Success Story

Modernizing Threat Intelligence at Scale with VAST DataBase



VAST

A leading cybersecurity and threat intelligence company reached a critical inflection point with its data storage and analytics infrastructure. With over 2 billion malware and PCAP files requiring fast, concurrent access for analysis and detection, their legacy systems were beginning to face challenges. The IT team needed a solution that could efficiently handle massive data volumes, support high-performance analytics, and simplify operations.

Their infrastructure, which relied on legacy storage and Elasticsearch clusters, had become increasingly difficult to manage. As the data footprint grew to 2 petabytes, managing disparate data lakes and the complex Elasticsearch environment created mounting challenges. Several key pain points emerged:

- **High operational complexity:** Managing large, fragmented datasets across different storage platforms and dealing with frequent scaling issues in Elasticsearch created significant overhead.
- **Costly cloud queries:** Relying on cloud providers for data queries led to high query costs, deterring analysts from fully utilizing the available data.
- **Inefficient data management:** The lack of a centralized platform for managing both structured and unstructured data made holistic data searches and analyses difficult.

Seeking a more modern and efficient solution, the organization turned to VAST Data to consolidate and enhance its threat research infrastructure. By implementing the VAST AI Operating System, they streamlined data management and enabled more powerful analytics. One engineer explained: “What enticed me about VAST is that it’s not just some behemoth that is picked up and dropped on top. It’s something that is new and well thought out, void of technical debt.”

> *What enticed me about VAST is that it’s not just some behemoth that is picked up and dropped on top. It’s something that is new and well thought out, void of technical debt.*
>
> **- Engineer, Anonymous Cybersecurity and Threat Intelligence Company**

Key Advantages

Consolidated Data Infrastructure: VAST enabled the organization to unify its data management, integrating both file and object with a multi-view system that supports NFS and S3 simultaneously. This removed the need for multiple storage solutions and reduced operational complexity.

Improved Data Reduction and Cost Efficiency: With advanced data reduction delivering a 5:1 ratio, VAST allowed the team to store and access large volumes of data more cost-effectively. They also replaced their existing storage with VAST at a lower upfront cost than legacy HDD-based solutions.

Enhanced Performance for Threat Detection: VAST's ability to handle high concurrency and rapid read/write operations proved essential for data-intensive applications, including malware analysis and Splunk-based querying. VAST's architecture, which decouples compute from storage, ensures scalable performance as data and analytics needs grow.

Operational Simplicity and Scalability: VAST's simplified data management significantly reduced the burden on IT. Adding capacity or compute resources is now straightforward, allowing the organization to scale infrastructure easily as data grows. The new, streamlined maintenance process is far less labor-intensive than it was with Elasticsearch and the legacy storage vendor.

The Implementation Process

The team deployed VAST across its on-premises data centers, migrating seven years' worth of critical threat research data. They then gradually transitioned and integrated more applications, including Splunk and other search tools, onto the VAST DataBase. One engineer described the process: "We moved our existing data set one-to-one from our legacy provider onto VAST. VAST was less expensive, reduced our footprint, and simplified management—eliminating the operational headaches of our previous system."

The results were transformative. The organization saw significant improvements in data reduction, cost savings, and operational efficiency. The VAST AI Operating System enables them to continue advancing their cybersecurity mission while providing a foundation for future growth. As another team member noted: "I'm building out tools to make research better. Querying massive amounts of data at scale is the primary thing we were after, and VAST allows us to build solutions that make researchers' work easier."

> *Running queries in the cloud is not cheap, and it scares people away from trying to query data. It's a bad combination—here's all the cool data, but you can't touch it because it's really expensive to look at.*
>
> - **Engineer, Anonymous Cybersecurity and Threat Intelligence Company**

VAST DataBase Deployment

The company's Elasticsearch-based on-prem stack had become increasingly constrained by limitations in scale and performance. As data volumes grew, so too did the complexity of managing fragmented, sharded clusters. To address this, the team began exploring the VAST DataBase as a more efficient way to catalog and query their structured and unstructured data. "We were looking to displace some databases we had in Elastic, and the debut of the VAST DataBase hit just right," said one engineer. With free queries and no additional licensing fees, the VAST DataBase simplified their cost model compared to previous Elastic and AWS-based solutions.

Their goal was to streamline operations, centralize analytics, and eliminate the pain points of their legacy setup. The deployment involved migrating threat research workloads from the legacy ELK Stack to the VAST AI Operating System, with a strong focus on deploying the VAST DataBase for structured and semi-structured data analysis. One engineer pointed out: "Running queries in the cloud is not cheap, and it scares people away from trying to query data. It's a bad combination—here's all the cool data, but you can't touch it because it's really expensive to look at." For datasets of this size, cloud queries typically reached the mid-five figures per query. By contrast, users running queries on the VAST AI Operating System avoid surprise costs entirely.

One infrastructure engineer recalled, "With Elasticsearch, every new sharded index added complexity and instability. Searches could time out or produce inconsistent results. The scalability just wasn't there. VAST, by contrast, allows us to keep everything in one place, with no need to break the data apart into different silos. It's truly a scalable data platform solution that adapts as our needs grow."

Another team member shared: "Migrating to the VAST DataBase was the clear choice for handling our structured queries that Elasticsearch previously managed. It freed up operational time and compute resources, without any added costs beyond storage. Now, we can focus on consolidating more data sources onto VAST. It is transforming how we manage data at scale."

> *Migrating to the VAST DataBase was the clear choice for handling our structured queries that Elasticsearch previously managed. It freed up operational time and compute resources, without any added costs beyond storage. Now, we can focus on consolidating more data sources onto VAST. It is transforming how we manage data at scale.*
>
> **- Engineer, Anonymous Cybersecurity and Threat Intelligence Company**



Core Architectural Benefits and Innovations

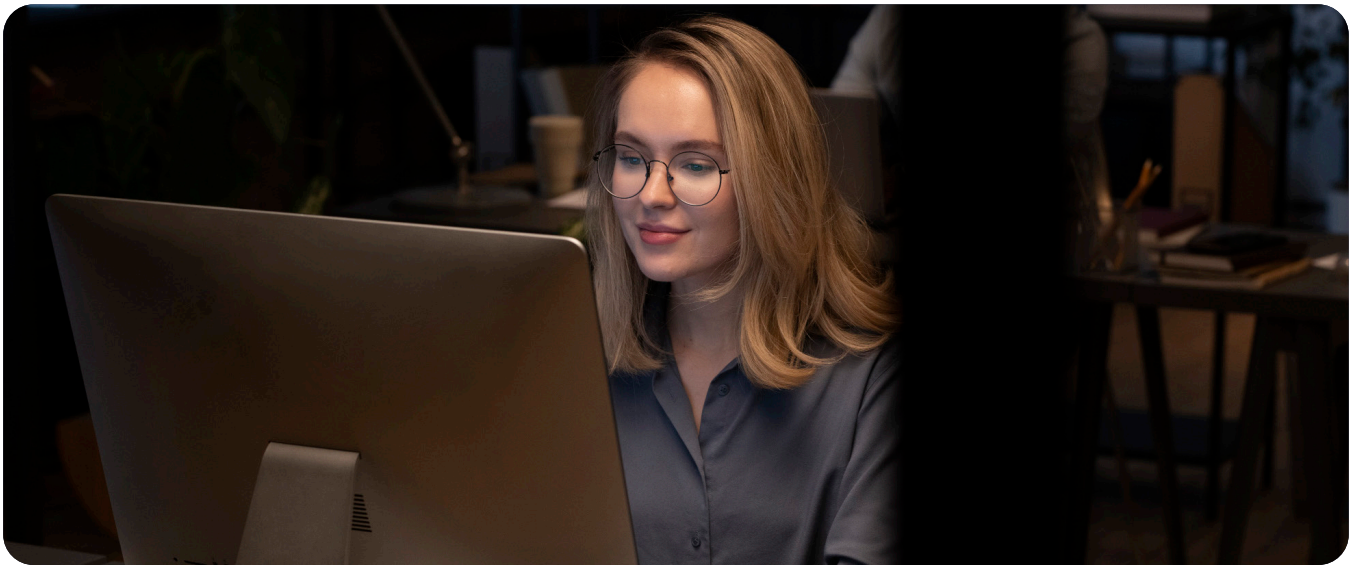
Advanced Threat Analysis with Structured Ingestion: Using the VAST DataBase, structured metadata from a third-party threat intelligence API is ingested and indexed by various identifiers such as threat ID, fingerprint, and contextual metadata. The data is stored in a columnar format optimized for filtering and accessed through both a custom-built web portal and VAST's Python SDK. This ingestion pipeline allows for real-time correlation of new intelligence with historical patterns and enables deep longitudinal analysis.

Unsharded, Exabyte-Scale Data Platform: Unlike Elasticsearch, which required manual sharding to manage large datasets, VAST's architecture enables storage of billions of records in a single logical table. This design reduces administrative overhead and improves query efficiency across petabytes of data. By eliminating delays introduced by distributed queries across multiple shards, it also improved consistency of results across repeated queries.

Optimize Cloud Costs: While the organization still leverages Amazon Athena for analytics, deploying the VAST DataBase has provided greater control over analytics costs. With no high-cost query paths for the threat intelligence deployment, analysts can query more frequently and experiment with new detection models, without fear of triggering massive cloud bills.

Multi-Interface Access for Varied User Needs: To support different analyst workflows, the company deployed a dual-access model. Threat researchers access data through a web portal, while others, such as network operations teams, use Splunk. Splunk connects to Trino using JDBC, and Trino retrieves relevant records from VAST DataBase. This hybrid model provides performance and flexibility without duplicating data, allowing the organization to leverage their existing tools while reducing their reliance on expensive proprietary ingestion pipelines.

Smart Filtering at the Source: One of the technical strengths of the VAST DataBase is its efficient filtering mechanism. When a query is executed via Trino, VAST's compute nodes (CNodes) perform the filtering locally, ensuring only the necessary data is transmitted over the network. This minimizes load on downstream analytics tools and boosts query responsiveness. This feature also significantly reduces time-to-insight and has enabled the development of automated alerting mechanisms based on query results.



A Real-Time, Scalable Threat Research Deployment

With VAST, analysts can search across years of historical data, correlating newly received indicators with past activity to uncover patterns and long-term threats. This long-range visibility is crucial in tracking persistent threat actors and coordinating proactive defense strategies.

To support rapid analysis, the organization also designed a user-friendly web portal that enables filtering by parameters such as timestamp range, threat fingerprint, and specific attack signatures. This accelerates response times and empowers threat researchers to pivot quickly when new discoveries emerge. The portal also supports custom visualizations, trend graphs, and exportable reports that improve cross-functional communication between security operations, IT, and executive leadership.

As one engineer put it, "We've eliminated the delays that used to plague our Elasticsearch stack. Now we're operating at a different level. We can look back across years of data in seconds, without worrying about system failures or cost overruns. We're able to work the way we've always wanted to—without compromise."

A Foundation for Growth and Innovation

The migration to VAST wasn't just an infrastructure update—it marked a shift in operational philosophy. Rather than managing a patchwork of disconnected systems, the organization now relies on a centralized, high-performance data platform that scales linearly with demand. This transformation has freed both the security and data teams to think bigger: expanding use cases, developing internal tools, and streamlining collaboration across departments. The newfound operational freedom has sparked innovation at every level of the organization, from junior analysts building new dashboards to senior architects designing AI-driven detection pipelines.

One engineer observed, "We've seen a lot of vendors come and go. VAST brings something fundamentally different—a blend of performance, scale, and openness that we haven't found anywhere else. It's not just a product—it's a platform that unlocks what's possible."

About VAST Data

VAST Data is the data platform company built for the AI era. As the new standard for enterprise AI infrastructure, organizations trust the VAST Data Platform to serve their most data-intensive computing needs. VAST Data empowers enterprises to unlock the full potential of their data by providing AI infrastructure that is simple, scalable, and architected from the ground up to power deep learning and GPU-accelerated data centers and clouds. Launched in 2019, VAST Data is the fastest growing data infrastructure company in history. For more information, please visit and follow VAST Data on X (formerly Twitter) and LinkedIn.

Scan here to learn more.

> *We've eliminated the delays that used to plague our Elasticsearch stack. Now we're operating at a different level. We can look back across years of data in seconds, without worrying about system failures or cost overruns. We're able to work the way we've always wanted to—without compromise."*
>
> - **Engineer, Anonymous Cybersecurity and Threat Intelligence Company**

©2025 VAST Data, Inc. All rights reserved. All trademarks belong to their respective owners.
06/2025