SECURITYSCORECARD.COM

SecurityScorecard Taps Confluent to Cost-Effectively Power Real-Time 360-Degree Security Prevention and Response at Massive Scale

SecurityScorecard

Headquarters New York, NY

Industry High Tech/Cybersecurity

Challenge Escalating costs and burden of managing Kafka on their own, while relying on Kafka services that lacked a complete platform for all their streaming needs.

Solution Used a combination of Confluent Cloud and Confluent Platform to build streaming data pipelines and scale faster, govern data better, and ultimately lower TCO by offloading the management of their data streaming infrastructure to Confluent.

Results

- $1M+ in TCO savings for their data streaming infra compared to managing in-house
- The ability to seamlessly scale data 10 to 100x, a game changer for the company
- Moving from Amazon MSK to Confluent Cloud enabled a 48.3% reduction in projected annual operating costs
- Support for streaming pipeline use cases—with a direct impact on revenue
- Increasingly granular data governance and security to confidently scale streaming to more teams
- Worldwide availability and reliability with the cloud deployment

Confluent Features Used

- Connectors (PostgresSQL source, PostgresSQL sink, S3 sink, Splunk sink)
- Cluster Linking
- Role-Based Access Control
- Stream Governance
- ksqlDB

When people want to know their credit risk, they consult one of three major credit scoring companies. And when companies want a holistic view of their security posture, they turn to SecurityScorecard. The global leader in cybersecurity ratings, and the first cybersecurity ratings company to offer digital forensics and incident response services, SecurityScorecard tracks potential weaknesses with certainty: exposed servers, suspicious IP addresses, breached employee accounts, malware-infected devices, out-of-date endpoints, and much more. The number of potential cyber risks today is enormous and growing, and SecurityScorecard analyzes over 200 issue types, and discovers billions of security issues weekly.

Accurate, up-to-date data is the lifeblood of their business, and it needs to come from a myriad of sources across the internet. "Our platform relies on data collection and processing being done extremely accurately, in real-time and at scale," says Jared Smith, Senior Director, Threat Intelligence, SecurityScorecard. This is what led the company's engineering teams to adopt data streaming and how they ended up using a novel combination of Confluent Cloud and Confluent Platform to build streaming data pipelines and scale faster, govern data better, and ultimately lower the total cost of ownership (TCO) by offloading the management of this key piece of infrastructure to Confluent.

"Since we built Horus, our global IPv4 scanning platform, on top of Confluent, we've saved over a million dollars compared to open source Kafka or MSK. Business resilience and ensuring no disruption to delivering customer value, all of that is enabled by having a system like Confluent that works securely and reliably to do data streaming."

— JARED SMITH, SENIOR DIRECTOR, THREAT INTELLIGENCE, SECURITYSCORECARD

Risks Assessed in a Matter of Days, Not Weeks or Months

SecurityScorecard is in the business of building solutions for customers that mine data from dozens of digital sources to discover security risks and flaws inherent to their businesses. This becomes increasingly important as companies grow and new sources of risks and threat vectors emerge. "As you become a larger organization, you can say with confidence that everyone's using, say, Microsoft or the Adobe suite," says Brandon Brown, Senior Software Engineer, data platform, SecurityScorecard. "But what payment processors are you using? Stripe? PayPal? The bigger the company you are, the smaller things are most likely what you miss."

Recently, a customer contacted SecurityScorecard with an issue to resolve:, "Hey, your automated vendor-detection product shows we use HubSpot, but we stopped using that ages ago." SecurityScorecard investigated and confirmed a very recent HubSpot ping. Sure enough, when the customer looked into it, the marketing team had started experimenting with another HubSpot instance on the side. Without the continuous data analysis that stream processing provides, the customer would not have known about this potential security risk. Brown confirms, "If you can automatically know about these things and not have to shoulder-tap a million people, that's huge company savings."

SecurityScorecard built their platform on open source Apache Kafka® because, as Smith says, "There's just not another system out there that gives you the fundamental tools to build literally whatever you want." Stream processing enables the company to process information in a matter of milliseconds, instead of weeks or months, so that detection of a website's security posture risk happens quickly in a landscape of constantly evolving security threats.

The Threat Intelligence and Research team led by Smith started off with self-managed Kafka, but quickly realized that the small team was spending up to eight hours a week to maintain Kafka—time that could be better spent focusing on new product features. This was the impetus behind the initial push to the Confluent Platform.

"Companies care more about cybersecurity than ever. It's just amazing how much more we can get done when we don't have to worry about exactly how to do things. We can trust Confluent to offer a secure and rock-solid Kafka platform with a myriad of value-add capabilities like security, connectors, and stream governance on top."

— JARED SMITH, SENIOR DIRECTOR, THREAT INTELLIGENC, SECURITYSCORECARD

The Ease of Building Streaming Data Pipelines via Confluent

There are two main teams using Confluent within the organization, and they work together in many ways. The first is Smith's team—Threat Research and partner team Signals Intelligence (SigInt)—which use Confluent Platform to collect data from various sources for risk analysis. The second is the Data Platform team led by Brown, which uses Confluent Cloud to collect and analyze the data Smith's team provides and create the summary analysis for customers.

SecurityScorecard's developers within Smith's team partnered with Confluent to build Horus, a global distributed system capable of running any agent-based code anywhere in the world, taking instructions off Confluent and producing data back to Confluent through real-time streaming pipelines. Dozens of battle-tested connectors feed data in and out of Confluent, and Horus acts as a management platform on top of it all. Smith's team only has to write Python-based applications that get deployed as agents on this system, and Confluent manages everything else, including the connectors that link all of these data sources and sinks. Currently, the agents perform various tasks, including IPv4 scanning, web crawling, vulnerability detection, and API integrations with partner data feeds.

This structure is important to Smith, who compares it to the previous version, built on RabbitMQ: "Our prior system, RabbitMQ, wasn't able to scale. With Confluent, we can scale seamlessly and deploy from Confluent's connector portfolio instead of building connectors on our own."

The two biggest fully managed connectors SecurityScorecard uses are the PostgresSQL (both source and sink) and S3 Sink, enabling the Threat Intel team, the Data Platform team, and other teams across the company to access streaming data for different purposes. They are also beginning to experiment with the Debezium CDC source connector. These connectors also enable the teams to create data archives that act as a history of assets and seamlessly link everything together in real-time to create a consistent data layer across the business.

Confluent is also critical to the company's scanning and web crawling functionalities. Billions of records of data from all of the databases that track breaches are pushed through Confluent so that data can be "replayed" by any team. Smith says, "Confluent's fully managed Amazon S3 connector enables teams to focus on development work by removing the operational burden and risk of managing and maintaining these connectors on their own."

This ease has given SecurityScorecard the flexibility to free up 2 FTE resources and time to think up new products as time goes on. One of them is the brand-new Internal Security Suite, a tool that pulls all the data from a customer company's various security vendors and internal intelligence tools and contextualizes it in a common place.

"For us, there's no desire to go build a bunch of managed things for Kafka when Confluent already does it."

— JARED SMITH, SENIOR DIRECTOR, THREAT INTELLIGENC, SECURITYSCORECARD

The Path to a Flexible Hybrid Cloud Deployment

When Brandon Brown came on board, one of his first projects was to upgrade his team from Kafka 2.6 to Kafka 2.7—a move that seems incremental, but was an enormous lift, partly because building data pipelines at SecurityScorecard was still entirely reliant on Amazon MSK at that point in time.

"MSK was not meeting our operational needs as doing something like a version upgrade was hard and very manual. We spent a lot of time trying to figure out cluster size and had some challenges with figuring out the number of brokers to set up." Because of this time-consuming process, Brown's focus was taken away from developing actual value-add business applications. Instead, he spent time trying to learn how to make MSK do what he needed it to do. By moving to Confluent Cloud, things that were otherwise overly hard—like cluster and connector management—became simple and more reliable.

Another obstacle for Brown was bottlenecks occurring with big JSON files that had to be pulled down off the database and filtered to be usable. It would take the team multiple days to process a single day's worth of data. Brown knew the data was in Confluent and that the process could be much faster, so he helped develop a fan-out process to subscribe to specific topics and consume the data from Smith's team's Kafka cluster. Now, Brown says, "It's super fast because it's binary messages and we don't have to do any filtering at all. We're getting the performance we want without having to worry about tweaking anything, which is really nice." Data is seamlessly passed from on-premises clusters from the Threat Intelligence team across data centers outside of the U.S. to Confluent Cloud clusters in the U.S., leveraging Confluent's innovative Cluster Linking capabilities to connect their hybrid environments.

"If you want to manage Kafka yourself, you get into a bubble of taking valuable time and resources to deploy everything out at once and learning how to configure all these things. With Confluent Cloud, it's very easy to experiment, and cost-efficient in the grand scheme of things."

— BRANDON BROWN, SENIOR SOFTWARE ENGINEER, DATA PLATFORM, SECURITYSCORECARD

Pay-As-You-Go as an Entry Point to Confluent Cloud

As Brown has shifted teams and projects from Amazon MSK to Confluent Cloud, an important enabler of this transition has been Confluent's pay-as-you-go (PayGo) pricing model. In the beginning, he says, "We started out with just the basic cluster, so our cost was pennies a day. That early experimentation of being able to actually stand up a cluster that was running—that you could deploy code to your environment and have it running without having it just be on your machine—that was a really big win and didn't cost a lot. Dollars a day, really."

Based on the promising early experiences with Confluent Cloud, Brown was able to calculate the costs of migrating from MSK to Confluent Cloud and make a strong business case for this move: "I was able to make some back-of-the-envelope calculations, to show how much running three clusters in MSK was costing us a day, and contrast that with our costs in Confluent Cloud. And the Confluent cost was about the same—or a little cheaper in some instances—but my management overhead was so drastically reduced with Confluent that it immediately showed value from a total cost of ownership perspective."

The Results

A flexible, scalable, and global data streaming platform powered by Confluent, enabling SecurityScorecard to innovate with data and build streaming data pipelines for real-time initiatives across the business.

Unmatched scalability

"We went from scanning the internet in a month and a half on 80 ports—and, remember, there's 65,000 ports—to being able to do what people said was literally not possible: over 1400 ports in a week and a half, at a level that no other system could produce, which directly led us to secure many more customers that need to see data across thousands of services. As the leading security ratings company, the data behind the scenes is critical to our success. Confluent has allowed us to simplify our system architecture and focus on building great data collection tools, which has, in turn, amplified our ingested and processed data 10 to 100x, and that's it, that's the game changer."

— JARED SMITH, SENIOR DIRECTOR, THREAT INTELLIGENCE TEAM, SECURITYSCORECARD

Enabling real-time security data collection at scale is the crux of what Confluent unlocks for SecurityScorecard. As just one example, in the previous system, which had been built with RabbitMQ, we did not know whether adding a port (or many ports) to be scanned could crash the entire system. The old architecture couldn't handle megabytes of throughput per second—something SecurityScorecard easily accomplishes now with one dedicated cluster in Confluent Cloud.

Support for numerous use cases

"The revenue boost is huge because we free our teams to go work on more real-time use cases."

— BRANDON BROWN, SENIOR SOFTWARE ENGINEER, DATA PLATFORM, SECURITYSCORECARD

Brown's team is building many new use cases for SecurityScorecard by tapping into streaming data pipelines. They're taking advantage of data sharing to downstream systems with minimal infrastructure overhead—for instance, managing a source of truth of data sent from upstream systems and applying logic on top of it, then sharing data from the PostgreSQL database, where this data lives, to downstream systems using a combination of polling and change data capture (CDC) connectivity. They've also built a new automated vendor-detection product that can onboard new data sources quickly via topics and refresh that data every hour. The impact of doing this in mere days has been huge, particularly in a constantly evolving landscape of security threats.

True confidence with data governance

Data governance is critical to SecurityScorecard—what data they have, where it comes from, what it looks like. Currently, they use a custom-built protobuf library to control who has access to sensitive data. The goal is to eventually use Confluent, so multiple teams can share the same Confluent source data, making it much easier to govern. Confluent's Stream Governance and role-based access control capabilities will enable Brown's team to act as gatekeepers for who has access to the cluster and what they do with it. In the future, he plans to also explore role-based access controls, where specific teams will have access to certain topics automatically.

Worldwide presence and hybrid cloud deployment

SecurityScorecard was unable to run all its cloud services globally, because certain regions were unsupported by cloud vendors. Confluent Platform solved that problem. Today, SecurityScorecard's Horus scanning engine has agents deployed on every continent except Antarctica. Moreover, with a hybrid cloud deployment where data from Confluent Platform clusters stream into Confluent Cloud clusters, different teams within SecurityScorecard can seamlessly and reliably get access to the same data in real-time.

Expert support when it counts

"When we couldn't figure something out and had to ask an actual expert, the partnership with Confluent gave us someone to provide an answer."

— JARED SMITH, SENIOR DIRECTOR, THREAT INTELLIGENCE TEAM, SECURITYSCORECARD

SecurityScorecard took advantage of Confluent Professional Services to talk through some of the problem-solving of building out use cases. Confluent's premium support channels also enable SecurityScorecard to talk to the experts behind the core technology, Apache Kafka, any time of day and under any circumstances.

A quantifiable TCO

"I definitely saved days of management every week just by not having to do anything to make sure our Connect cluster is working. With Confluent Cloud, it just wor*ks."*

— BRANDON BROWN, SENIOR SOFTWARE ENGINEER, DATA PLATFORM, SECURITYSCORECARD

"Since we built Horus on top of Confluent, we've been able to save over a million dollars in what we were paying other vendors for, not to mention the additional operational overhead we're now able to offload. It makes so much of a difference. Business resilience, customer value, all of that is enabled by having a system like Confluent that just works securely and reliably to do streaming. We have since built a number of other systems on top of Confluent Cloud, all enabled by the simple push/pull from Kafka topics and then relying on Confluent Connect fully-managed connectors to handle getting our data and insights to the right places."

— JARED SMITH, SENIOR DIRECTOR, THREAT INTELLIGENCE TEAM, SECURITYSCORECARD

Just for the Confluent Cloud versus Amazon MSK part of the equation, Brown's team estimates they save about $125K a year. There's also the harder-to-quantify but very real operational overhead that has been alleviated with Confluent, where the team estimated an 80% reduction in Day 2 operational burdens, resulting in an overall 48.3% reduction in projected annual operating costs.

"I wanted Kafka, and I didn't want to deal with management, and there's only one company for that: Confluent. I'd much rather go with the platform built by the actual experts themselves for things that are going to be relied upon on, for the core of our business."

— JARED SMITH, SENIOR DIRECTOR, THREAT INTELLIGENC, SECURITYSCORECARD

What's Next

The Threat Intelligence and Data platform teams within SecurityScorecard are excited about the potential of streaming data pipelines to unlock the true value of their data. The vision is to ultimately become a data organization where people can discover data, understand where it came from, and share data they generate with other teams to facilitate easier reuse. ksqlDB and stream processing are also a big part of their future plans to replace spark heavy batch jobs so joins of data can happen in real time and data quality issues can be proactively addressed. Another place Brandon and Jared hope to leverage ksqlDB is to lower the amount of custom service deployments for simple join tasks, which will allow SecurityScorecard to have consolidated observability and save on infrastructure costs.

"We want to shift our thinking to model the flow of data rather than the implementation detail of the systems that have to make this data flow. We are moving away from sequential thinking, where the next "step" is triggered after the previous step is completed. Instead, our data is continuously flowing, so our systems and applications can react to the flow of data and not the other way around."

— BRANDON BROWN, SENIOR SOFTWARE ENGINEER, DATA PLATFORM, SECURITYSCORECARD

Learn More About SecurityScorecard

https://securityscorecard.com

Get Started with Confluent for Free