SPONSORED CONTENT

AWS Insider.net

Highlights from a webcast about a revolutionary new approach to log management

HOW HUBSPOT SOLVED THE CHALLENGES OF AN OVERWHELMED ELK STACK

A Digital Dialogue based on an Amazon Insider Webcast demonstrating how HubSpot used ChaosSearch to markedly improve its DDoS defenses.

Den Rise / Shutterstock.com

Protecting 78,000 customers in 120 countries from distributed denial of service (DDOS) attacks requires HubSpot to collect, store and query mass quantities of data.

In a recent webcast, Stephen Salinas, engineering lead at HubSpot, and Dave Armlin, vice president of solution architecture and customer success at ChaosSearch, explored how HubSpot turned to ChaosSearch to take control of the issue.

HubSpot offers a full stack of marketing and sales services with a free customer relationship management solution at the core. Keeping DDoS attacks against their customers at bay has been a major priority for the company.

"We want to address things as soon as they start, not when it's completely down," Salinas said.

The company uses the Cloudflare Content Delivery Network (CDN) as a primary defense against DDoS. In order to store the log entries and other data generated by Cloudflare from such a large base of customers, HubSpot had built and been managing their own ELK Stack. ELK stands for Elasticsearch, Logstash and Kibana—three open source projects. Logstash takes the data from multiple sources, ingests it and transforms it. It then stores the data in the Elasticsearch search and analytics engine. Finally, Kibana enables the team to visualize

> "Looking at the numbers, it was roughly a 40% savings for being able to search the same data set in a very similar Kibana UI." *–Stephen Salinas, Engineering Lead at HubSpot*

the data with charts and graphs.

In HubSpot's case, Cloudflare's log data went into Amazon S3, and the company used the ELK Stack along with Kafka to ingest, transform and analyze the data. However, over time HubSpot's engineering team found that they were growing out of the ELK implementation that they were using to firefight DDoS attacks.

"This particular use case got so large that it actually first needed to be partitioned off into its own cluster," Salinas explained. "But there was still too much data a lot of times for even the singular Elasticsearch cluster to hold. Every time HubSpot would grow, we would have these conversations of, do we throw more compute at the problem, which ends up costing us a lot more money? Or do we shorten our data retention, which ended up being the solution a lot of the times."

Adding to the stress on the DDoS firefighting ELK cluster was the DDoS attacks themselves, which by their nature produce significantly more data. "At the time that we need the data the most, we're actually trying to shove more through an already stressed pipeline," Salinas said.

"If the ELK Stack starts to get lags during one of those critical time periods, you can imagine that our feedback loop on 'Did our block for this attack work?' gets slower and slower and slower. We really needed to find a solution that was not going to reduce our retention to a point where it was so small, it wasn't useful anymore, but also be able to keep up with the throughput at a fairly low latency," he said.

HubSpot started evaluating other solutions, even including a hard look at using Amazon Athena to query data directly inside Amazon S3. "However, the kind of very fast iterative narrowing down use case was something that was not quite as well served by Athena as by an Elasticsearch-style interface, like interacting with Kibana," Salinas said.

Through continuing Google searches, the HubSpot team soon found ChaosSearch, which describes itself as a massively scalable ELK-compatible log analysis platform that is delivered as a fully managed service.

"We deliver insights directly out of the customer's S3, and our security model and our basic compute fabric and ability to index and query data provide a disruptive performance and price solution," ChaosSearch's Armlin said.

For HubSpot, the new architecture resulting from the ChaosSearch solution is drastically less complicated than it was under the old ELK Stack approach.

"The architecture is drop dead simple, really, which is one of the really nice things," Salinas said.

The Cloudflare Logpush still goes to Amazon S3. But ChaosSearch reads from S3, and HubSpot can do almost all of the things it was previously doing on its own in the ELK Stack from within the ChaosSearch managed service.

"From our side of things, it's really quite a touchless system. We don't really have to manage the scaling of the ingestion components. We don't have to manage the scaling of the query side components. That's all taken care of by ChaosSearch," Salinas said. "Overall, the new setup has just been really easy to get going and really touchless to maintain, which has been excellent."

For HubSpot, the bottom line of moving from the ELK Stack to ChaosSearch has been a huge savings.

"One of the biggest things that was a selling point to a lot of people at HubSpot was the cost of the system versus maintaining one ourselves. Looking at the numbers, it was roughly a 40% savings for being able to search the same data set in a very similar Kibana UI," Salinas said. "Probably the second biggest thing from the HubSpot perspective was the increased data retention. We had gotten to some points with our ELK setup where retention was down to, four or five days. Now, I get 30 days or more!"

SPONSORED BY:

CHAOSSEARCH

Find out more:
https://www.chaossearch.io/