Gather Once. Use Many. Moving to Data-Centric Operations

Kurt Steege *CTO, Thundercat Technology* | **Will LaForest** *Field CTO, Confluent*

In construction the saying, "measure twice cut once" is a mantra that helps avoid re-work and wasted material. IT operations professionals are finding that flipping that saying into "gather once use many" can reap similar rewards. Until recently hardware, tools, and applications have been focused not only on the tasks for which they were designed for but also on using and managing data for these tasks. This means that teams continually pull and replicate the same data over and over in order to do the analysis needed to perform their specific tasks, which has resulted in duplicative and siloed information throughout the organization.

This continual recall, storage, and processing of data is expensive in terms of costs and resources and with the current rate of data expansion is quickly becoming untenable. These costs can be mitigated by flipping the focus from tool-centric to data-centric.

Data-centricity not only makes business and operational sense, but (for government organizations) is part of the M-21-31 guidance. That document's focus on centralized access and visibility for security operation centers (SOC) and the need to quickly share information will require a shift in how data is accessed and used.

Tool-centric legacy

Over time, organizations have implemented a wide variety of tools to provide specific security safeguards or to address performance-related issues. Each of these tools consume and understand data related to the roles of the teams that are using it. In order to share that data across other related tools, a good deal of integration must be developed. Similarly, existing hardware has been limited by the cost, capacity, and compute power it was built with which likely did not account for the way today's security and operations teams need to consume and use data.

Algorithms can help bridge gaps that exist between tools and hardware, but they too are limited by the properties of the systems they run on. As a result, teams only have access to small amounts of the total pool of data to make complex decisions. To get a fuller picture, data must be copied multiple times across multiple tools resulting in high data processing costs and introducing potential for errors within that data and analysis. This continuous copying of data back and forth between these different teams onto diverse platforms makes it inherently less safe and more difficult to get a clear picture of the entire environment for a common operating model.

Data-centric future

The current state of technology allows for a different perspective on data usage. In the past couple of years, computational capacity, data generation, and algorithmic maturity has exploded. It allows us to change our perspective to focus on managing the data as the central entity, streaming it from multiple sources, enriching and contextualizing it as a data plane and in turn providing the ability to view it securely to a set of tools. The tools used by teams are then able to be focused on their tasks and not on data management.

Data Centralization also enables algorithms to work more effectively, with access to more information and working at the speed of machines to provide deeper insight in near real time. Together with the advances mentioned above we stand at a unique inflection point where hardware, software and machine learning can augment our teams to the point where the most relevant information can be bubbled up to the people who can take action in a just in time manner securely and with full situational awareness.

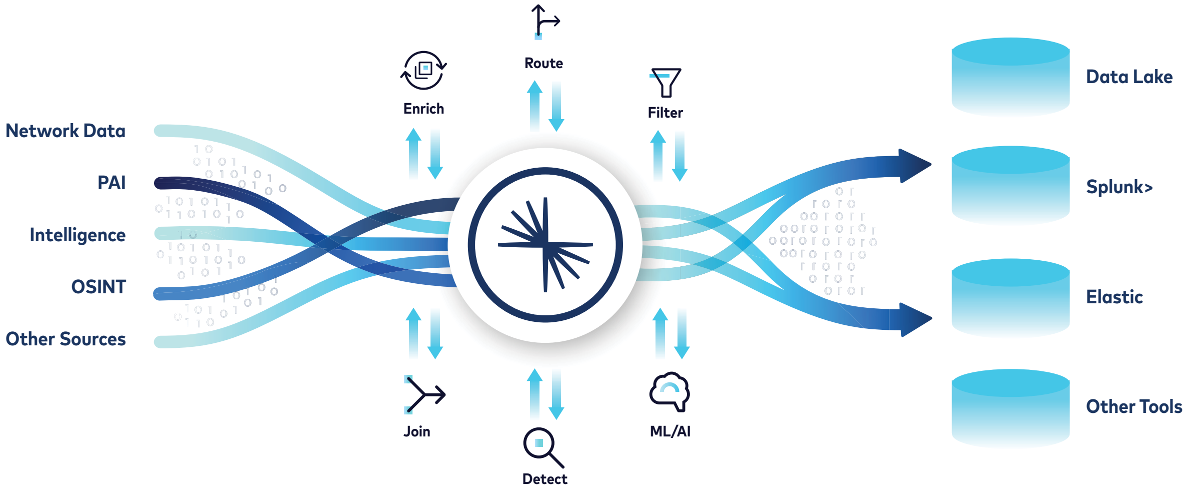
How do we get there?

In addition to changing the mindset of data being a product of a tool or role to the ideal of data as a central asset, independent of the people that use it. Moving towards data-centric culture requires a measured approach.

- Choose an event streaming platform like Confluent/Kafka to connect and process disparate data streams. Look for features like pre-built and validated Kafka connectors, enterprise ksqlDB, Stream Designer, and Stream Governance.
- Connecting the information together into a data collection flow within the streaming platform provides the visibility into the data that you have in your environment.
- Information can be enriched and contextualized for incorporation into a centralized data plane.
- The enriched, pure signal can then be placed in an appropriate place built for elastic capacity within a single namespace which in turn can have broader context applied to it through analysis and machine learning techniques.
- Once completed, the organization can then drive automation and orchestration of tasks that leverage information for response and configuration (X as Code).

In doing this, we can move the right data to the right place at the right time and allow it to be viewed in context in a way that makes sense to the teams looking at it thus enabling quicker discovery and action. Teams will be able to quickly see the data that provides value without having to sift through mounds of irrelevant information and build automated responses to situations. Data is automatically categorized into relevant buckets where what needs to be investigated further, what is required to be kept and stored and what data has no organizational value and removed to free up resources.

With this Data-centric "gather once use many" philosophy organizations can transform away from a limited view of an organization's performance and compliance though a miasma of tools to a more transparent platform for organizational observability that is secure, actionable, and useful.



Ready to get started? Contact a Confluent expert today

Email us at publicsector@confluent.io | Or visit www.confluent.io/government/ for more details