

Moving at the Speed of the Mission with Data Fabric

As the DoD presses forward with Joint All-Domain Command and Control (JADC2) programs and architectures the Air Force is working to stand up technology centers that will not only allow for the sharing of data but for the sharing of data in motion. Our warfighters and peacekeepers need real time data to make decisions in the field and the traditional database structure, modernized with data lakes cannot operate at the speed of the mission.

What is Data Fabric?

Gartner defines data fabric as a design concept that serves as an integrated layer (fabric) of data and connecting processes. This layer utilizes continuous analytics over existing, discoverable, and inferred metadata assets to support the design, deployment, and utilization of integrated and reusable data across all environments. It continuously identifies and connects data from

disparate applications to discover unique, business-relevant relationships between the available data points. A Data Fabric approach enables software factories that practice continuous integration/continuous delivery (CI/CD) to deploy faster and automate version control with their internal and external data integrations.

Why do we need it?

There is an acute need to avoid the same mistakes of past data modernizations efforts. Attempting to create yet another centralized data lake with all an agency's data would just add to data sprawl as existing databases aren't likely to be going anywhere. More importantly, data at rest in a datastore isn't useful when it isn't being actively queried. Instead, as DoD has shown with JADC2, there is a real need to start thinking about creating a central connective tissue that conducts data across organizations. Doing so allows for data to stay in motion, becoming available across the enterprise as it is created and available to those who need it precisely when they need it.

"Data-in-motion is really about inverting the longstanding dynamic of data at rest. Rather than storing the data away in

silos where it's static and asking retroactive questions, what you want to do is publish the data as a stream and constantly deliver it for real-time analysis," said Will LaForest, public sector chief technology officer at Confluent.

By removing the storage resting place, you are decoupling the producers of the data from the consumers. Data Fabric provides the connective tissue of decoupling all the different actors within an organization, so they can all produce and consume independently. This is key to making solutions scalable across a large organization – even one as big as the DoD.

How do we weave data into fabric?

A U.S. Air Force Small Business Innovation Research (SBIR) Phase III contract was recently issued to focus on deploying a Data Fabric in the Air Force. Run out of the Chief Architect Office (CAO) of the Department of the Air Force (DAF), this project aims to invigorate innovation for data centralization using event streaming architecture and integrated solutions for easier data analysis.

The CAO mission objective is to ease data visibility with an enterprise data architecture prototype that allows users to access previously siloed data, platforms, or producers by changing users' interaction with data. A tactical Data Fabric will become a global "source-of-truth" that will make it easier to integrate disparate applications by unlocking mission silos and sharing data in real-time between analytics and mission systems.

The CAO solution is utilizing Kafka, an open source distributed event streaming platform capable of handling trillions of events a day. Kafka's enterprise scalability has made it the open-source industry standard in not just messaging, but event streaming, allowing for real-time processing, curation, and transformation of the data in motion. To support the enterprise use of Kafka, they are implementing Confluent, engineered and released by the original creators of Kafka. Confluent is the mission-ready form of Kafka with significant enhancements for developers, operations and administrators along with critical security features required by Federal programs and supported 24/7.

Turning Data Fabric into a full outfit

The Confluent Platform provides security, encryption, and monitoring capabilities including Confluent Control Center, Role-Based and Policy-Based Access Control, and FIPS 140-2 Compliance. These capabilities allow for the handling of the most sensitive data sets for critical mission operations while preserving the speed, flexibility, and scalability of Kafka.

Connectors, pre-delivered integration points to external sources of data, are key to the functioning of a data fabric. Confluent provides and supports a large collection of connectors to make integration with the data-fabric a low-no code endeavor and accelerate mission data integration from existing and emerging programs. These connectors, unlike legacy systems dependent on slower transactional REST APIs, allow for real-time streaming of data and avoid time consuming and costly creation of custom data endpoints.

A Fabric Future

A data fabric architecture meets all eight guiding principles in the DoD Data Strategy from viewing data as a strategic asset to collective stewardship and enterprise access to being designed for compliance. By "dressing" DoD systems in data fabric, our military can better utilize data to meet the vision of JADC2 and share information at speed and scale for operational advantage and increased efficiency.

For more information on Confluent's Government solutions, please visit confluent.io/government