Datadog bolsters observability data stream processing, rounds out platform with Vector

May 7 2021

by Liam Rogers, Fernando Montenegro

In February, Datadog picked up Timber Technologies and Sqreen to boost its capabilities around observability data stream processing and application security. The integrations are still a work in progress, but give Datadog a useful tool for unifying disparate data sources and delivering that data to multiple destinations and personas.
Introduction

Datadog’s 2020 acquisition of Undefined Labs brought a dose of ‘shift left’ into the observability vendor’s portfolio and kicked off a small acquisition spree – the vendor has since announced its second and third acquisitions in the past year. In February Datadog picked up Timber Technologies and Sqreen to boost its capabilities around observability data stream processing and application security and further round out its platform. The integrations are still a work in progress, but Vector (from Timberline) gives Datadog a useful tool for unifying disparate data sources and delivering that data to multiple destinations and personas.

THE 451 TAKE

The capabilities from Vector provide an onramp to easily bring more data into the Datadog platform, as well as an element of agnosticism or symbiosis since users can continue to move data to other platforms as needed. Customers won’t be reducing their observability spending to one vendor while excluding all others, but if Datadog can position itself as the vendor that provides the glue between different platforms and personas, it could get a leg up on competitors that are driving home the message of consolidation to a single pane of glass at all costs. Because Vector enables data to be transformed as it is routed to multiple destinations, it has potential to be the connective tissue between different personas, piping logs and metrics useful to site reliability engineers, developers, security and IT ops teams alike to the destination that best matches their use cases. However, at a time when the effort to consolidate is driven by concerns of cost and complexity, adding more pipelines to the mix could be seen as a complexity contributor for some organizations.

Context

Datadog finished its fourth quarter (ended February 11, 2021) with $177m in revenue, almost 9% higher than consensus estimates, according to S&P Global Market Intelligence. Full-year revenue reached $603.5m, up 66% year over year. The vendor has also touted growth in large customers, ending calendar year 2020 with 97 customers with $1m-plus in annual recurring revenue, compared with 50 one year prior. The recent acquisitions play into the vendor’s strategy as it looks to maintain its growth and expand its total addressable market. 451 Research’s Voice of the Enterprise data shows that 83% of organizations would prefer to buy monitoring and incident response tools from a single supplier when possible. Like its peers, Datadog intends to meet as many customer needs as possible and be a platform that can unify the disparate data sources within an organization.

Vector and observability data stream processing

Vector is a project that was established by startup Timber Technologies. Timber was originally developing a log management platform (hence the name), but pivoted to data stream processing after seeing the glut of log management services available on the market and that user challenges often revolved around data movement itself. The team consisted of 15 people prior to the acquisition, and the company had taken in $6m in funding over the course of several seed rounds.

Vector is based on the concept of connecting sources, transforms and sinks. It is an aggregator that allows users to take logs and metrics from any number of different sources (could be AWS ECS metrics, Kubernetes logs, Nginx metrics, etc.) and send them to an equally varied set of end destinations, whether it be an observability platform, a data lake or just cold object storage. Once operational data is collected, Vector structures it by event type, either a log or metric, and then can transform and enrich data before it is sent to a downstream service. There are 13 types of transforms available that are fairly broad in their utility, and their conditions can be tweaked based on a customer’s needs (users are also able to write their own custom transformations). Examples of transforms include providing the ability to reduce multiple logs into a single log event based on specified conditions, sampling events at a rate configured by the user, and setting a limit on the cardinality of tags for metrics – to save space and, ideally, achieve significant cost savings at scale.
From a security and compliance perspective, the data transformation and filtering capabilities enable users to obscure sensitive information in data as it is sent to its chosen destination, enabling the scrubbing of information that shouldn’t be accessible to certain audiences even if an organization wishes to grant them access to parts of the data. Additionally, Vector was built using Rust, which the team asserts has security implications because of the language’s memory safety and the fact that Vector is a single binary with no dependencies or runtime, thereby reducing the potential attack vectors. The lightweight nature of Vector also makes it easy to deploy in a range of environments, including microcomputers such as Raspberry Pi (Vector is x86- and Arm-compatible), presenting edge opportunities for transforming and moving data.

The Vector team maintains that, even as part of Datadog, openness and vendor agnosticism are critical to the project and ensuring that as many data sources and destinations as possible are accounted for. The roadmap includes maintaining the open source version alongside a commercialized SaaS version for enterprise customers, which will be made available as part of the Datadog portfolio. However, the details, including pricing model, for the enterprise version have yet to be made public.

Sqreen and app security
Monitoring vendors have been emphasizing the utility they can offer for security practitioners, and this has driven a number of acquisitions in the space, with Datadog’s purchase of Sqreen among the most recent. Sqreen is a SaaS security vendor that provides an agent for monitoring application runtimes. Founded in 2015, the company took in $28m in funding, mostly from a series A of $14m from Greylock in 2019. It had accrued 800 customers prior to the acquisition.

Architecturally, Sqreen uses what it refers to as a ‘microagent’ – a library that can be installed in a given app, microservice or API by developers before it is deployed. The microagent is compatible with Go, Java, Node.js, Ruby, PH and Python. Once deployed, the agent then monitors the app during its lifespan and communicates metadata to the Sqreen SaaS platform, which can then issue configuration updates and similar changes to mitigate threats detected by the security engine. The multi-layered security engine leverages a mix of runtime application self-protection (RASP), in-app web application firewall and user data protection.

This will augment Datadog’s portfolio and the existing threat-detection capabilities of its Cloud Security Platform offering to provide an added layer of security to customer apps, looking to capitalize on the importance that organizations are placing on security. Datadog says that, for customers of the Datadog APM tools, it will be a simple choice to enable security detection and protection capabilities on already-instrumented services. According to 451 Research’s Voice of the Enterprise, DevOps Organizational Dynamics 2020 survey, security is identified as the top concern and obstacle for those deploying cloud-native technologies. Like Vector, how this service is ultimately integrated into the Datadog platform and dashboard remains to be seen.

**Competition**
Due to the breadth of its portfolio, likely competition for Datadog continues to be other vendors in infrastructure monitoring, such as Instana (IBM), LogicMonitor, New Relic, ScienceLogic, SolarWinds, Sumo Logic and Zenoss. With regard to Vector, startup Cribl specializes in observability data pipelines and also enables users to transform data from different inputs en route to a chosen destination. AWS Firehose also routes streaming data from various inputs to different destinations, although one such destination is Datadog, so there will be synergistic deployments, making Firehose less likely to be a direct source of competition. For Sqreen, there are numerous vendors offering runtime protection and related capabilities, including but not limited to Palo Alto Networks, Trend Micro, Aqua Security, Imperva, fastly (via the acquisition of Signal Sciences), Contrast Security, Dynatrace and Cisco (via AppDynamics).
## SWOT Analysis

<table>
<thead>
<tr>
<th>STRENGTHS</th>
<th>WEAKNESSES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Datadog is an established vendor with a well-rounded portfolio. Its recent acquisitions demonstrate that it’s willing and eager to have its platform cater to a wider audience and span more of the application lifecycle than just what ends up in production.</td>
<td>It’s still early for these acquisitions, and Datadog will have to iron out the finer points of integrating these technologies with its existing portfolio, including how to price Vector to retain open source users and draw revenue from Datadog’s larger enterprise clients.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OPPORTUNITIES</th>
<th>THREATS</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ability to transform data en route as it shifts to a variety of destinations could be a valuable tool for opening up new partnerships outside the normal bounds of Datadog operations, as data is piped to workload-specific personas such as data scientists.</td>
<td>Perception of open source in observability is somewhat in flux, with prominent software moving to a closed source model, and there will be some skepticism about any vendor-acquired open source project remaining truly open and competitive with its commercialized version.</td>
</tr>
</tbody>
</table>