BAI Communications Applies Large-Scale Log Analytics to Optimize Performance Across Some of the World's Largest Wireless Networks

LOG ANALYTICS AT BAI COMMUNICATIONS

BAI Communications (BAI) is a leading communications infrastructure provider focused on delivering the next generation of connectivity solutions for transit operators, governments, broadcasters, and mobile network operators (MNOs). BAI specializes in designing, building, and operating communications infrastructure in some of the world's most challenging environments.

Bringing robust connectivity to a metropolitan transit system is incredibly complex and enabling technology doesn't happen overnight. The complexity of BAI's projects require specialized proficiency in high capacity, high availability, multi-user communications networks.

BAI invests in a consistent IT profile across all of its solutions, which range from broadcast transmissions in Australia to delivering services to MNOs in markets such as Canada, Hong Kong, Australia, the UK and across the US. Keeping up with the demands of each business region, as well as constant wireless technology innovations, is no small feat. A big part of adapting to this level of change is removing the analytics team's constraints around data storage, retention and log analytics—allowing them to discover trends and make continuous improvements across a diverse range of projects.

At the heart of BAI's log analytics strategy sits ChaosSearch—the data lake platform that allows BAI to retain more than a year's worth of log data for long-term analysis, enabling improved network performance while maintaining regulatory compliance, at 0.1% of the cost of other leading technology stacks.

bai communications

Industry
Communications

Location
Global

Use Cases
- Long-term network performance analysis
- Troubleshooting network performance issues; forensic analysis
- Long-term security analysis, complementing SIEM solution

Impact
- Up and running within a day
- 99.9% cost savings over legacy Elasticsearch stack
- No data movement or duplication needed for full text search or analytics on log data
- No unplanned downtime; no clusters to manage

Scale
- Analytics on a year of log data, vs. 2-3 weeks
- The Toronto network alone logs ~10,000 events per second, amounting to about 50GB per day

THE CHALLENGE: UNLOCKING LONG-TERM, DIVERSE ANALYTICS IN A COST-EFFECTIVE MANNER

The biggest challenge for the BAI analytics team is also its biggest resource. BAI projects operate at massive scale, and each region has diverse requirements. For example, managing a nationwide broadcast signal is completely different from delivering a high-quality subway system Wi-Fi network. Even so, each business unit has insights locked in long-term data that can help improve services across the entire organization.

"Our analytics team likes to look beyond real-time data. While in-the-moment knowledge of our networks is crucial, you need to change how you design and write software to anticipate changing needs," said Jeremy Foran, head of Data Analytics at BAI. "Instead of just solving one problem and being 'done,' you need to consider how your customers' needs will change in the next week, quarter, year, and beyond."

Goal: Maintain Searchability within Log Analytics

Before adopting ChaosSearch, the team found itself constrained by the lack of cost-effective storage for its volumes of long-term data. The team had to store massive volumes of log data both on-site and in the cloud for both Payment Card Industry (PCI) and ISO 27001 security compliance purposes, which required retention periods of at least one year.

In an interview with Datanami, Foran said, "We had to go from what we needed operationally, maybe two or three weeks' [worth of data], to a year-plus. The cost of having that much logging went up dramatically. We weren't going to be able to afford to buy all of those disks." [1]

The team processes between 10 and 100 gigabytes of data an hour, depending on the day's network activity. Saving the data in a data warehouse would be cost-prohibitive, but locking it away in cold storage would make it nearly impossible to query, making it difficult to parse long-term trends across its global network infrastructure.

When the COVID-19 pandemic hit, the need to query this data and track trends via dashboards and reports became even more urgent. "We didn't know the value of long-term log retention, until there was a long-term unprecedented trend that hit the world," said Foran. "For example, storing data for multiple years helps you evaluate the performance of the network before COVID, and understand the impacts of a global pandemic on geographic regions around the world."

Goal: Reduce Long-Term Data Retention Costs

When evaluating whether to build or buy a system to store and query long-term data, the BAI team considered multiple options. Some solutions would require the team to learn an entirely new set of terminology and processes and for its data pipelines and were far too costly for the organization's long-term data retention needs.

The team wanted something simple that they could use with existing, low-cost storage options, such as Amazon Simple Storage Service (S3) buckets. Instead of implementing a massive volume of solid-state drives (SSDs) to write logs, the team needed a simpler and more cost-effective solution that would keep cloud infrastructure in place for availability and geodiversity across markets.
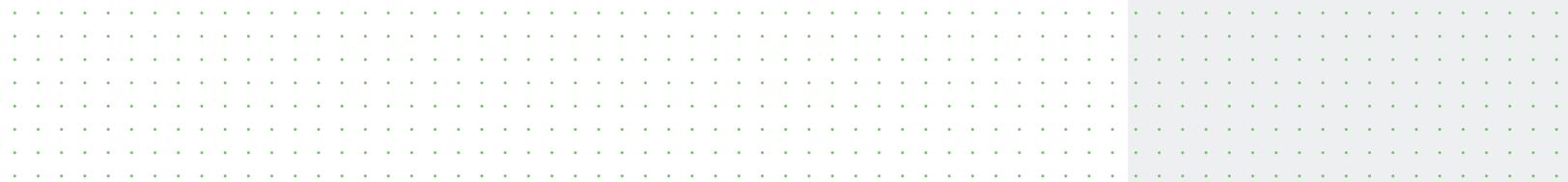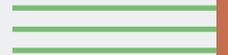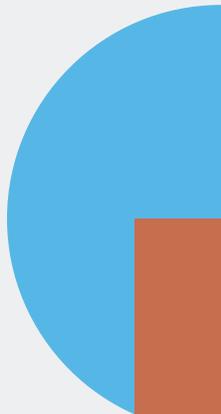
Users
- Infrastructure and IP Network Operations
- Analytics and Network Forensics teams

Data Management Environment
- Data Lake Platform: ChaosSearch
- Cloud Object Storage: Amazon S3
- Analytic Tool(s): Kibana
- Application Monitoring and Observability Tools: Logstash

[1] https://www.datanami.com/2021/08/11/log-storage-gets-chaotic-for-communications-firm

SOLUTION: CHAOSSEARCH DATA LAKE PLATFORM FOR LOG ANALYTICS AT SCALE

The ChaosSearch Data Lake Platform quickly became a critical part of the BAI team's log analytics strategy. The team deployed ChaosSearch on top of its global Amazon S3 data lake in less than one day. ChaosSearch's service allowed BAI to tap into the industry-leading scalability, data availability, security, and performance provided by Amazon S3 and simultaneously take advantage of the exceptional security, governance and granular data access control provided by the Identify and Access Management (IAM) functionality. What's more, ChaosSearch integrates seamlessly with the organization's existing analytics infrastructure, including a security information and event management (SIEM) solution for threat detection logs.

This fully managed, highly available service plugs directly into cloud object storage, delivering a cost-effective option that applies patented indexing technology to activate Amazon S3 as a hot analytics environment.

"ChaosSearch is fundamental within any long-term retention strategy," said Foran. "Anyone can capture a small amount of data within a small window of time. But when you ramp up the amount of data and store it for a very long time, ChaosSearch is the best option. It's like an S3 bucket, but better."

Value Above and Beyond Low-Cost Storage

While the BAI team started with an immediate need to store large volumes of data, they quickly realized value above and beyond low-cost storage alone. The analytics team can also work within its existing ChaosSearch indices to create Kibana dashboards and visualizations.

"We deal with so many devices coming in and out of our network hourly. There are no best practices guides for operating a network of this scale," said Foran. "As a result, we need to have a lot of insight into our network operations, so we can do forensic analyses on performance issues around the world. ChaosSearch helps us easily perform this level of log analytics without data movement, so we can learn from global trends to fully optimize the delivery of our networks."

While no one plans on having to operate through a crisis, it's important to have a contingency plan in place in the event of a network outage or high latency. With ChaosSearch, if something happens to a network, the team can easily pull historical data logs and perform deep analysis.

> "It's an order of magnitude difference. It's the difference between getting an Uber and buying a car."

Jeremy Foran
Head of Data Analytics at BAI
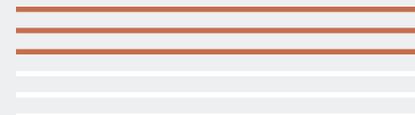
THE IMPACT: BIG DATA ANALYTICS AT A FRACTION OF THE COST

Today, ChaosSearch helps the team store and query long-term data at 0.1% of the cost of other leading technology stacks.

"At the end of the day, if we wanted to put this in another tech stack, it would have cost tens of thousands of dollars per month," Foran said. "If we went to put it in ChaosSearch… It's an order of magnitude difference. It's the difference between getting an Uber and buying a car." [1]

"Because ChaosSearch is cost effective, we feel quite comfortable applying it to new projects. So I have no reason not to use ChaosSearch on our next data pipeline," said Foran.

In the future, the team will be using ChaosSearch as a part of its data retention and analytics strategy for many of its other large-scale projects with long-term data retention requirements.

The team also plans to evaluate upcoming ChaosSearch SQL API integration capabilities to tap into business intelligence platforms like Tableau. These tools will help the team further investigate network performance based on Internet Protocol Flow Information Export (IPFIX) or NetFlow data. In addition, the team is looking forward to seeing the addition of anomaly detection on top of their existing ChaosSearch platform (via machine learning APIs coming in 2022).

"One of the most exciting things about ChaosSearch is that it removes constraints. We're able to cost-effectively store all of our data, run queries against it, build dashboards and gain critical insights. We can do all of these projects in ChaosSearch and use it for even more future requirements we haven't realized yet," said Foran.

[1] https://www.datanami.com/2021/08/11/log-storage-gets-chaotic-for-communications-firm

> "Because ChaosSearch is cost effective, we feel quite comfortable applying it to new projects. So I have no reason not to use ChaosSearch on our next data pipeline."

Jeremy Foran
Head of Data Analytics at BAI

To begin your free trial of ChaosSearch, visit **chaossearch.io**

ABOUT CHAOSSEARCH

ChaosSearch helps modern organizations Know Better™ by activating the data lake for analytics. The ChaosSearch Data Lake Platform indexes customers' cloud data, rendering it fully searchable and enabling analytics at scale with massive reductions of time, cost and complexity.

ChaosSearch was purpose-built for cost-effective, highly scalable analytics encompassing full text search, SQL and machine learning capabilities in one unified offering. The patented ChaosSearch technology instantly transforms your cloud object storage (Amazon S3, Google Cloud Storage) into a hot, analytical data lake.

For more information, visit ChaosSearch.io or follow us on Twitter @ChaosSearch and LinkedIn.

info@chaossearch.com | www.chaossearch.io