Datadog

the challenge:
How do you turn 110 million events into actionable intelligence that helps customers stay up and running?

the solution:
By using Elasticsearch to provide true real-time visibility combined with powerful historical search

CASE STUDY HIGHLIGHTS

Boost search performance

- Reduce query response time from 20 seconds to less than a second by migrating from Postgres to Elasticsearch
- Enable much more complex, in-depth queries
- Deliver indexing of up to 1,000 events per second automatically

Scale to handle massive growth

- Grow from 10 million to 110 million events
- Manage data set that expands by 250K events per day
- Take on larger customers with 500x more events

Making sense out of 110 million events

Datadog is a SaaS (Software-as-a-Service) monitoring service for IT, operations and development teams. Datadog enables them to turn the massive amounts of data produced by their applications, tools and services into actionable insights, and ultimately to keep their applications and services up and running. Data is sent to Datadog directly from the customer's infrastructure and through services like GitHub. Datadog aggregates the data, analyzes it, and delivers it back to the customer through an easy-to-use monitoring dashboard accessible via a web browser.

Datadog collects two types of data: metrics and events. The metrics are time-series data, and the events are incidents occurring across the customer's infrastructure. Both types of data are critical to providing visibility into IT operations. Datadog has always been able to provide metrics data to customers, but before deploying Elasticsearch, the company faced major challenges in providing access to their events data.

With a database of 110 million events from customers, growing at a rate of 250,000 events per day, search is an essential function for the Datadog service. In many cases the customer must look into the history of what has happened in their infrastructure in order to identify and solve a performance problem, but without full text search there is a limit on what customers can query. Without the ability to search, these millions of events became useless.

"Initially, we had all the events stored in a Postgres database," recounts Conor Branagan, Software Engineer at Datadog. "We had no ability to do any search beyond simply querying the database, and as searches became more complex, the queries became too slow. We do a lot of rolling up of the events, and it was creating a bottleneck – and that was even before we had so many - we're now at 100 million events. At that point we started having troubles, we were at 10 million events, and it was already too slow. I can't even imagine what it would be like now."

"Updating was also an issue for Datadog," he continues. "When the old database was indexing, it would become very slow, and it would make everything else slower. If it is slowing down on Write, then it is going to slow down Read, and we would get behind on processing data. At that point, we were losing the real-time aspect of the data."

"As we were getting more and more customers, the number of events was very high and getting higher," Branagan adds. "We needed to change our search technology because it wasn't going to scale."

Elasticsearch makes data actionable

To solve the problem with events, Datadog chose Elasticsearch. Today, all the customer events are stored and indexed in Elasticsearch. When customers login to Datadog, the standard system view shows the last 30 events of the week, delivered as a query through Elasticsearch, and a timeline-style bar graph, which is built using the histogram facet in Elasticsearch.

"So Elasticsearch is both querying a specific term and rolling up and presenting data over a longer period of time, which is very useful to our customers because they can quickly see the events at a high level," says Branagan. "Instead of us having to do the heavy lifting of rolling up that data, we are using Elasticsearch."

Boosting query performance by 20x

"Performance, specifically speed of queries, was the main reason we moved to Elasticsearch in the first place," Branagan says. "In Postgres, an average query was 10 or 20 seconds. Now a query takes a second or less with Elasticsearch, and that is including all the rolling up of events. "

"Customers are definitely happy with the performance," he continues. "The increase in performance is a huge advantage for our customers, because they are often in the middle of an issue and they need to know what is going on now, not 20 seconds later when the query responds."

Real-time visibility is an important aspect of the Datadog product, so updating the events data on a timely basis is also vital. Before Elasticsearch, however, there was a limit to how often Datadog could update. The query response time was already slow, and if Datadog conducted constant updates it would hurt performance even further.

"Now we are able to update more often with Elasticsearch," Branagan confirms. "Elasticsearch gives us much higher performance, with the ability to index hundreds or even a thousand events per second. The data comes in fast, and our customers are able to get true real-time visibility into events."

Scalability to handle larger customers with 500x more events

"The biggest advantage Elasticsearch gives us is scalability," says Branagan. " We were at a point before where we were growing so rapidly that we were afraid of the growth. Now we have no fear of our growth. We know that we can extend Elasticsearch easily – we can simply provision new nodes into the cluster."

"Elasticsearch has enabled us to take on bigger customers, which is a big advantage," Branagan concludes. "Before Elasticsearch, a large customer might send in 100 requests per day. Now a large customer is 5,000 to 10,000 requests per day. That is 500 times the number of events. It is a massive difference, and without Elasticsearch we would not be able to handle that."

Datadog's benefits using elasticsearch

- **Increased query performance**
  Elasticsearch has dramatically improved the speed of Datadog's response time for queries on events.
- **True real-time visibility**
  Because Elasticsearch can continuously update data, Datadog can provide customers with real-time visibility into the events from across their IT infrastructures.
- **Improved customer satisfaction**
  Datadog customers are much happier with the system now that Elasticsearch has improved query speed, accelerated updating and expanded the search capabilities.
- **Scalability to handle large customers**
  Elasticsearch has given Datadog the capability and confidence to take on new customers 500 times larger than before.
