Building a Data-centric DoD Starts with Search

BY WOODY WALTON

The Defense Department (DoD) is continuing to build out a truly data-centric approach based on the DoD Data Strategy, presenting new opportunities for transforming the way data is collected, analyzed and leveraged.

On a global scale, multi-domain, multi and hybrid cloud deployments need to continue to interoperate with legacy systems, even while the framework is updated and enhanced. Siloed data—all too often duplicated across systems—is essential, but difficult to extract, combine and analyze. Decisions that need to be made in an instant require answers in real time, but existing big data systems are unable to return queries quickly enough for real-time analytics. And with growing data being queried by more connected users than ever before, it's getting increasingly challenging to maintain fast reaction times.

The DoD Data Strategy aims to remedy the core deficiencies inherent in this situation with a blueprint for how data should be managed and accessed by the services, ensuring that trusted information gets to the right destinations at the right time. The challenge now is to identify and implement solutions that can deliver answers to critical questions using all the relevant data in time for it to be useful.

A search-powered data platform that's built for instant availability, distributed operations, true interoperability and uncompromising security offers exactly what's needed to treat data as a truly strategic asset in the fight against the adversary.

Near-instant availability

Information that arrives too late is the same as not having arrived at all. Queries that take hours to run have the effect of limiting imagination and restrict the ability to easily run hypothetical scenarios. The ideal solution returns answers in seconds, even when searching vast stores of data, which allows for pivoting investigations at the speed of thought.

Beyond typical web searches (where a question returns links to thousands of sources), enterprise-wide search enables greater understanding of a situation or requirement. The ability to redraw a map as it is dragged across a screen, update ETAs based on constantly changing weather and terrain information, or analyze the source of cyber attacks as they occur, are all dependent on near-real time, iterative search results.

Distributed operations

The first step of any data operation, and often the difference between success and failure, is a search for all potentially relevant data. Getting the full data picture across globally distributed systems has been a giant hurdle for running large-scale trend analysis and machine learning, much more so for achieving real-time situational awareness. Traditional approaches involve duplicating data to monolithic, centralized repositories so all the information can be analyzed at once. Transporting the data to centralized repositories is slow and inefficient. To house that amount of information, the technologies that must be used by big data lakes cannot provide the analytic speed to run real-time analytics. And the larger the big data lakes get, the slower they become.

It's long been a dream to overcome these barriers, and it turns out the simplest approach is also the most effective: index the data at or near wherever it is generated and access it remotely. A search platform that can connect and operate across globally distributed nodes makes this dream a reality. That distributed search ability also enables faster innovation and adoption of technologies like large-scale machine learning and artificial intelligence, which rely on building and testing against huge datasets to refine models. A speed layer over all data lets that modeling process occur much faster, and a single distributed data layer gives those models a platform to be applied universally.

True interoperability

But there's another issue: the lack of consistency in data formats presents a huge barrier to multi-domain operations, impacting joint missions.

DoD information is generated and stored in a dizzying array of formats across disconnected proprietary systems, which limits the ability to include a large proportion of data when searching for answers, and leads to an incomplete data picture when making decisions. Additionally, each branch of service maintains its own processes and policies across a spectrum of disparate systems, making retrieval of all useful data from all relevant sources nearly impossible.

A comprehensive solution would also use a common data schema to draw from all relevant sources—so that no matter the origination or location, the data all speaks the same language. With a standards-based common schema—such as the one being developed by the DoD's Schema One program—diverse datasets can be accessed and analyzed no matter the original format.

Uncompromising, yet flexible security

There's a natural concern that stems from ever-increasing access to data: security and privacy must be protected, while also allowing users with a legitimate need and clearance to access information. In order for a distributed, search-based platform to be considered secure, it must have fully integrated security that can be defined and applied by dynamic policies that can be changed on the fly as needed to support changing data access requirements.

Make data a force multiplier

Along with the sheer speed of access a search platform provides, a comprehensive, unified view of all relevant data gives a clearer picture of the real-time situation. A single source of truth for all organizational data streamlines operations and delivers insights and answers that are both relevant and timely to all levels of the organization—amplifying situational awareness, strengthening cybersecurity, and empowering automation and innovation.

Transforming data from a mass of disconnected bits and bytes to information that empowers organizational change, rapid decision making and mission readiness is an exciting challenge. With a fully distributed data access speed layer, the DoD can harness the power of search to make data a force multiplier—delivering actionable information wherever needed, at mission speed.

For more information
visit elastic.co/federal
or email federal@elastic.co

Woody Walton is a Principal Solution Architect at Elastic.

Powering the Eight Guiding Principles with Elastic

The DoD Data Strategy cites the need for "enterprise data management to ensure that trusted, critical data is widely available." This can be related directly to the famous OODA Loop (Observe, Orient, Decide, Act): events have to be Observed in real-time so that Orientation can be continually updated, so that Decisions and Actions are made based on all available information. Bringing knowledge to the point of need, precisely when needed, enables greater readiness and responsiveness. Data access is essential to speeding up the OODA Loop, making it a tactical advantage, and crucial to empowering concepts such as JADC^2 (Joint All-Domain Command and Control).

Elastic is a search-powered distributed data platform that includes a suite of enterprise search, observability and security solutions built on a single technology stack. It is built to ingest and query massive amounts of diverse datasets in near-real time across a distributed environment. Already trusted and used by every service branch of the military as well as many Fortune 500 companies, Elastic supports all of the DoD Data Strategy's 8 Guiding Principles:

Data is a strategic asset - With Elastic, globally distributed organizations like the DoD can use data much more effectively because they can bring their questions directly to the data, collecting it in place or regionally while allowing access from anywhere in near-real time. Security controls remain intact, and data can be cost-effectively leveraged throughout the full retention life cycle.

Collective data stewardship - Schema One (which is based on the Elastic Common Schema) enables data to be used interoperably across all data sources and services. When managing data access and retention, stewards, custodians and managers only need to define the policies related to their function and let the Elastic platform move the data through each lifecycle phase automatically.

Data ethics - Elastic supports monitoring and auditing data access and usage to ensure ethical oversight is maintained. It is much easier to maintain, monitor and apply standards to a single data system than the multitude of disparate systems currently in use.

Data collection - Elastic includes tools to capture data at the point of creation, integrates seamlessly with other tools, and can automate the creation of pedigree and attribution tags. Subsequently combined or created products can then have the appropriate tags applied to them for data management and assurance concerns.

Enterprise-wide data access and availability - Elastic has the ability to create a global data mesh where data from all sources is immediately available via a single, yet distributed, search-based platform. Access controls can be dynamically created, updated or removed at a granular level to support standing or ad-hoc missions.

Data for artificial intelligence training - Elastic's ability to create a speed layer over all organizational data sources simplifies and accelerates the most tedious aspects of creating, training, testing, and applying AI and supervised ML. Those models can then be applied to the production data stored in Elastic. Fully integrated unsupervised and supervised ML capabilities for building custom analytics are built into Elastic, as well as hundreds of pre-built detection rules and ML jobs for solution-specific activities related to security and observability.

Data fit for purpose - At its core, Elastic is an extremely flexible general-purpose data access layer; it provides Data-as-a-Service. This means that Elastic does not shoehorn data into a particular solution or viewpoint (e.g., everything does not suddenly take on the characteristics of a security problem in order to fit into a SIEM solution). When all potentially relevant data is made available together, data is not only fit for purpose, it also becomes a mission enabler.

Design for compliance - With all data available and monitored within a single platform, compliance and auditing becomes much easier and truly comprehensive. Compliance becomes yet another function that can leverage the common data platform to become more effective, standardized and automated.