DATA IN MOTION

Securing Data In Motion: Working at the Speed of the Mission

Overview

Introduction

Everyday user experiences are being completely reimagined. Requesting a ride through an app is better than calling a cab. Buying online is changing the way we go shopping. And gone are the days of waiting in long lines at your local bank.

Data within these modern businesses is in constant motion, flowing, in real-time, across systems, environments, and applications every time a user interacts with the system.

The most common characteristic of today's highly successful organizations is how they are powered by Data In Motion. The impact of using Data In Motion can make an even bigger impact than getting a ride when you need it when applied to mission critical intelligence operations. For Intelligence data to be effectively provided to support current and emerging missions, it is time to think about how we unlock the value of data to react to the speed of the mission.

Delivering the experiences customers demand today, whether in industry or government, requires a new way of thinking and that comes with a new set of requirements for your data infrastructure. Government agencies must be ready to adapt to meet the needs of citizens living is this data-driven, digital service world. In doing this, it is important to incorporate the policies that exist today, such as data governance, role-based access control (RBAC), and attribute-based access control (ABAC), into this shift to a data-centric mindset. By addressing policy along with real time access to data, organizations can work at the speed of their mission.

What's Data At Rest?

For decades, the focus of information technology (IT) has been on how to build and deliver applications based on a database (or lake or warehouse) where data passively sits at rest. This data is then periodically queried on demand, or in batch done hourly, nightly, weekly, etc.

This has worked remarkably well for a long time and many tools and solutions have been built around this data at rest concept. However, DoD Intelligence missions are not static, but continuously in motion, handling the streaming "now" but in context of historical information. We need to be proactive, not reactive, with the increasing technological advances and threats of our adversaries. This requires solutions to support the most-to-least connected user, across all relevant data sources, 24x7, anywhere, in real-time. These critical requirements, in the face of emerging threats and signals across Cyber, OSINT, All-Source, and MASINT domains, are not attainable with traditional, historical databases filled with data at rest that are periodically queried.

The Data Architecture Consequence

Databases remain an important part of your architecture, but they have a critical limitation—they were not designed to support the continuous real-time flow of data. The foundational assumption of every database is data at rest. As a result, agencies are burdened with architectures that become tightly coupled with disconnected data and slow, periodic batch processing that does not match the real-time operational needs. This negatively impacts an agency's ability to respond to adversary dynamics and pivot to other adversaries or AORs. This architecture will see increasing IT budgets diverted to sustainment costs rather than innovation. **These operations and maintenance costs spiral because systems built on data at rest architectures are being asked to do things they were never designed to do.**

Performing data intensive actions that are commonly seen in mission applications, such as time series analysis, multi-source integration and fusion, and AI/ML operational analytics, is extremely difficult. This is especially true when data is distributed and non-transactional, requiring disparate systems to retain caches of data in order to meet requirements. Maintaining this requires a heroic level of development skills which greatly limits the pool of contractors with sufficiently skilled staff.

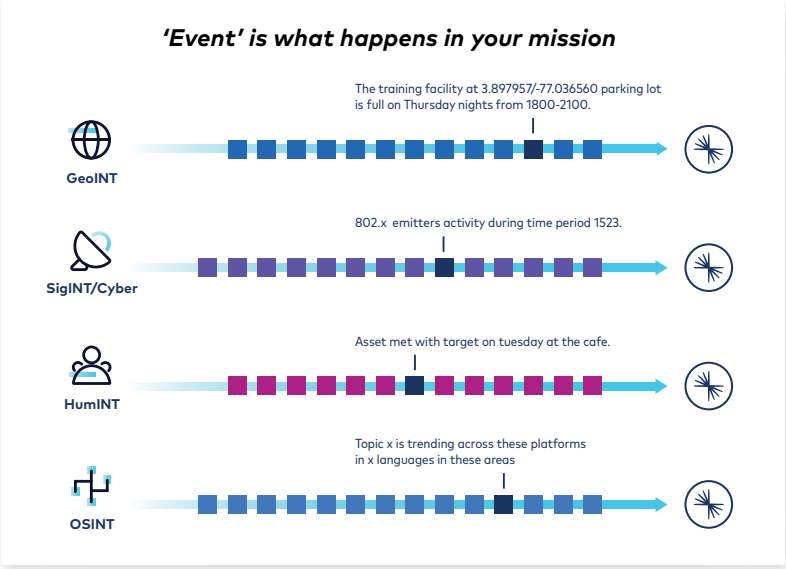
To support the missions that DOD Intelligence organizations face, a Data In Motion approach looks increasingly attractive.

Data In Motion

In order to achieve the vision defined in the DOD Data Strategy, and to be effective in surpassing our mission needs, we need to break down the silos of data management across organizations and across different lines of business including different Government agencies and programs. Equally, agencies require the ability to constantly act upon the data as it is received, whether it is from customers, citizens, sensors, or soldiers. This is Data In Motion.

The eight guiding principles in the DoD Data Strategy align well with the Data In Motion paradigm. **A Data In Motion architecture puts the emphasis on what users do with the data and not where or how it is stored and processed.** It anticipates an operating environment where sensors, applications, analytics, and missions are dynamic and changeable. In the DoD, the data architecture should be built to support activities that are frequently joint, stretched across different communities of interest, and agencies and even nations. Lastly, resilience, access control, and security are not afterthoughts, they are built-in at the atomic level.

In isolation, each reading from a sensor, each message that is ingested, each activity that is surveilled, or each observation recorded is a discrete event. Data In Motion architectures capture these events, communicate them in immutable logs, and express them as meaningful topics. These topics can comprise the output from multiple sources and can themselves be combined as vital intelligence streams. Stream processing can reveal how mult-int data relates to specific events.



Besides reducing data latency, and architectural complexity, a Data In Motion approach will provide analysts with the kind of immediate data flow they've gotten used to from social media. From a data science standpoint, the gap between observation and interpretation can be greatly shortened through stream processing.

Data In Motion enables new mission Outcomes

In motion		At rest
Find the problem now		Find the problem later
Detect threats as they happen		Detect threats too late
Real-time tagging of cyber threats	or	Batch indexing and analytics of obsolete cyber data
Real-time reporting and situational awareness		Late or out-of-date reporting and limited situational awareness
Enhanced operating picture with integrated data		Data exists in silos limiting the operating picture

Challenges of Data In Motion

There are challenges that need to be considered, specifically around **data governance and security. We need to modernize** how we apply security, access control, and dissemination rules on the Event stream rather than having to wait for this data to be fed into yet another repository. Most data governance strategies and security implementations have grown up around data at rest. In fact, traditional Security Incident and Event Management (SIEM) solutions are optimized for data at rest and search rather than data in motion. High volume ingestion and data sharing is inefficient and comes with a hefty price tag, making them ineffective for large scale event logging and processing.

A Data In Motion architecture deployed at the enterprise or community echelon with RBAC and ABAC will pave the way to a responsive, secure, and scalable approach on which Intelligence operations can be transformed.

How do you implement these critical features in the stream?

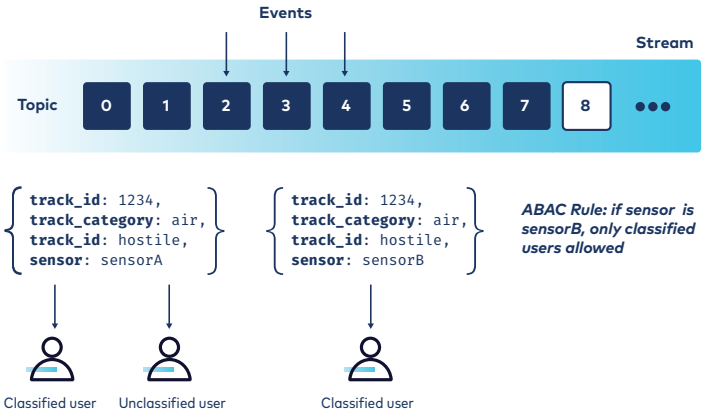
Access controls to tables and databases are typically controlled by access control lists (ACLs) and role-based access controls (RBAC). While this may be more complex to manage in the stream, the concepts are very similar, albeit at a more granular level.

Within the DOD and IC, almost every key intelligence collection, information sharing, and analysis/exploitation system has Apache Kafka—he most widely used data streaming capability, as a foundational component. With Apache Kafka, events are presented as streams of data that model a specific or combination of operations of the organization or mission. In Apache Kafka, streams contain a series of events which are grouped into topics. Authentication and accreditation can be managed using ACLs and/or integrated into an organization's RBAC solution on a topic-by-topic basis. These controls can be enforced at the source providers (producers) or the downstream systems, applications, or users (consumers). When a producer or consumer requests a topic, access can be validated and allowed or denied based on the policy defined.

What if you need to handle information that is not only controlled by the consumer rights and permissions but instead attribute by attribute?

In many of the existing solutions, attribute-based access control (ABAC) is implemented by replicating tables in databases and creating a specific view that contains the applicable fields for a user. This can lead to a very large amount of tables and easily become unmanageable.

With ABAC, authorization occurs at a granular level less than the topic in the stream. The requirement restricts access to fields within the event based on attribute types, combinations, and user roles. This will allow access to topics for all users, but also restrict certain fields based on varying classification levels or need-to-know access.



Summary

The world is becoming software defined and as a result, user experiences and data management strategies need to evolve. The data at rest paradigm has served the systems of the past well, and many important strategies around data governance and security continue to be a critical part of that evolution. Today's users expect to have immediate answers to their questions, and this is no different than the expectation within government software and missions where response times are even more crucial. With a Data In Motion architecture, our infrastructure and mission critical applications can incrementally evolve to meet the expectations of users at the speed of the mission.