Building better data pipelines for DevSecOps

Continuous security compliance hinges on robust data flows and the power of observability

Nick Heudecker

Cribl

Digital transformation allows agencies to improve productivity, deliver services more efficiently and automate repetitive tasks. Achieving this kind of transformation requires rethinking how software is delivered and maintained, and many agencies have adopted DevOps to meet these challenges.

Despite its advantages, DevOps often conflicts with security requirements. Security compliance checks occur late in the development process, creating unacceptable slowdowns and making it difficult to comply with new or revised security policies. In response, practitioners have adopted a "shift left" mindset, with security as part of an automated process integrated into DevOps workflows. The result is DevSecOps.

DevSecOps isn't just a concern at build and deployment time. After an application has been deployed, a key part of DevSecOps is providing continuous feedback on application performance and security. For complex applications running on modern architectures, traditional monitoring methods may not be dynamic enough. Instead, agencies should consider observability tools that pull together data from multiple sources to give them insight into the internal state of an application and ask questions they may not have considered initially.

Building a strong observability discipline in a DevSecOps environment requires having a good understanding of all the different data touchpoints. Teams must be able to analyze a lot of different data from a lot of different sources and do it in real time or near-real time. Therefore, observability should be delivered where developers will encounter it. These capabilities must be embedded into their toolchains.

Essential features: Integration and manageability

When building data pipelines for DevSecOps (or any activity), agencies should focus on three key elements: integration, manageability and scalability.

Integration can easily represent 80% of the work involved in building a data pipeline. There are a multitude of integration points, and a data pipeline needs to understand all data sources, protocols and destinations natively. Sources include the build system, the ticketing system, testing, logging, monitoring, auditing and governance. And agencies must decide how data will move between those sources and destinations.

Next, agencies must answer central questions about managing those pipelines. Who will be in charge? Do they have the necessary skills? Do they understand the data they'll be consuming?

Data pipelines are always evolving, so it should be as easy as possible to manage and add to them. That includes addressing governance issues such as deciding how to redact personally identifiable information in a consistent way, how to share subsets of data with different parts of the organization and how to change the rules under which the pipeline operates, when necessary.

David Clode

> Teams must be able to **analyze a lot of different data from a lot of different sources** and do it in real time or near-real time."

The importance of scalability

Adding in DevSecOps-related data can rapidly increase flows that might already number in the hundreds of terabytes a day. In addition, agencies often deal with bursts of activity — for example, tax filing season — when 30 terabytes a day turns into a 300-terabyte or even a multi-petabyte workload. Therefore, agencies should optimize for the common case but ensure that their data pipelines can handle demand outliers.

Building data pipelines from scratch and managing all the integrations can take a significant amount of effort and time, perhaps as long as a year. By contrast, agencies can buy a product from a trusted partner and be up and running in days or weeks, with the added benefits of built-in observability tools and ongoing expert support.

In digital transformation, there are no prizes for second place. All government agencies should have the ability to move forward quickly and securely to provide the apps and digital services their users need. ■

Nick Heudecker is senior director of market strategy at Cribl.