# Using Confluent to Improve the Nation's Cybersecurity

Executive Order 14028, "Improving the Nation's Cybersecurity" is an important step towards protecting against the increasing volume and danger of cyber attacks. The fact sheet that accompanied this release appropriately noted that "*Recent cybersecurity incidents such as SolarWinds, Microsoft Exchange, and the Colonial Pipeline incident are a sobering reminder that U.S. public and private sector entities increasingly face sophisticated malicious cyber activity from both nation-state actors and cyber criminals.*"

This mandate reinforces the need for log and event management and explicitly notes the importance of collecting and storing event data. The Department of Homeland Security (DHS) and Department of Justice (DoJ) provided guidance regarding "*requirements for logging events and retaining other relevant data within an agency's systems and networks*" and is found in the adjunct OMB memo mandating a maturity model for event log management, and establishing requirements for agencies and government-wide compliance with new mandates.

While executive order 14028 is unquestionably setting the right direction, it brings with it challenges around cost, speed of delivery, and an ambitious scope of requirements that will confront all agencies. There are many helpful cybersecurity tools in the market today but no single tool or vendor has a complete solution, forcing trade offs between cost, flexibility, and completeness.

## Bolstering Log and Event Management is Key to Cybersecurity

A central piece of the executive order is the collection, logging, and sharing of events in near real-time across tools and organizations. Collecting and getting a view across these events is the only way to have visibility into all the attack vectors adversaries can take. As with anything else, it's important to use the right tool for the problem. Traditional Security Incident and Event Management (SIEM) vendors focus on their own agents directly pushing all data into their proprietary stores, making it difficult to leverage tools outside their ecosystem. Since SIEMs 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.

The Confluent platform, powered by Apache Kafka, is purpose-built for scalable, durable, and highly efficient event logging and processing. Confluent enables customers to create a central

nervous system so that tools and organizations can rapidly get the right data when and how they want it. With widespread use across industry and all segments of government, Confluent has become a standard in large scale cyber defense architecture.

## Cyber Data Fabric with Confluent

Modern cyber defense architecture has moved to using an event streaming platform to provide a data fabric for receiving, logging, processing, and sharing data with cyber defense tools like SIEM, SOAR, and Machine Learning. This significant progression has come as a result of lessons learned focusing on a single vendor or tech stack: ArcSight connectors straight to the ArcSight SIEM, Splunk forwarders straight to the Splunk SIEM, etc. This 1-to-1 direct data path is ineffective for Enterprises which typically have multiple tools across vendors, open source, and operational environments that need to be able to selectively access data. Organizations have also suffered from a rigid and vendor-locked architecture making it difficult to adopt new tools and approaches.

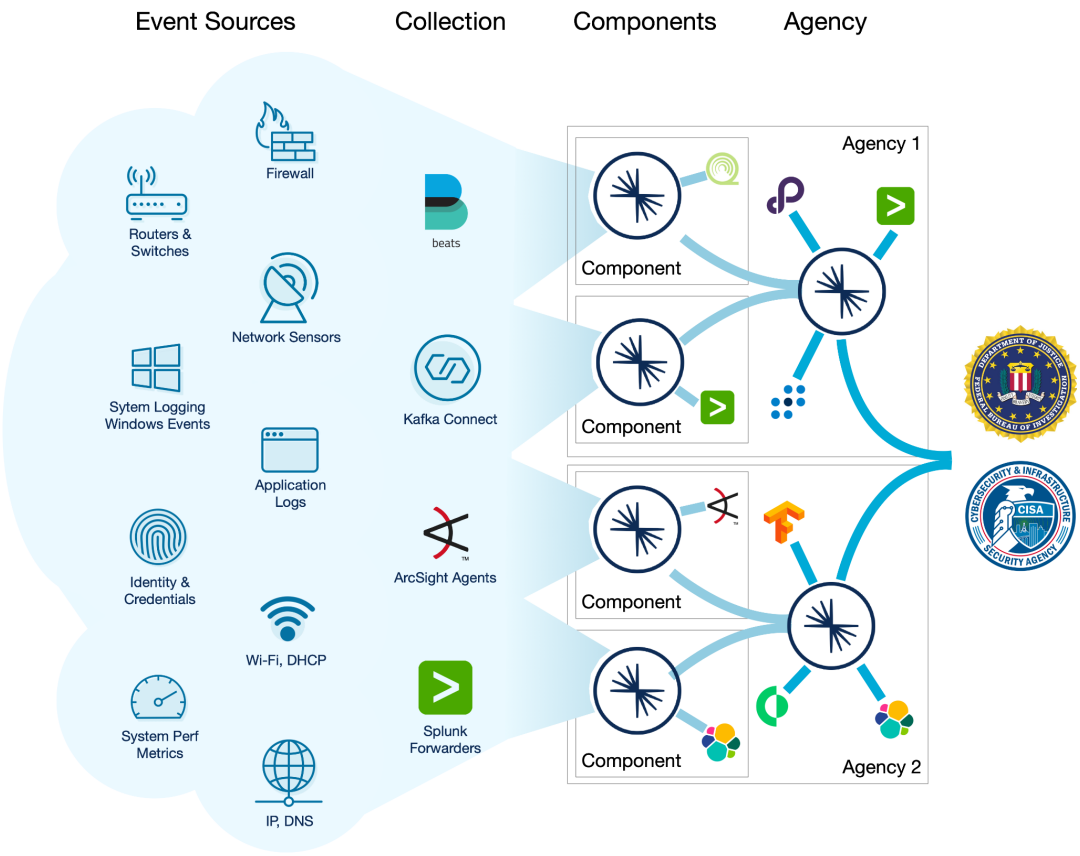


Figure 1. Data Flow for Improving Cyber Defense

The need for an event oriented data fabric in the cyber defense architecture is reflected in the policy released by OMB for the Executive Order. It specifically calls out the requirement for event forwarding and event log management in collecting, aggregating, routing, and sharing data. Figure 1 provides a high level diagram of typical data flow in such an architecture mapped to the

requirements of the executive order. Confluent serves a number of important roles in optimizing and modernizing your cybersecurity posture.

- **Collection:** Event data from the vast array of sources is received at the collection layer of the data fabric which can be made up of connectors and agents from Confluent and any variety of different vendors and open source technologies. Critically, Confluent integrates with any of the different tool ecosystems so organizations can extend the value of investments they have in place and pivot to a new technology incrementally, or mix and match.

- **Curation:** With a vast variety of different sources comes a vast variety of different formats and schemas. As noted by the executive order it is critical to normalize the data and leverage a standard set of fields laid out by the OMB memo. Confluent's native stream processing capabilities improve the quality of events with in-stream transformations to normalize, enrich, aggregate, and compress data.

- **Routing:** Stream processing provides a spectrum of techniques from rules and conditions to a cyber threat domain specific language (SIGMA) to coded algorithms which can organize data into topics which provide the structure for routing. This is critical for sending the right data to the right tools but also for determining what should be kept at the component level of the data fabric and what should be sent up to the agency level.

- **Sharing:** Data is made available in real-time for federal authorities such as CISA and FBI to subscribe to, provided they have been granted permission via Confluent's Rules-Based Access Control (RBAC). As the OMB memo requires, the schema for published data must be available for CISA and other external partners. Confluent's Schema Registry provides this automatically via APIs for all data that flows through the data fabric. Confluent ensures that as data flows through the organization, it is secured and governed every step of the way, ensuring expanded usage without bypassing requirements for risk management and compliance.

## Reduce Costs

Given the data demands of effective cyber defense and requirements of the executive order, it is critical to design for cost efficiency. The number of data sources are vast, the rates of many are high, and the OMB guidelines require 30 months retention for all but a few. Aside from just the sheer amount of storage required, significant demands will be made on the network given the distributed nature of where data is coming from and the variety of different locations for the consumers. Confluent drives significant savings in three areas:

## Save on Infrastructure Costs

Confluent's tiered storage engine automatically and transparently shifts "colder" data into inexpensive object storage behind the scenes. Indeed, the OMB memo specifically cites the need for a tiered approach but conservatively targets 12 months as the point at which to leverage cold storage. Confluent's efficient utilization of object storage makes it practical to offload all data after days or even hours, which provides an immense cost savings. Aside from huge storage savings, Confluent's tiered engine greatly reduces compute requirements by decreasing the number of nodes required to manage the higher data densities.

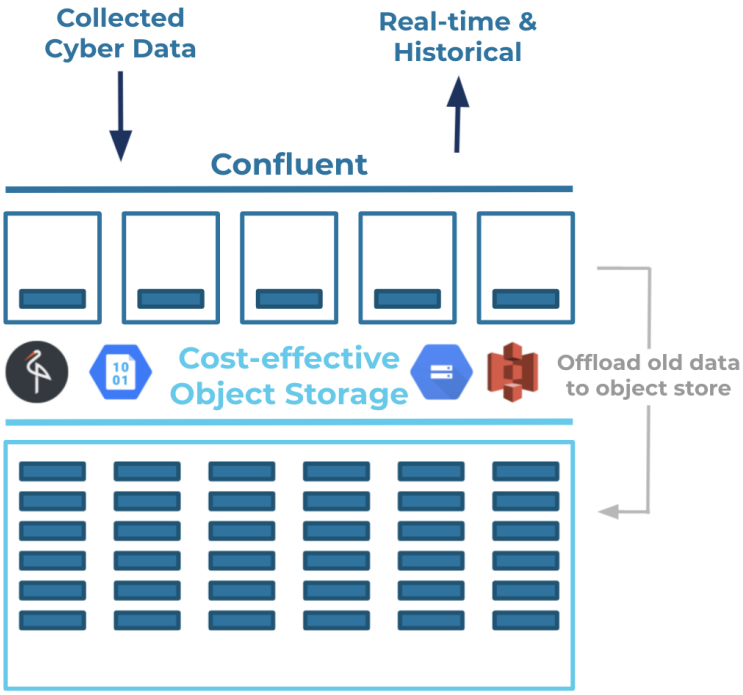


Figure 2. Confluent's Tiered Storage

## Save on Software Costs

A key tenet of this new cyber defense architecture is maximizing the efficiency of the analysis tools being used. This is accomplished by intelligently routing the right data to the right cyber defense tools. Software costs can be greatly reduced because most cyber tools license based on data volume ingested. For those tools that are not licensed this way or are open source, their costs are also directly proportional to data volume: more nodes translates to more compute. Finally, many of these tools are index-heavy to make retrospective query and analysis effective. This makes them very sensitive to data rates. It is cost ineffective to scale their compute layer based upon **peak** rates. Confluent readily absorbs peaks and variations in data rates so that indexing tools can receive data as fast as they are able and no faster. No data is dropped and this enables organizations to size indexing for average data rates rather than spikes or peaks.

### Save on Human Costs

Architecting and engineering a tiered data strategy that spans the wide variety of tools that will be used across agencies and components would be a large and expensive effort in and of itself. It's easy to write data into tiered cold storage, it's hard to integrate this across hundreds of different source technologies and scores of cyber defense destinations. Building a different process for each set of vendor tools to move data to cold storage and designing special mechanisms for fetching data from cold storage and ingesting it into the tools that might need it is hard. By using Confluent for your data fabric, no additional work needs to be done to support cold storage requirements. Any data that flows through the already integrated fabric is automatically written off to cold storage and no new special integrations and mechanisms are required to access it.

## Design for Better Cybersecurity

Saving money and complying with mandates are important but irrelevant if efforts do not result in a more secure government. A cyber data fabric built on Confluent helps enhance capabilities provided by SIEM vendors.

### Enable Faster Response

The faster you detect and act on threats the more likely you are to stop them or mitigate the damages of attacks. Confluent enables one to push threat patterns to the data fabric for real-time detection and routing via stream processing as well as the open source threat detection DSL (SIGMA). This enables organizations to distribute the processing at the point of collection which can avoid unnecessary latency and network costs.

### Tap Into High Velocity Data

Another lesson learned from the trenches of cyber defense is that some data sources have rates and volumes so high that it is impractical to apply traditional SIEM principles. It is not unusual for packet capture, net flow, and even DNS data to fall on the floor based upon expense or inability to ingest. Confluent delivers the ability to push threat detection into real-time streams of data, enabling the examination of data streams which otherwise would have been neglected.

## Partnering For Success

A data fabric built on Confluent enables a more effective and affordable cyber defense operation. Confluent helps organizations from across Government and Industry modernize and improve their cyber security posture while extending the value of their existing investments by embracing and leveraging their chosen cyber security vendors and ecosystem. The first step towards joint success starts with a Confluent architectural assessment to devise a plan for how Confluent can help you reach your goals.

## Confluent Capabilities Enable

- Reduced storage, network, and licensing costs
- Enables tiered and cold storage out of the box to support OMB requirements
- Aggregation at the Component and Agency level through Confluent replication
- Enables agencies to leverage their individual cyber investments via connectors
- Makes data available to FBI and CISA in realtime out of the box
- Provides real time threat detection abilities at edge
- Enables open ecosystem and prevents vendor lock-in
- Built in auditing to support requirements on tracking usage of data
- Provides RBAC to lock down data to only those who have a job related need
- Provides Schema catalog with APIs for sharing of data schemas with consumers