evolution.

1. Enterprise Search

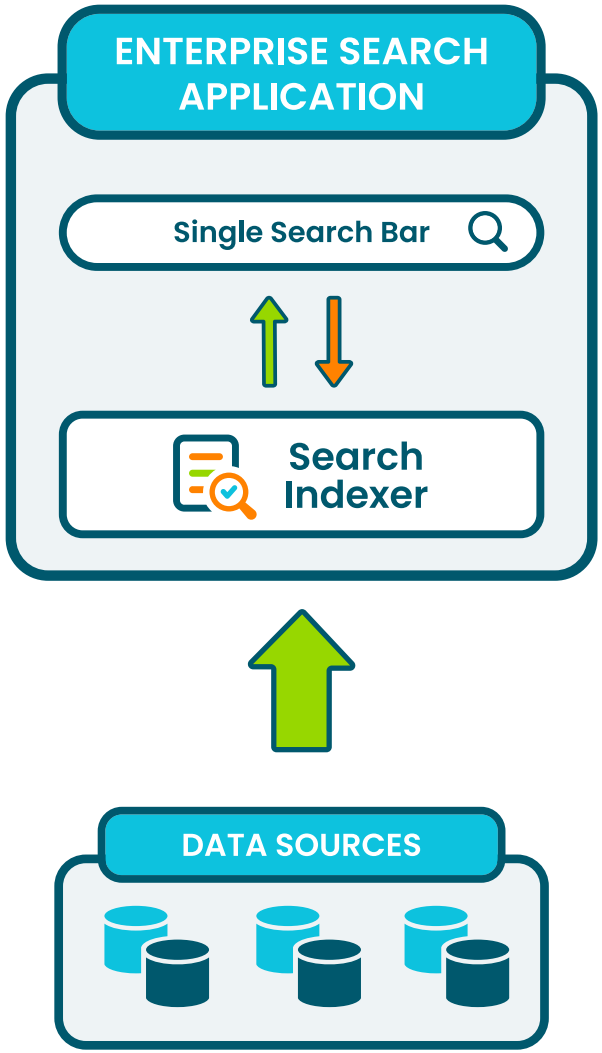
"Enterprise search is the practice of making content from multiple enterprise-type sources, such as databases and intranets, searchable to a defined audience."
- Enterprise search - Wikipedia[2]

What it is:

- The most basic way to search the enterprise's internal data
- One vendor's technology stack
- Data needs to be indexed into that vendor's search index

What it is not:

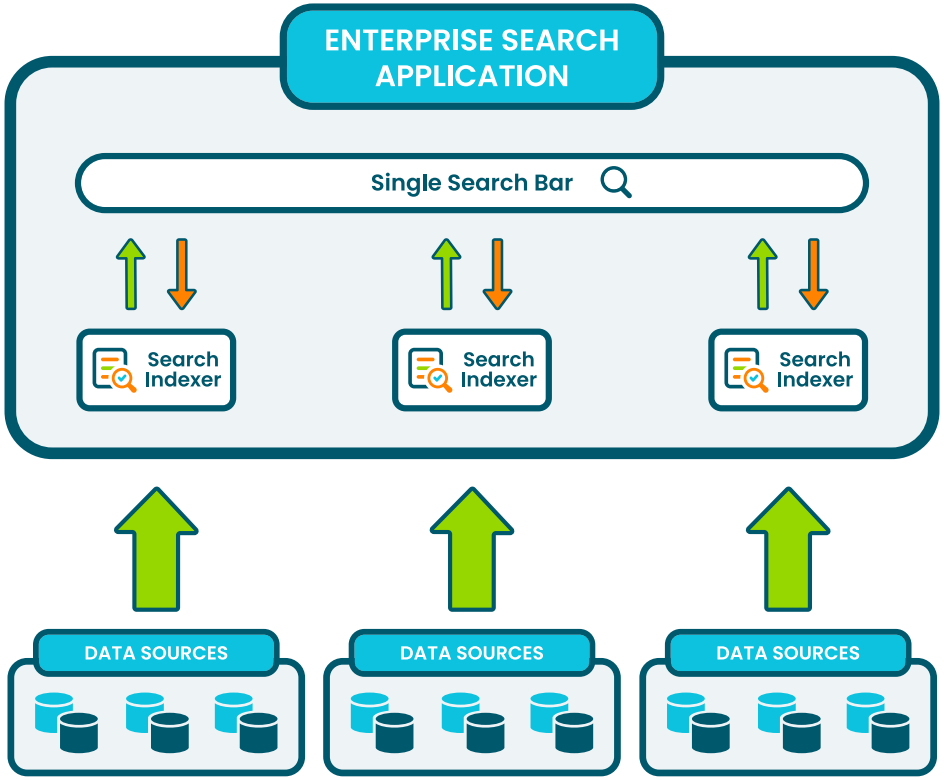
- Federated search, because it cannot take the user's input and search live from data stored or indexed external to it in other platforms.



2. Distributed Search

"A distributed search engine is a search engine where there is no central server. Unlike traditional centralized search engines, work such as crawling, data mining, indexing, and query processing is distributed among several peers in a decentralized manner where there is no single point of control."
- Distributed search engine - Wikipedia[3]

Distributed search comes from one search engine stack working in collaboration with multiple of its nodes. It is essentially a search across multiple instances of that vendor's Enterprise Search technology.



What it is:

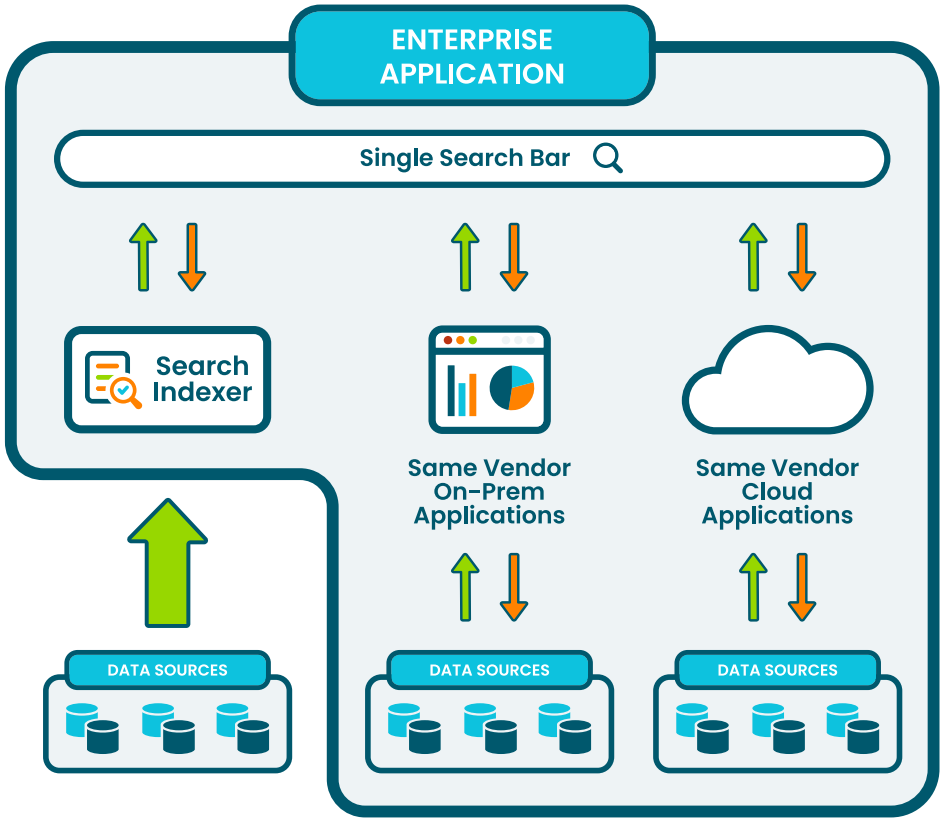
- One vendor's technology stack across homogeneous instances.
- Data needs to be indexed into that vendor's stack.

What it is not:

- Federated search, because it cannot take the user's input and search live from data stored or indexed external to it in other platforms.

3. Closed Federated Search

Closed federated search, or single vendor federated search, allows security teams to execute a unified search across a single vendor's toolsets. It is most often seen today in some SIEM solutions that have created a search bridge between their cloud and multiple on-premise instances. This enables security teams to search on security event data across their cloud-to-cloud, on-premises-to-cloud, and on-premises-to-on-premises instances of that single vendor's solution.



What it is:

- Relies on the traditional approach of moving and centralizing the organization's data within one vendor's solutions.
- Requires security teams to pivot and individually search in other relevant environments, such as AWS and Elastic, and in other vendor data sources.

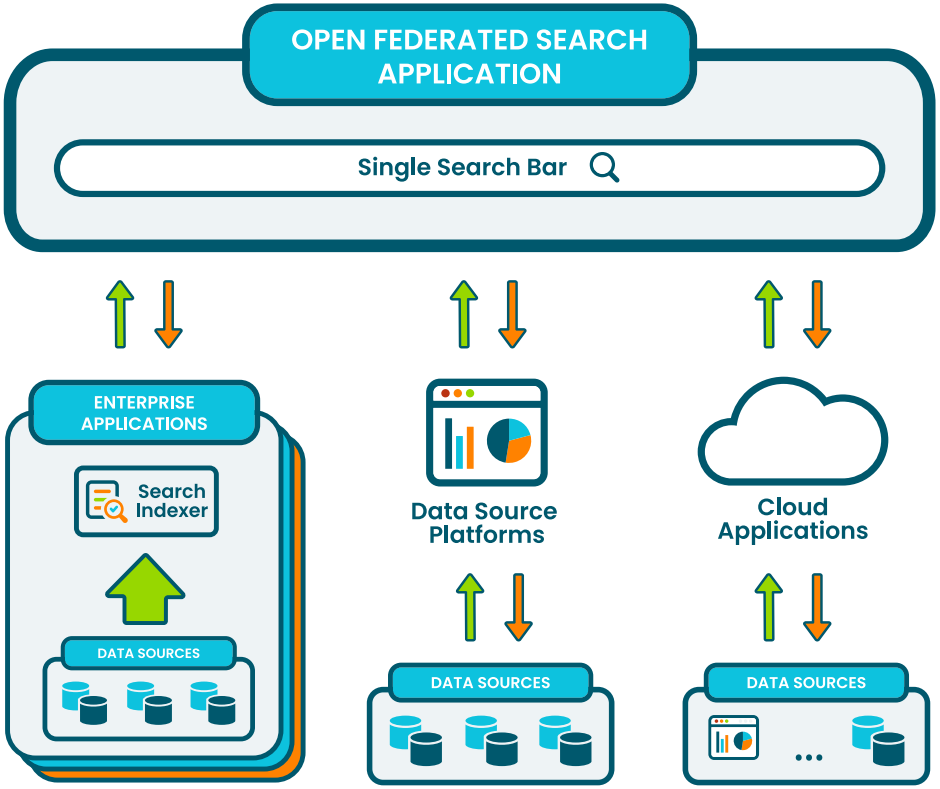
What it is not:

- Vendor agnostic
- Transportable
- Open federated search

4. Open Federated Search

While some vendors freely use “Federated Search” to describe their Closed Federated Search systems, an open federated search system is:

“Federated search retrieves information from a variety of sources via a search application built on top of one or more search engines. 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.”
- Federated search - Wikipedia[4]



What it is:

- Integrates seamlessly across other vendors’ search engines, databases, query APIs, and storage systems.
- Allows customers to query from any data sources they choose.
- Vender agnostic

What it is not:

- Searching only across a single vendors’ products/ecosystem.
- Does not index or store data. Rather it leverages indexes and data stored in 3rd-party products/systems.

5. Open Federated Search for Security

Open Federated Search for Security understands the object context of the search run by the security analyst, such as investigating a file or a user's activity. With knowledge of the cybersecurity data model, it can automatically run focused queries and follow-up queries to show the information the analyst needs for their investigation.

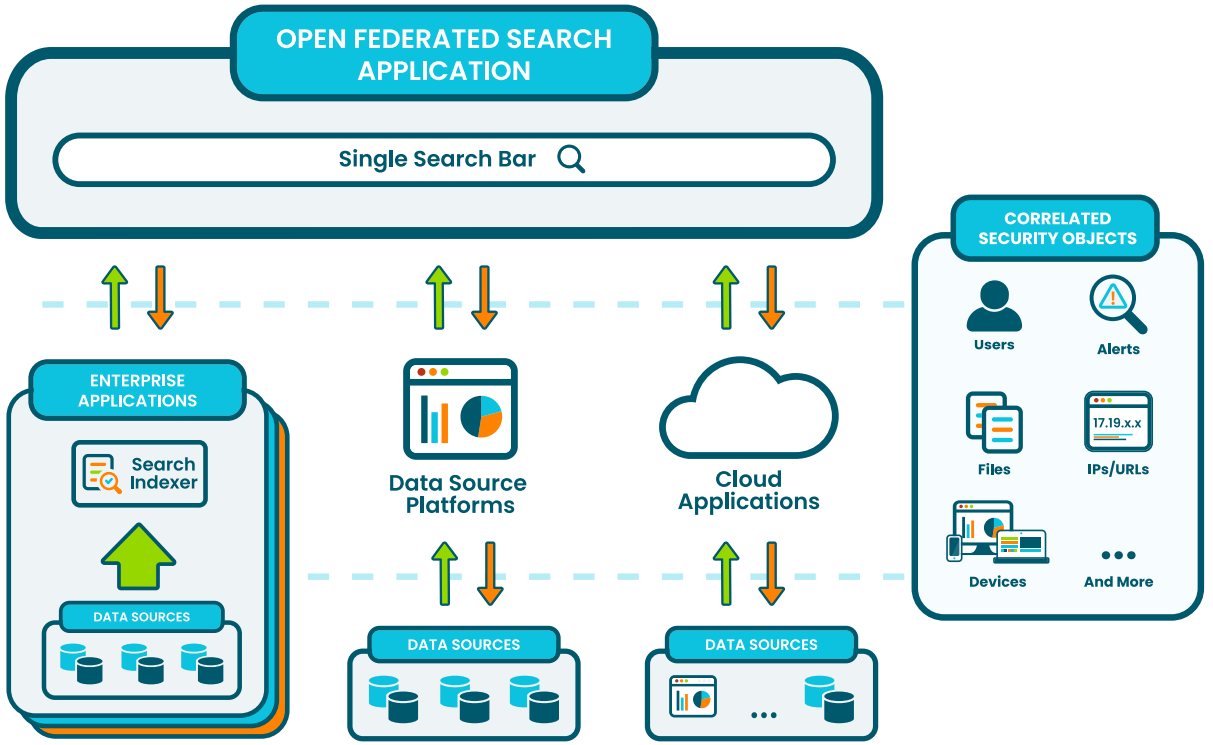
Open Federated Search for Security "retrieves information from across vendor solutions and environments. It uses API integrations with third parties to perform a unified search across the data sources that are participating in the federation, and it does this without requiring data transfer or centralization. This approach also provides the flexibility to choose and integrate the best-of-breed security solutions vs having a single-vendor lock-in." - Query[5]

What it is:

- Understands cybersecurity entities and can correlate information across data sources.
- Can normalize from the vendor's native format to an open cybersecurity schema like OCSF.
- Can automatically initiate multiple followup and lookup queries to complete context.

What it is not:

- Raw text match results. While it does support that, it goes beyond to understand the result and perform relevant follow ups.
- Does not index or store data. Rather it leverages indexes and data stored in 3rd-party products/systems.



What to look for in an Open Federated Search for Security Solution

When evaluating vendors for federated search for security, you have multiple categories of features to consider. **Each organization will have its own hierarchy of needs regarding their federated search solution,** including security, scalability, and ease of use.

Before you are able to evaluate multiple vendors, you should understand **the goal(s) of your organization** (and the challenges of your current situation), **its environment** (including all data storing technologies), and **any specific requirements** to your organization.

The following questions and bullets outline what you can expect from a true open federated search for security solution. The boxes are designed to act as a checklist throughout your evaluation process.

Query

Where can it search?

- **Across relevant cybersecurity data sources, including:**
  - **Security Technologies**
    - SIEM
    - EDR
    - Email Protection
    - Threat-Intel
  - **IT and Business Technologies**
    - Ticketing
    - Asset Inventory
    - Identity Stored
    - Human Resources
    - Business Applications
    - Custom Developed Applications
- **Across your infrastructure, including:**
  - Public cloud
  - Private cloud
  - 3rd-party SaaS
  - On-prem data sources
- **Across vendors with no vendor restrictions**

How does it search?

- **Able to run multiple parallel simultaneous queries:**
  - Parallel simultaneous queries across all data sources AND
  - Also into any given data lake storing data from multiple sources
- **Beyond vanilla text-based search without standard cybersecurity entities like files, users, devices, and IPs. It should understand:**
  - object schema and
  - any related contextual activity
- **Able to understand cybersecurity data, such as:**
  - **Entities**
    - User
    - IP
    - File
    - Device
    - Domain
    - Other
  - **Contextual relationships between entities**
  - **Security significance**
- **Able to automate dependency lookups to create the relevant cybersecurity context for the investigated entity**
- **Able to merge cross-platform data to a common cybersecurity schema**

User Interface: What does it provide the analyst?

- **Provide cybersecurity use-case based context construction to aid analysts investigation**
- **Graphically visualize cybersecurity entity relationships and lead the analyst to understand and navigate**
- **Allow the analyst to pivot and explore from initial results to take the investigation in a desired direction**
- **Provide analysts a full end-user solution (not a building block that needs engineering)**
  - Collaboration features to allow users to share content like saved searches
  - Rich visualizations for context
- **Allow the analyst to do both easy google-like searching and also allow building a complex conditional query logic**

Is it secure?

- **SOC2 Certified**
- **Multi-factor User Authentication**
- **Role Based Access Controls (RBAC)**
- **Single-Sign On support**
- **Integrations' credentials stored in vault**
- **Encrypted and secure TLS connection to data sources**
- **Does not store your data**

Summary

Evaluating federated search for security solutions is a multifaceted process. When researching vendors, it is important to **clarify ambiguity regarding terms, features, and functionalities**. Ensure the goals of your security organization align with the capabilities of the federated search for security provider by asking the specifics within **“Where can it search?”**, **“How does it search?”**, **“What does it provide my analysts?”**, and **“Is it secure?”**

Query: Making open federated search for security a reality

Query aims to deliver visibility into all relevant data for security teams. We provide a **federated search solution** that allows operators to **access data at the source** and in your data lakes, creating opportunities for more nimble and cost efficient data storage architectures.

Our customers are using Query to expand visibility for security investigations, threat hunting, and incident response. They are drastically **reducing the time and complexity** of repetitive search tasks and **improving outcomes for investigations.** Expose your security data with Query.

Learn more

Ready to **expedite your security investigations** with open federated search for security?

For more information visit:
www.query.ai

[1] *Panaseer Survey of 200 CSOs - https://panaseer.com/reports-papers/report/cyber-insurance-trends/*
[2] *Enterprise search, Wikipedia - https://en.wikipedia.org/wiki/Enterprise_search*
[3] *Distributed search engine, Wikipedia - https://en.wikipedia.org/wiki/Distributed_search_engine*
[4] *Federated search, Wikipedia - https://en.wikipedia.org/wiki/Federated_search*
[5] *Query - https://www.query.ai/wp-content/uploads/2023/04/QWP-001_Federated-Search-for-Security-1.pdf*