Confluent Enabling Data in Motion for DOD

The acquisition and use of data in DOD is more important than ever. In addition to the traditional sources, there is an explosion of additional data coming from an exponential increase of new mobile devices and sensors. The warfighter needs capabilities to leverage all this data in a fast-evolving, constantly changing environment. Managing data in the DOD's highly-distributed, global operational environment—often working in remote areas with limited or even no bandwidth—comes with its challenges.

The traditional approach of simply accumulating data and knowledge into data stores where it sits at rest, passively being queried, is no longer tenable. Data needs to be handled "in motion" as it happens, it needs to be immediately brought to those who need to act or glean insight. The technical term for this is called event streaming.

Real-time streaming of data to warfighters is complicated by the sheer size of the DOD's networks and its continuous change and constant expansion through the addition of new mobile devices and the Internet of Things (IoT). Maintaining the speed, flexibility and security necessary in a fast-moving agile environment made up of a wide range of data sources and formats requires a readily scalable, flexible, durable and secure platform that supports the creation, delivery, use and repackaging of applications and microservices across the enterprise.

In industry, Apache Kafka has become the go-to technology for event streaming and particularly for requirements involving edge environments and the IoT. However, on an enterprise scale with DOD's security requirements, occasional low-bandwidth realities and global reach, a more comprehensive platform is needed, which is why the DOD and other organizations have turned to Confluent.

Out to the Edge and Back

Living on the edge, in terms of communication, is nothing new for the DOD. Global operations have long been a way of life for the department. "DOD has been dealing with 'edge computing' challenges since before the term became popular in industry." said former USCYBERCOM Chief of Staff Joseph Brendler, Major General, US Army (Retired). "Once, it was sufficient to say 'tactical environment' to convey an understanding of the extreme difficulty of providing reliable communications in those environments and thus, of the challenge of assuring the performance and availability of IT systems on which modern operations now routinely depend." But the steady transition to a software-based, cloud enabled military presents new sets of difficulties, particularly in remote areas with intermittent connectivity, limited bandwidth, and/or high latency.

"The deployed forces can't always use public communications infrastructure. Sometimes it is destroyed by friendly or enemy forces to eliminate the advantage its availability might provide to the other side. Our forces often times only have the communications systems they deploy with -- including satellite and line-of-sight radios, etc. that must be shared among all deploying elements, according to priorities. Thus, some forces may have relatively little such support." Major General Brendler said.

The DOD, of course, also has strict requirements on security. The private sector also has security requirements, which can vary by country or region, but they also have a certain amount of freedom, for the most part. "Tech companies can take the data, correlate it and use it how they like," said Kai Waehner, technology evangelist and edge specialist at Confluent. "In government, you have very specific requirements for encryption and end-to-end security," as

well as a host of compliance requirements.

In a distributed, bandwidth-constrained environment, units need to be able to collect data, store it securely and replicate it to the cloud at a later date. "So that, for example, you can secure something at the edge, and encrypt it, and replicate it in an encrypted way to the data center in the cloud," Waehner said.

Flexibility at Scale

Like tens of thousands of organizations around the world, the DOD has adopted Apache Kafka as part of its product stack. The open-source event-driven streaming platform connects data sources while decoupling applications from the sources, allowing for asynchronous data exchanges between processes, application and servers.

But although Kafka can process up to trillions of events a day, it does have its limits when it comes to security and operations.

Confluent, founded by Kafka's creators, is enterprise-grade Kafka, enabling large-scale mission operations with a secure, high-speed, cloud native back plane that integrates microservices in an event driven fashion. It can be deployed on bare metal, virtual machines or on Kubernetes pods, automated via Confluent's Operator, or provided as a managed service in the cloud. Confluent also provides out of the box integration with more than 100 supported and Confluent-verified connectors already implemented. In DOD environments, it brings flexibility and security to support highly distributed, complex operations, including in areas with sporadic connectivity. "Kafka alone gives you no way to get the data from an edge instance to any sort of central location," said Will LaForest, Public Sector CTO, Confluent. For that matter, "where the bandwidth is extremely constrained, you don't *want* to send everything from the edge to a central location."

Confluent's ksqlDB allows streams of data to be processed in motion so that organizations can route and filter data, find any anomalies and identify what's important in real time before that data is sent to a central location. This enables DOD entities to take action based on data at the edge as well as cost optimize the dissemination of said data. Along with security, flexibility and durability, Confluent also seamlessly supplies two essential elements of event streaming: the means of getting data to the cloud and the means of picking which data to replicate in the cloud. In all, it can help the DOD toward realizing the concept of a global data fabric.

The DOD is also employing Confluent as part of its move toward agile DevSecOps, a larger initiative that has embraced DevSecOps, CI/CD and Continuous Authority to Operate (CATO) in order to improve the service's agility and responsiveness to ever-changing circumstances. FeedOne, for example, is the canonical event streaming service for PlatformOne and the Air Force and is built around Confluent.

The solution gives the DOD flexibility with merging data sources and formats at speed, while scaling easily to accommodate the rapid growth in data throughput, and connecting data sources and targets with low- or no-code requirements. Automation is another feature that's essential to a dispersed, constantly changing and growing environment, when changes are being pushed out to thousands of devices and edge locations where software engineers are in short supply, LaForest said.

Confluent's highly efficient persistence delivers durability and resilience even in settings with low and/or intermittent bandwidth, as well as in the face of outages involving entire data centers.

And most importantly for the DOD, Confluent provides a full array of security controls, even for sensitive data. Having already been through multiple ATO processes with DOD, the solution is FIPS 140-2 compliant, encrypts authentication traffic and provides Role-Based Access Control which integrates with identity management tools such as Active Directory, or Lightweight Directory Access Protocol (LDAP). It has been through the DOD's Security Technical Implementation Guides (STIGs) process and has been deployed on classified networks and in disconnected, intermittent and limited (DIL) environments.

Confluent also allows the DOD to take an iterative approach to implementing Kafka, which is advisable with almost any new technology. "Start small, and pay attention to lessons learned," Waehner said, a point echoed by LaForest, who warned against trying to over engineer things up front. "There are too many variables," he said, so it's best to try it, see what works and then build on that. Kafka is a cloud native platform that scales seamlessly whenever new brokers are added to a cluster, and Confluent allows the DOD (or any other organization) to focus on iterative milestones before scaling to the size of a full mission.

Kafka for the Enterprise

The DOD's commitment to DevSecOps promises to bring a great deal of agility and capability to the operational environment, but it needs to ride on a secure, available and reliable platform. The Confluent Platform, which extends the industry-standard Apache Kafka to mission-ready form, accelerates and secures data that applications and capabilities need. Ultimately, it helps ensure that the right tools get into the hands of those who need them.

"The soldiers and the devices at the edge have the same requirements," LaForest said. "They need to be able to send and receive regardless of whether the bandwidth is great or not, or whether there is a disruption in communications." Confluent provides that at an enterprise scale.