Elastic vs. Splunk
A Strategic Comparison for Enterprise Data Management

Patrick M. Hayes
Chief Strategy Officer

Shane Thorson
Chief Technology Officer

THIRD WAVE
Innovations

October 2024

Notice

Third Wave Innovations, LLC (Third Wave) has made every reasonable attempt to ensure that the information contained within this whitepaper is current as of the date of publication. Readers acknowledge and agree that Third Wave assumes no responsibility for errors that may be contained in or for misinterpretations that readers may infer from this document.

Trademark & Copyright Notice

Third Wave Innovations, the Third Wave logo, and other trademarks, service marks, and designs are registered or unregistered trademarks of Third Wave Innovations, LLC in the United States and in foreign countries. © 2024 Third Wave Innovations, LLC All Rights Reserved.

Table of Contents

Abstract

This white paper offers a thorough comparison of Elastic and Splunk, both recognized as prominent enterprise data management platforms. Our analysis focuses on essential aspects such as system architecture, scalability, data lifecycle management, query languages, artificial intelligence and machine learning capabilities, ecosystem extensibility, and pricing structures.

One of Elastic's standout features is its distributed architecture, which allows for greater flexibility when managing large datasets. This design enables businesses to scale horizontally, accommodating growth without the constraints often seen in more traditional systems. In contrast, Splunk uses a centralized approach, which may limit its adaptability as data volumes fluctuate. While Splunk is still a strong option, its architecture may not support the rapid changes in data environments that many organizations face today.

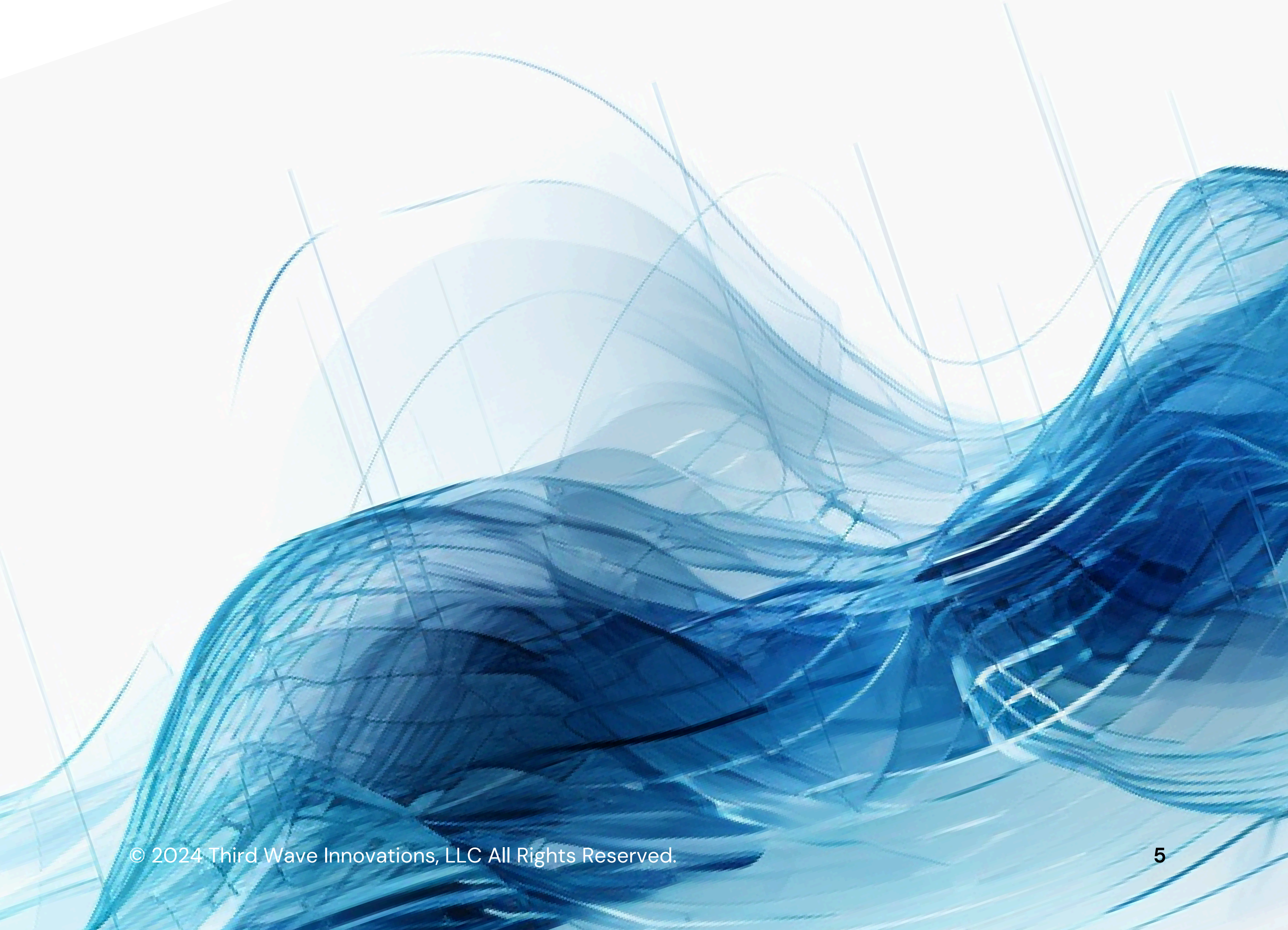
Another critical factor in our evaluation is the ability of each platform to integrate artificial intelligence (AI) and machine learning (ML). Elastic excels in this area, offering advanced capabilities such as natural language processing and vector search, which enhance its data analytics functionalities. These features make Elastic particularly appealing to organizations aiming to leverage the latest technology in their data strategies. Splunk does provide AI and machine learning tools, but they are not as deeply integrated or as innovative as those offered by Elastic, potentially limiting their effectiveness.

COMPARISON FOCUS

- System Architecture
- Scalability
- Data Lifecycle
- Management
- Query Languages
- Artificial Intelligence
- Ecosystem Extensibility
- Pricing structures

Cost considerations also play a significant role in the decision-making process. Elastic's open-source foundation generally leads to a more favorable pricing model, which can be especially beneficial for businesses looking to manage their expenses. In contrast, Splunk tends to have higher costs, particularly as organizations scale their data operations.

By understanding the strengths and weaknesses of each platform, IT leaders can make informed choices that align with their specific needs and long-term goals in data management. This paper looks to provide decision-makers with the insights they need to navigate the complexities of both platforms.



Introduction: Divergent Approaches

Elastic and Splunk are two of the major enterprise players in data analytics, observability, and security information event management. Each offers distinct ways to handle large-scale data. Ultimately, understanding each's strengths is key to leveraging data effectively. Whether opting for Elastic's flexibility and scalability or Splunk's out-of-the-box functionality, the right choice can significantly impact how an organization uses its data.

Elastic's architecture for data management involves a distributed, scalable approach using Elasticsearch, which handles both structured and unstructured data. It employs a tiered storage system to manage data efficiently across hot, warm, cold, and frozen tiers, optimizing for cost and performance. Elasticsearch clusters consist of nodes that provide horizontal scalability and fault tolerance, with data being distributed across shards and replicas for redundancy.

In Elasticsearch, data streams manage append-only time series data across multiple indices while providing a single named resource for requests. They are ideal for continuously generated logs, events, and metrics.

Data streams automatically route indexing and search requests to the appropriate backing indices, which are hidden and auto-generated. They facilitate efficient data management by using index lifecycle management (ILM) to automate the rollover of indices and transition them through different data tiers for cost optimization.

Index Lifecycle Management (ILM) integrates with data streams in Elasticsearch by automating the management of backing indices, which store the stream's data. ILM policies define actions such as rollover, retention, and deletion based on index age or size. When a data stream's write index meets rollover conditions, ILM creates a new backing index, ensuring efficient data organization and cost management. This integration allows for seamless transitions of data through different storage tiers, optimizing resource use while maintaining data accessibility.

Splunk has built a reputation for out-of-the-box operational intelligence, security information, and event management (SIEM) functionality. Its platform transforms machine-generated data into insights based on pre-built functionality that teams can act on. Splunk's search processing language (SPL) provides query capabilities, and its ability to correlate events from different sources in real-time makes it useful for cybersecurity and IT operations, helping detect threats and respond quickly.

Splunk's architecture is designed to efficiently handle large volumes of machine-generated data by collecting, processing, storing, and analyzing it. The system comprises three main components: forwarders, indexers, and search heads.

Forwarders are lightweight agents installed on data sources to gather and send data to Splunk indexers. There are two types of forwarders: universal forwarders, which collect and transmit raw data without processing, using minimal system resources. On the other hand, heavy forwarders can parse and process data before sending it along, which helps reduce the workload on indexers.

Indexers form the backbone of Splunk's infrastructure. They receive data from forwarders, process it into searchable events, and store it in indexes. They also compress the data and create efficient indexes to enable fast search capabilities.
Search heads act as the user interface, allowing users to search and analyze the data. They manage search requests by distributing them to indexers and consolidating the results for display.

Splunk's architecture is typically organized into distinct tiers. The collection tier consists of forwarders that gather data from various sources. The indexing tier comprises indexers that process and store this data. The search tier includes search heads for querying and visualizing the stored information. Additionally, a management tier controls overall system configuration, including components like deployment servers and cluster managers.

Both Elastic and Splunk integrate well with a wide range of data sources and third-party tools, but they differ greatly in deployment, integration, and pricing. Elastic's open-source model gives it more flexibility and allows for greater customization. Elastic's commitment to open telemetry development at its core gives it significant data processing power. Splunk offers a more polished user experience but can become expensive for larger implementations, especially when looking for capabilities that extend beyond pre-existing functionality.

Choosing between the two depends largely on the organization's needs. Elastic is often chosen for its capability to cost-effectively scale in diverse data environments, from operational data to security event logs and external data enrichment sources. Large enterprises especially value Elastic's data processing and ability to customize. Splunk, with its focus on out-of-the-box operational functionality, is favored for IT operations, security log analysis, and compliance.

Both platforms are evolving as data grows in complexity and volume. They are enhancing machine learning capabilities, expanding cloud offerings, and improving user interfaces to make advanced data analysis more accessible.

Elastic: The Distributed Search and Analytics Engine

At its core, Elastic functions as a distributed search and analytics engine optimized for rapid querying and analysis of extensive datasets. Its architecture is fundamentally designed to deliver real-time insights, making it an invaluable asset for organizations grappling with structured and semi-structured data. The versatility of Elastic extends its applicability across a broad spectrum of industries and use cases, from powering sophisticated content management systems to driving operational intelligence in high-volume data environments. Elastic is ideal for organizations with complex infrastructures requiring integration across multiple endpoint detection and response (EDR) platforms and threat intelligence enrichments.

Splunk: The All-Encompassing Platform for Machine Data

Splunk positions itself as an all-encompassing platform for real-time monitoring, search, and analysis of machine-generated big data. Its log management and security analytics capabilities are meant for environments where system logs and security event data are important. This makes Splunk suitable for organizations that find the out-of-the-box functionality suitable for their environments and require little customizations or need more technical staff that can build out customizations.

Conclusion

While both Elastic and Splunk offer robust solutions for data analytics and management, Elastic emerges as the more versatile and scalable option, particularly for organizations with complex infrastructures and diverse data environments. Its distributed, open-source architecture provides the flexibility to handle both structured and unstructured data efficiently, while the tiered storage system optimizes cost and performance. The ability to customize and seamlessly integrate with multiple data sources makes Elastic an ideal choice for enterprises that need to scale cost-effectively without being tied to expensive, pre-built functionalities. For organizations looking to future-proof their data management strategies with a platform that can adapt and grow with their needs, Elastic proves to be the superior option

DIVERGENT APPROACHES COMPARISON

ELASTIC

- Offers flexibility and scalability, with a distributed, scalable architecture.
- ·Handles both structured and unstructured data and employs a tiered storage system for cost-efficient data management.
- Favored by organizations needing to scale across diverse data environments and requires customization for complex infrastructures.
- Open-source model allows for more customization and cost-effective scaling,
- Suited for environments with complex data streams.

SPLUNK

- Its strength lies in real-time monitoring and analysis of machine-generated data.
- Provides out-of-the-box operational functionality, making it easy to use for quick deployment with minimal customization.
- Interface offers a user-friendly experience but can be more expensive for larger deployments.
- Pre-built capabilities that require less technical staff for implementation.

Usability and Learning Curve

Elastic: Flexibility at the Cost of Complexity

While offering greater flexibility and customization potential, Elastic generally demands a higher level of technical proficiency for effective utilization. This increased complexity can initially result in a steeper learning curve, potentially extending the time to value for organizations adopting the platform. Elastic also has the ability to use Business Data and Logic to influence detections, actions, and reporting. This can be said across the board with the ability to use any relevant data in enrichment data to influence.

Elastic, typically used with Kibana for visualization, focuses on data discovery and exploration. It includes pre-built data visualizations and analysis tools like the Discover, Visualize, and Dashboard apps. Elastic provides more flexibility for customization, which can be tailored more precisely to specific needs with some additional effort.

Splunk: Intuitive Interface and Pre-Built Features

From a usability perspective, Splunk has a reputation for its intuitive interface and streamlined deployment process. If managed in-house, this user-friendly approach can reduce the time and resources required for initial setup and ongoing management. However, it is reported to be fairly rigid and costly. This makes Splunk suitable for organizations that find out-of-the-box functionality suitable with little customizations or lack the technical staff to build out customizations.

Conclusion

While both Elastic and Splunk offer powerful solutions for data management and analysis, Elastic stands out for organizations seeking greater flexibility and customization. Although it demands a steeper learning curve and higher technical proficiency, Elastic’s ability to leverage business data, logic, and enrichment for tailored insights makes it a more versatile platform. For teams willing to invest in technical expertise, Elastic offers a highly customizable and scalable solution that can be finely tuned to specific organizational needs, making it the preferable choice for those prioritizing long-term adaptability over ease of initial setup.

USABILITY AND LEARNING CURVE COMPARISON

✓ ELASTIC

- Offers greater flexibility and customization potential.
- Requires higher technical proficiency, leading to a steeper learning curve.
- Uses business data and logic to influence detection, actions, and reporting.
- Flexible in using enrichment data for tailored solutions.
- Focuses on data discovery and exploration with tools like Discover, Visualize, and Dashboard apps (usually paired with Kibana for visualization)
- Customization is possible but demands additional effort.

✓ SPLUNK

- Splunk has a moderate learning curve, especially when it comes to building expertise for more specialized analyses.
- Its proprietary Search Processing Language (SPL) may take some time to master for complex queries.

Data Ingestion and Processing

Elastic: Versatility in Data Processing

While highly efficient in processing data, Elastic may require more upfront configuration and schema design to handle unstructured data sources effectively. It is important to note how Application Performance Monitoring (APM) and OpenTelemetry integration offer substantial benefits as part of Elastic's data processing ecosystem.

APM provides continuous and detailed insight into application performance, enabling teams to improve stability, reduce incidents, and resolve issues faster. By monitoring key metrics like response time, throughput, and error rates, APM helps identify and diagnose performance issues quickly, enhancing user experience and faster root cause analysis. APM solutions also offer powerful machine learning and analytics capabilities that can automate anomaly detection and correlate APM data with other observability data, significantly reducing mean time to detection and resolution for issues.

OpenTelemetry integration brings standardization and flexibility to data collection across different services and applications. By supporting OpenTelemetry, Elastic allows organizations to avoid vendor lock-in and future-proof their observability investments. The seamless integration of logs, metrics, and traces from various sources into the Elastic Stack enables a unified view of all observability data. This standardized approach facilitates better collaboration between development and operations teams and allows for efficient collection, processing, and export of large volumes of telemetry data.

Elastic's unique advantages in APM and OpenTelemetry include a unified observability platform that integrates all data types into a single view, powerful search and analytics capabilities for deep analysis, and built-in machine learning features for anomaly detection and predictive analysis. Elastic also offers flexible deployment options, supporting cloud, on-premises, and hybrid environments to meet diverse infrastructure needs. With its commitment to open source, Elastic ensures continuous innovation and community-driven improvements. By leveraging APM and OpenTelemetry within the Elastic ecosystem, organizations can achieve comprehensive observability, improve application performance, and gain valuable insights into their systems' behavior.

Splunk: Efficiency with Structure

Splunk's data ingestion and processing system is designed to handle large volumes of machine-generated data from various sources efficiently. The process begins with data collection, utilizing multiple methods such as the ingest service, forwarder service, HTTP event collector, and various connectors for cloud and streaming data sources. Splunk can accommodate a wide range of data inputs, making it versatile for different organizational needs.

Splunk's indexing system offers several key features, including immutability of indexed data, flexibility in handling various data types, automatic field extraction, and scalability through distributed indexing. Once the data is indexed, it becomes available for querying using Splunk's Search Processing Language (SPL). This enables users to perform complex searches, create visualizations, and build dashboards, facilitating rapid analysis of large datasets. The process transforms raw data into searchable events, allowing organizations to quickly gain valuable insights from their machine data.

Splunk also supports OpenTelemetry but maintains a more distinct separation between It and its proprietary tools. While Splunk offers its own distribution of the OpenTelemetry Collector, its approach seems to be more focused on providing compatibility and support for OpenTelemetry data alongside its existing proprietary agents and data collection methods.

Conclusion

Elastic emerges as the more versatile and forward-thinking data ingestion and processing solution. Its seamless integration of Application Performance Monitoring (APM) and OpenTelemetry within a unified observability platform provides organizations with a powerful toolset for deep performance insights, real-time anomaly detection, and predictive analysis. The open-source nature of Elastic ensures continuous innovation and flexibility, allowing companies to avoid vendor lock-in and future-proof their observability investments. With strong support for diverse deployment environments, Elastic is well-positioned to meet the evolving needs of modern businesses, making it the superior choice for organizations seeking comprehensive, scalable, and flexible data processing solutions.

DATA INGESTION AND PROCESSING COMPARISON

ELASTIC

- Versatile in processing structured and semi-structured data but requires more upfront configuration for unstructured data.
- Offers built-in APM with deep performance insights, machine learning, and automated anomaly detection for faster issue resolution.
- Seamlessly integrates OpenTelemetry, offering a unified observability platform that prevents vendor lock-in and simplifies data collection.
- Known for powerful search and analytics with built-in machine learning for predictive analysis.
- Supports cloud, on-premises, and hybrid environments with open-source and open standards.

SPLUNK

- Efficiently handles large volumes of machine-generated data from various sources, with flexible data input methods.
- Provides robust analytics but separates OpenTelemetry from its proprietary tools, focusing on compatibility rather than full integration.
- Supports OpenTelemetry but keeps it more distinct from proprietary systems, offering its own distribution of the OpenTelemetry Collector.
- Features a robust indexing system, immutability of data, and the Search Processing Language (SPL) for complex queries and visualizations.
- Focuses on distributed indexing and scalability with a wide array of data connectors.

Data Lifecycle Management

Elastic: Distributed Architecture

Elastic's data lifecycle management is built around its distributed architecture and flexible indexing system. Data ingestion offers versatile options for data collection, transformation, and enrichment, with storage leveraging Elastic's distributed nature to support hot-warm-cold-frozen tiers for cost optimization. Index Lifecycle Management (ILM) automates retention policies, allowing seamless migration between tiers. Elasticsearch's Query DSL and Kibana provide robust search and analysis capabilities, while compliance features like document-level deletion and data anonymization further enhance its appeal. The platform's scalability and performance make it ideal for large-scale deployments.

Splunk: Proprietary Indexing

Splunk's data lifecycle management centers around its proprietary indexing system and bucket-based storage model. Data ingestion relies on Splunk Forwarders and the HTTP Event Collector, with options for parsing and transformation during indexing. The storage system uses hot, warm, cold, and frozen buckets, with cloud storage integration and flexible data retention policies, supporting automatic archiving and data roll-up for long-term storage. Data access is facilitated through Splunk's Search Processing Language (SPL) and search interface, which supports real-time and historical analysis with compliance features like role-based access control and audit logging.

Conclusion

Elastic offers a versatile and scalable solution for data lifecycle management, ideal for organizations with large-scale data needs. Its distributed architecture, flexible indexing, and automated Index Lifecycle Management (ILM) optimize costs and performance across data tiers. Advanced data access capabilities, including Elasticsearch's Query DSL and Kibana, provide robust search options, while compliance features like document-level deletion and data anonymization enhance its appeal. For flexibility and comprehensive data management, Elastic stands out as the more robust choice.

DATA LIFECYCLE MANAGEMENT COMPARISON

ELASTIC

- Distributed architecture with a flexible indexing system.
- Offers versatile options for data collection, transformation, and enrichment.
- Supports hot, warm, cold, and frozen data tiers for cost optimization.
- Index Lifecycle Management (ILM) automates data retention and migration between tiers.
- Uses Elasticsearch's Query DSL and Kibana visualization for search and analysis.
- Offers role-based access control and audit logging.
- Includes document-level deletion and data anonymization.
- Scalable and performant for large-scale data management.

SPLUNK

- Proprietary indexing system with bucket-based storage model.
- Uses Splunk Forwarders and HTTP Event Collector for data ingestion with parsing and transformation during indexing.
- Uses hot, warm, cold, and frozen buckets with cloud storage integration.
- Flexible retention policies with automatic archiving to frozen tier and data roll-up.
- Offers Search Processing Language (SPL) and a user-friendly search interface for both real-time and historical analysis.
- Offers role-based access control and audit logging.
- Suited for real-time and historical data analysis with flexible long-term storage.

Scalability: Horizontal vs Vertical Expansion

Elastic: Horizontal Scaling for Flexibility

Elastic excels in horizontal scaling, allowing easy addition of more nodes to a cluster. This is particularly effective for handling large data volumes and complex searches. Horizontal scaling enables organizations to increase their data processing power by simply adding more nodes to their clusters as data volumes grow. This scalable design offers flexibility and long-term cost-efficiency, especially for businesses managing large and ever-expanding datasets. By expanding incrementally, organizations can manage resource demands more effectively without the need for massive upfront infrastructure investments.

To optimize costs, Elastic employs a multi-tier storage strategy that organizes data into hot, warm, cold, and frozen tiers. Each tier is designed to align with specific data access needs and storage costs. The hot tier is dedicated to frequently accessed data, leveraging high-performance hardware such as SSDs to ensure fast read and write operations. The warm tier is used for less frequently accessed data stored on more cost-effective solutions while maintaining reasonable performance.

For infrequently accessed data, Elastic's cold and frozen tiers significantly reduce storage costs by utilizing cheaper options like object storage or snapshots. While access times may be slower in these tiers, the strategy helps to balance overall performance and costs, ensuring that critical data is stored efficiently according to usage patterns. This tiered structure allows organizations to optimize both performance and budget, making Elastic a highly adaptable choice for varying data management needs.

Splunk: Vertical Scaling and Its Challenges

Splunk's vertical scaling approach involves increasing the resources of existing servers or nodes to handle growing data processing and search demands. This method, also known as scaling up, typically involves adding more CPU, RAM, or storage to existing Splunk servers. It's often the first step in scaling Splunk infrastructure due to its relative simplicity and immediate impact on performance. Key aspects of vertical scaling in Splunk include resource expansion, performance boost, and simplified management. By upgrading hardware components on existing indexers, search heads, or other Splunk components, organizations can improve individual nodes' processing power and capacity to handle increased data volumes or more complex searches.

Vertical scaling in Splunk comes with several challenges. There's an upper limit to how much a single server can be upgraded, eventually hitting a ceiling regarding available hardware upgrades. Concentrating more resources on fewer machines increases the risk of a single point of failure, where the loss of a heavily upgraded server can significantly impact the entire Splunk deployment. Upgrading hardware often requires taking the machine offline, leading to service interruptions unless a backup server is available. Moreover, at a certain point, continually upgrading a single machine becomes less cost-effective than adding new servers.

While Splunk can scale horizontally, it typically does so by adding more indexers or search heads, which poses challenges with licensing costs as data volume increases.

Conclusion

Elastic’s horizontal scaling approach provides greater long-term flexibility, scalability, and cost efficiency compared to Splunk's vertical scaling model. With the ability to incrementally add nodes and a multi-tiered storage strategy that optimizes both performance and costs, Elastic allows organizations to seamlessly handle growing data volumes without the constraints of hardware limitations or the risk of a single point of failure. Its adaptable design makes it a more sustainable solution for businesses that require robust, scalable, and cost-effective data management as they grow. While offering immediate performance boosts, Splunk's vertical scaling ultimately faces limitations that make Elastic a more future-proof choice.

HORIZONTAL VS VERTICAL EXPANSION COMPARISON

ELASTIC

- Prioritizes flexibility, scalability, and cost-effective incremental growth.
- Expands by adding more nodes to a cluster.
- Ideal for handling large data volumes and complex searches by increasing data processing power incrementally.
- Long-term cost-effective due to incremental expansion without large upfront infrastructure costs.
- Uses a multi-tier storage model (hot, warm, cold, frozen) to balance performance and costs, optimizing data access based on usage patterns.
- Flexible, scalable, and well-suited for businesses with growing datasets.

SPLUNK

- ·Offers immediate performance improvements but faces challenges with resource limits, downtime, and escalating costs with continued growth.
- Increases resources (CPU, RAM, storage) on existing servers (vertical scaling).
- Improves individual server performance by upgrading hardware components.
- Hits a ceiling with hardware upgrades, risking a single point of failure and making continuous upgrades cost-inefficient.
- Upgrades often require downtime unless backup systems are in place.
- Can scale horizontally by adding indexers/search heads, which adds licensing costs as data volume increases.

Query Languages: Power vs Accessibility

Elastic: Complex but Powerful

Elastic's Domain Specific Language (DSL) offers powerful and flexible search capabilities, allowing complex query constructions to target and analyze data precisely. However, the depth and complexity of the DSL may require substantial training and experience to master.

Elasticsearch Query Language (ESQL) is a new piped query language developed by Elastic to simplify querying, aggregating, and visualizing large datasets with advanced search capabilities and concurrent processing, improving speed and efficiency. Using a familiar pipe syntax, ESQL allows users to perform complex data transformations, enrichments, and calculations in a single query. Integrated with Kibana, it enables seamless data visualization and analysis from a single interface.

ESQL offers significant speed advantages over traditional Elasticsearch queries due to its concurrent processing capabilities and the use of a new query engine. This allows ESQL to handle complex data transformations and aggregations more efficiently, resulting in faster query execution times. ESQL's piped syntax simplifies the chaining of operations, further enhancing performance by reducing the need for multiple separate queries. Additionally, ESQL's ability to execute queries directly within Elasticsearch without translation to Query DSL contributes to its speed and efficiency.

ESQL handles large datasets more efficiently than traditional Elasticsearch queries by leveraging concurrent processing and a new compute engine designed for performance. This allows ESQL to perform complex data transformations and aggregations directly within Elastic without translating to Query DSL, enhancing speed and scalability. ESQL's piped syntax enables chaining multiple operations, simplifying data analysis and reducing query complexity. Additionally, ESQL supports asynchronous querying, allowing users to run long-running queries without blocking resources, further optimizing performance for large datasets.

Splunk: Accessible but Proprietary

Splunk's primary query language, Search Processing Language (SPL), is a versatile tool for searching and analyzing machine-generated data. It combines search terms with commands using pipes, allowing users to filter, modify, and analyze data. SPL supports Boolean operators, wildcards, and field extractions, offering flexibility for data manipulation tasks. However, it is a proprietary language developed and maintained exclusively by Splunk.

As a key component of Splunk's ecosystem, SPL is specifically designed to work with Splunk's data indexing and search capabilities. It is optimized for searching and analyzing machine-generated data stored within the platform. This proprietary nature allows Splunk to tightly integrate SPL with its architecture.

The proprietary nature of SPL reflects Splunk's broader closed-source software model, allowing the company to maintain control over the language's development and integration. While this approach ensures a cohesive and optimized experience within the Splunk environment, it also means that SPL skills are not directly transferable to other platforms or systems outside of Splunk's ecosystem.

Conclusion

Elastic’s query languages, particularly the new ESQL, stand out for their power and efficiency, making them a superior choice for advanced data analysis. While Elastic’s DSL has a learning curve, its flexibility allows for intricate query constructions that precisely target data. ESQL enhances this by simplifying complex data transformations and leveraging concurrent processing for faster execution.

The ability to perform complex operations within a single query and asynchronous querying capabilities optimize performance for large datasets, making Elastic ideal for dynamic data environments. Unlike Splunk’s proprietary Search Processing Language, Elastic empowers users with a versatile query language adaptable to various data sources and structures. This flexibility, along with continuous improvement and integration with Kibana, positions Elastic as a leading choice for organizations aiming to unlock the full potential of their data.

POWER VS ACCESSIBILITY COMPARISON

ELASTIC

- DSL offers flexible search capabilities, enabling complex queries for precise data targeting.
- Requires significant training due to its complexity.
- ESQL is a new piped query language simplifying querying, aggregating, and visualizing large datasets.
- Supports complex transformations and calculations in a single query with familiar syntax.
- Concurrent processing enhances speed and efficiency over traditional queries.
- Piped syntax reduces complexity by minimizing multiple queries.
- Asynchronous querying allows long-running queries without resource blocking.

SPLUNK

- Search Processing Language (SPL) is designed for searching and analyzing machine-generated data, combining search terms with commands using pipes for flexible data manipulation.
- It is proprietary, developed and maintained exclusively by Splunk, which means skills are not easily transferable to other platforms.
- SPL is tightly integrated with Splunk’s architecture, ensuring optimal performance for data indexing and searching.
- The proprietary model allows for continuous development aligned with Splunk's product roadmap and customer needs.
- The closed-source nature of SPL limits adaptability to external systems.

Artificial Intelligence and Machine Learning

Elastic: Native and Comprehensive

Elastic has integrated a suite of AI and machine learning features into its Elasticsearch platform, including native capabilities for anomaly detection, forecasting, and data classification. The platform also supports vector search and integration with large language models, enabling advanced natural language processing tasks like sentiment analysis and text classification.

Elastic has also invested heavily in unsupervised learning and preconfigured models for identifying observability and security issues, along with the ability to apply supervised models to incoming data at ingest. Elastic's recent focus on generative AI is evident in its AI Playground and the Elastic Search AI Platform, which support refining semantic text queries and prototyping conversational search experiences.

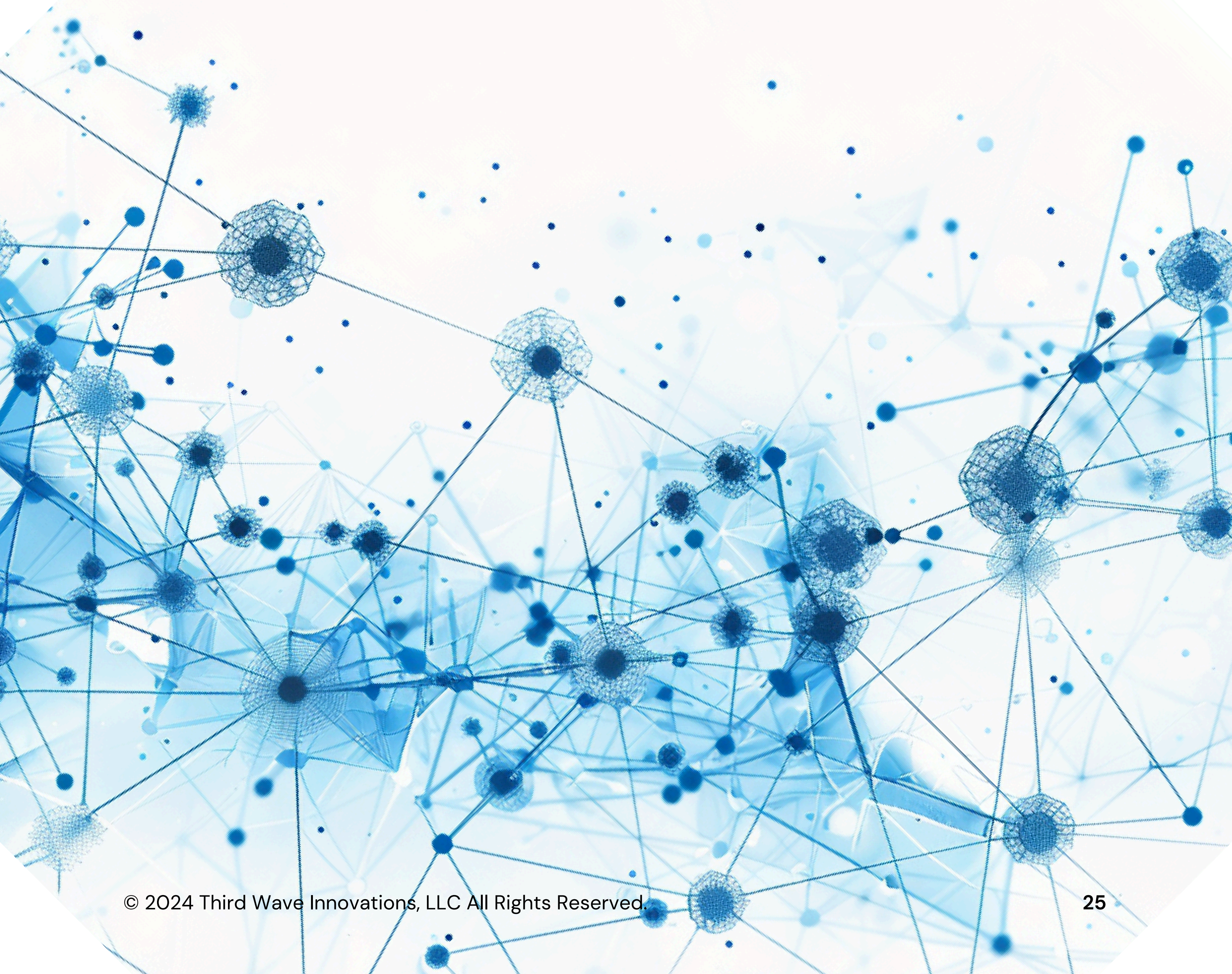
Splunk: Proprietary Toolkit

Splunk provides its AI and ML capabilities primarily through the Machine Learning Toolkit (MLTK) and other integrated features. The platform offers strong anomaly detection and predictive analytics capabilities, particularly for IT operations and security use cases. The MLTK supports common machine learning operations like prediction, outlier detection, forecasting, and clustering. For more advanced data science use cases, Splunk offers the Data Science and Deep Learning app (DSDL), which supports tools like Jupyter Notebook, PyTorch, and TensorFlow.

Splunk has also been developing AI-powered assistive features, such as the Splunk AI Assistant, which can generate SPL queries from natural language prompts. The company is working on Security and Observability guided assistants that leverage generative AI to streamline incident investigations and analysis. Splunk also supports external model integration, allowing customers to extend its models or bring their own.

Conclusion

Elastic stands out with its deeply integrated and comprehensive suite of native AI and ML features. Its focus on advanced natural language processing, vector search, and seamless integration with large language models makes it a more versatile choice for modern data-driven applications. Additionally, Elastic's investment in unsupervised learning, preconfigured models for security and observability, and its forward-looking approach to generative AI, including tools like the AI Playground and Elastic Search AI Platform, position it as the more innovative and future-ready solution.



ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING COMPARISON

ELASTIC

- AI/ML features are integrated directly into Elasticsearch.
- Includes capabilities for anomaly detection, forecasting, data classification, and semantic search.
- Supports vector search and integration with external large language models.
- Advanced NLP tasks: sentiment analysis, text classification, and named entity recognition.
- Focus on unsupervised learning, preconfigured models for observability and security.
- Recent emphasis on generative AI with tools like AI Playground and Elastic Search AI Platform.

SPLUNK

- AI/ML primarily through the Machine Learning Toolkit (MLTK) for IT and security use cases.
- Offers prediction, outlier detection, forecasting, clustering, and predictive analytics.
- Advanced data science use cases supported by Data Science and Deep Learning app (DSDL).
- Splunk AI Assistant for generating SPL queries from natural language.
- Supports external model integration for customization.

Ecosystem and Extensions

Elastic: Community-Driven Innovation

The Elastic ecosystem is built on an open-source foundation, with Elasticsearch, Kibana, and other core components freely available. This approach has fostered a large community of developers and contributors, leading to rapid innovation and extensive customization options. Elastic offers a wide range of out-of-the-box integrations through Elastic Agent, Beats, and connectors, simplifying data ingestion from various sources. Elastic is highly flexible and can be deployed on-premises, in the cloud, or in hybrid environments. Elasticsearch's design for horizontal scaling allows for easy expansion of clusters to handle growing data volumes. Kibana serves as the primary visualization tool, providing powerful capabilities and an extensible user interface. For cloud deployments, Elastic Cloud offers managed services across major providers like AWS, Google Cloud, and Azure.

Elastic's open-source framework has fostered a large and active community, resulting in a diverse array of tools, plugins, and custom solutions. This ecosystem diversity can be a significant advantage for organizations looking to extend the platform's capabilities or integrate it with existing systems. This approach also provides community provided integrations, pipelines for parsing, and the addition of Kafka pipelines for high-volume workflows, along with the opportunity to pre-parse and route the data to different storage, or even other tooling.

Splunk: Curated Marketplace

The Splunk ecosystem is built around proprietary software, which results in a more controlled but potentially less flexible environment. Splunk's core strength lies in its powerful data indexing and search capabilities, complemented by the proprietary Search Processing Language (SPL) for querying and analyzing data. The ecosystem is enhanced by Splunkbase, a community-driven marketplace for apps and add-ons that extend Splunk's functionality for specific use cases. Splunk Cloud offers managed services on major cloud platforms, with tools available for migrating on-premises deployments to the cloud.

Splunk, with its curated marketplace, provides a more structured approach to extensions and integrations. The platform offers a wide range of pre-built applications and connectors designed to accelerate specific implementations and use cases.

Conclusion

Elastic stands out as the more flexible and adaptable platform, especially for organizations that prioritize customization and community-driven innovation. Its open-source foundation provides a rich ecosystem of tools, plugins, and integrations, allowing users to tailor their solutions to specific needs. Elastic's scalability, versatility in deployment options, and strong community support offer significant advantages for those looking to manage complex, growing datasets across varied environments. For businesses seeking agility, rapid innovation, and extensive customization, Elastic offers a compelling and future-proof solution.

ECOSYSTEM AND EXTENSIONS COMPARISON

✓ ELASTIC

- Open-source, community-driven ecosystem.
- Core components (Elasticsearch, Kibana) are freely available.
- Wide range of integrations (Elastic Agent, Beats, connectors) for data ingestion.
- Flexible deployment: on-premises, cloud, or hybrid.
- Designed for horizontal scaling to handle data growth.
- Elastic Cloud offers managed services on AWS, Google Cloud, Azure.
- Ecosystem enriched by diverse community tools, plugins, and custom solutions.
- Supports integrations with pipelines like Kafka for high-volume workflows and data routing.

✓ SPLUNK

- Proprietary, curated ecosystem.
- Strength in data indexing and search, with SPL (Search Processing Language) for querying.
- Splunkbase marketplace offers apps and add-ons for specific use cases.
- Managed cloud services available on major platforms.
- Structured approach to extensions and integrations through pre-built applications and connectors for faster implementation.

Cost Structures:
Open-Source vs Proprietary Models

Elastic: Enterprise

Elastic Enterprise employs a tiered pricing structure with four main levels: Standard, Gold, Platinum, and Enterprise, starting from $95 to $175 per month, respectively. This model is primarily based on resource usage, particularly compute capacity, and allows for unlimited data ingest without volume-based restrictions. Users gain access to more advanced features, enhanced support, and increased customization options as they move up the tiers. The Enterprise tier offers sophisticated security, machine learning, and AI capabilities.

Elastic's licensing model is straightforward, with licenses tied to the chosen tier. This approach is more suitable for organizations seeking a straightforward, predictable pricing structure with unlimited data ingest.

Splunk: Multi-model Pricing

Splunk offers three main approaches: entity-based pricing, workload pricing, and ingest pricing. The entity-based pricing, designed for Splunk Observability Cloud, is determined by the number of devices or hosts monitored, with tiered pricing for predictable costs. Workload pricing is based on the compute capacity used for search and analytics, measured in Splunk Virtual Compute (SVC) units for cloud deployments or vCPUs for on-premises installations. This model allows for flexible scaling across various use cases. The ingest pricing model, which is the traditional volume-based approach, charges per GB/day of data ingested and is available for both cloud and on-premises deployments.

Splunk offers annual subscriptions for cloud solutions and term licenses for on-premises products. While Splunk's multi-pricing model offer more flexibility, allowing organizations to choose a model that best aligns with their specific use cases and growth patterns. However, it compounds through scaling or expanding service features.

Conclusion

Elastic's tiered pricing model stands out for its simplicity and predictability. With a straightforward cost structure tied to compute capacity and the benefit of unlimited data ingest, it offers a clear and scalable solution for organizations looking to avoid the complexities of traditional volume-based pricing. As users move up the tiers, they gain access to enhanced features without worrying about escalating costs as data volumes grow. For businesses seeking a balance of cost-effectiveness, flexibility, and advanced features like AI and machine learning, Elastic is the more attractive option compared to Splunk's multi-model approach, which can become more complex and costly as services expand.

PRICING COMPARISON

	ELASTIC	SPLUNK
Pricing Structure	Tiered, resource based	Multiple models available
Data Ingest Limits	Unlimited	Varies by model
Scalability	Increases with tiers	Flexible across models
On-premises Option	✓	✓
Cloud Option	✓	✓
Support	Included, improves with tiers	Standard included, Premium available

Support and Knowledge Base

Elastic: Comprehensive Tiers

Elastic offers tiered support for both cloud and self-managed deployments. Cloud support ranges from Limited (basic, 3-day response) to Premium (24/7/365). Self-managed deployments have Platinum and Enterprise levels with round-the-clock support. Elastic's global team across 22 countries and 30 time zones ensures continuous assistance, allowing customers to choose the support level that suits their needs and budget.

Elastic's knowledge base includes comprehensive official documentation for all their products and solutions, a searchable Support Hub with articles and technical documentation, free training courses on topics like the Elasticsearch Relevance Engine (ESRE), and community forums for peer-to-peer support and knowledge sharing. This multi-faceted approach caters to different learning styles and user preferences.

Splunk: Basic and Premium

Splunk offers a simpler support structure with two main programs: Standard support, which is included with all licenses, and Premium support, available as an upgrade. Splunk provides various support channels, including self-service options, community forums, official documentation, knowledge base articles, video answers to common queries, and direct support for error messages, unexpected behaviors, and outages. This approach balances basic included support and a premium option for those requiring more extensive assistance.

Splunk's knowledge base offers detailed official documentation for all Splunk products, a community-driven Q&A platform called Splunk Answers for fast responses, a Splunk Community platform for users to connect and share ideas, a video library with answers to common queries, and a Support Portal containing a comprehensive collection of knowledge from Splunk technical support engineers. This diverse range of resources ensures that users can find the information they need in their preferred format.

Conclusion

Elastic provides a more comprehensive and flexible approach. Its tiered support structure, which includes 24/7/365 availability for both cloud and self-managed deployments, along with a globally distributed support team, ensures continuous assistance for critical issues. Elastic's diverse knowledge base, including official documentation, free training courses, and community forums, caters to different learning styles and user preferences. This multi-layered support system, combined with extensive global reach and flexibility, makes Elastic the stronger choice for organizations seeking tailored and responsive support solutions.

SUPPORT COMPARISON

	ELASTIC	SPLUNK
Support Structure	Offers four tiered support levels for cloud deployments, ranging from basic (Limited) to Premium (24/7/365). Self-managed deployments have two tiers (Platinum and Enterprise), all providing round-the-clock support.	Features two main support programs: Standard (included with all licenses) and Premium (available as an upgrade).
Response Time	Provides varied response times based on the tier, with Premium offering the fastest.	Offers direct support, including self-service options and priority support for Premium users.
Global Availability	Global support team distributed across 22 countries and 30 time zones, ensuring continuous availability.	No specific mention of global distribution, but direct support is available for critical issues.
Knowledge Base	Comprehensive official documentation, searchable Support Hub, free training courses, and forums.	IDetailed documentation, a community-driven Q&A platform, a video library, and a Support Portal.
Community Engagement	Community forums for peer-to-peer support and knowledge sharing.	Community platform, Splunk Answers, and a Q&A forum.
Training Resources	Offers free training courses	Video library and other self-service materials.

Elastic vs Splunk: The Verdict

Declaring a Winner

Elastic emerges as the clear winner with its flexible, scalable architecture, advanced query capabilities, integrated AI/ML features, and open-source ecosystem. It offers a robust and cost-effective solution. In contrast, Splunk’s vertical scaling and complex pricing pose long-term challenges, making Elastic the better choice.

Decision Criteria

Elastic's distributed architecture and horizontal scaling offer exceptional adaptability for managing growing datasets. This scalability, combined with a tiered storage approach, allows organizations to balance performance and cost efficiency. In contrast, Splunk's vertical scaling may provide immediate performance boosts but presents limitations and potential cost challenges over time.

Elastic’s query languages, especially the new Elasticsearch Query Language (ESQL), provide advanced analytical features that surpass Splunk's proprietary Search Processing Language (SPL), excelling in environments with dynamic data demands.

Elastic’s integrated AI and machine learning features, focusing on natural language processing, vector search, and large language models, give it a competitive edge in modern data-driven applications.

Elastic’s open-source foundation fosters a robust ecosystem of tools, plugins, and integrations, enabling customization and community-led innovation—ideal for organizations with evolving data needs.

Financially, Elastic’s transparent, tiered pricing with unlimited data ingestion offers more predictability and potential cost savings compared to Splunk's complex, multi-model pricing, which can become expensive as services expand.

Migration

While the migration process can be complex, many businesses find the long-term benefits of switching to Elastic outweigh the initial challenges, particularly for large-scale data operations and cost-sensitive environments.

Reasons for Migration

Several companies have migrated from Splunk to Elastic for various reasons. Splunk customers are increasingly migrating mission-critical workloads for observability, security, and enterprise search to Elastic. Large enterprises have migrated their Security Information and Event Management (SIEM) systems from various offerings, including Splunk, to Elastic.

Companies that have switched report several advantages:

- ·Increased data capacity: One user mentioned jumping from 100GB/day on Splunk Cloud to several TB/day with Elastic while doubling retention time.
- ·Improved performance: Users report faster data importing and searching compared to traditional databases like MySQL.
- ·Better handling of large datasets: Elastic excels at managing multi-billion row datasets.

Migration Support

Elastic has partnered with Third Wave to offer a fully managed Elastic service offering, including migration for observability and security. This offering provides incentives and support for Splunk customers looking to switch. With artificial intelligence (AI) features like automatic import, we can help automate the development of custom data integrations, reducing migration effort and transition time.

About Third Wave Innovations

n today's complex risk landscape, understanding and managing risk is more critical than ever. Third Wave Innovations stands at the forefront of this challenge, providing a comprehensive, clear, and accurate view of your organization's risk profile. Headquartered in Frisco, TX, we've been pioneering cybersecurity and risk management solutions since our inception in 2017. Our approach combines cutting-edge technology with robust cybersecurity defenses, ensuring comprehensive protection for every aspect of your business. We don't just respond to threats; we anticipate them.

.

- **Always Ready:** We provide our customers with the assurance that their environment is being monitored, engaging them as necessary, and ready to act when they need us.
- **Network and Security**: Our Cyber Network Operations Center (cNOC) leverages experience managing networks and equipment, to monitoring performance and cybersecurity threats. We work together to make our customers more resilient.
- **Simplified Complexity:** We understand that legacy systems may exist in our customer environments. By integrating these systems, we have the ability to securely extend the life of these systems, while monitoring them for exploitation and threats.
- **Integrated Security:** We understand the complexities of our customer environments, and that content is as important as the threat of attack. Our capabilities span network, cybersecurity, compliance and risk; for our customers we translate it, and we action it.

About the Authors

Patrick M. Hayes
Chief Strategy Officer
Third Wave Innovations

Patrick is a seasoned security strategist, certified enterprise security architect and an accomplished executive. With 30 years of experience, Patrick has built a reputation as a thought leader and innovator. Patrick has been credited with building and commercializing several successful Security-as-a-Service platforms in SIEM, XDR and Risk Analytics, managing as many as 500 billion security events a day. Throughout his career, Patrick has been recognized for his contributions to the cybersecurity community. He has authored numerous articles on best practices and has been invited to speak at industry conferences around the world.

Shane Thorson
Chief Technology Officer
Third Wave Innovations

Shane is an experienced technology leader with a demonstrated history of accomplishments spanning communications, operations and cybersecurity in the government, healthcare, and financial services industries. Shane is the architect behind the C4 Intelligence Platform and leads the ongoing engineering and development.

www.3rdwave.io

THIRD WAVE Innovations