elastic | The Search AI Company

Leveraging AI-driven observability to deliver better applications

Table of contents

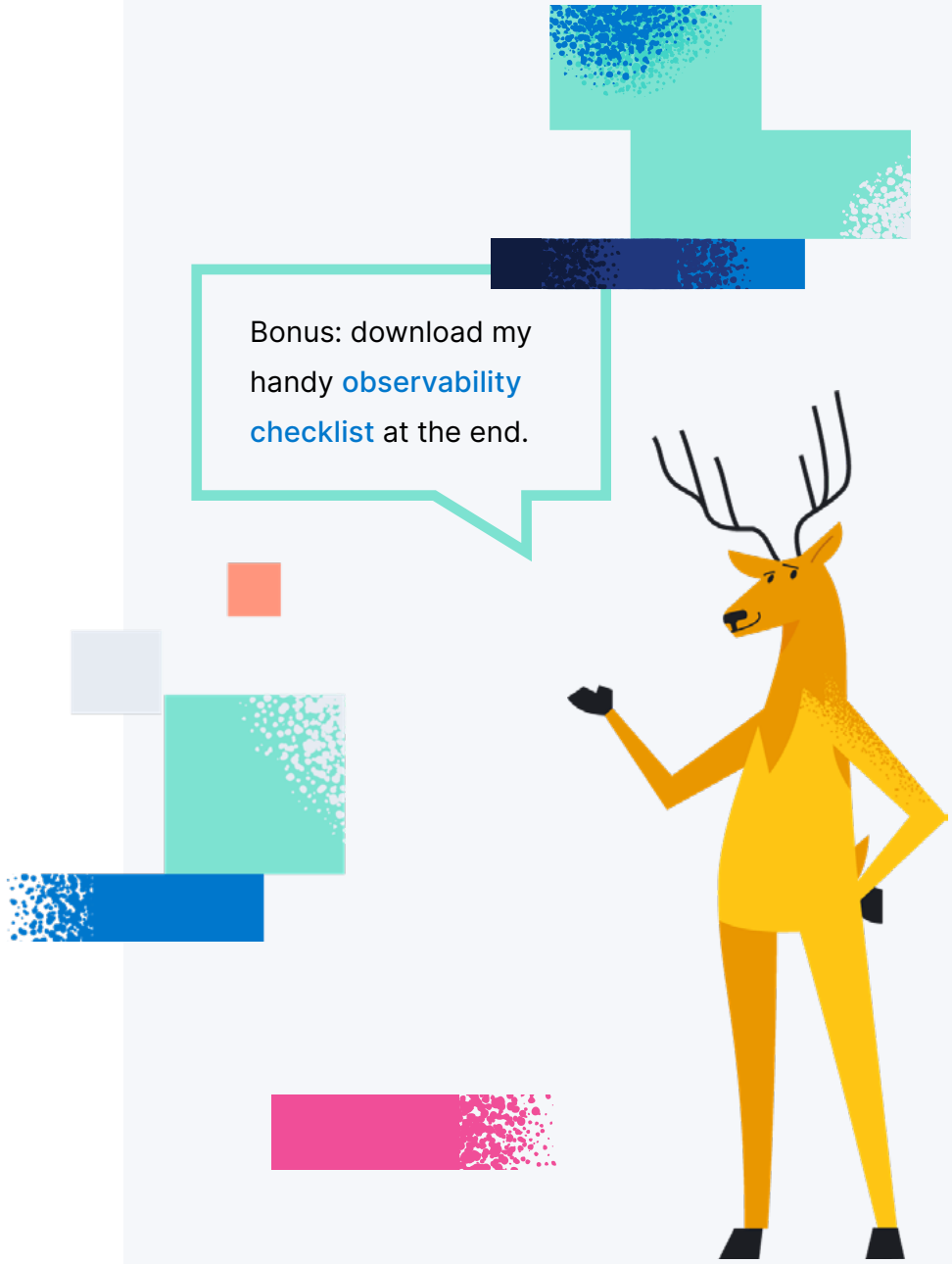
Start here if you're already familiar with observability

Introduction

Today, the average cost of IT downtime for medium to large businesses is $9,000 per minute, which can amount to approximately $540,000 per hour[1]. That's a big dent in your organization's bottom line. Without the ability to surface actionable insights from vast amounts of data quickly – in the growing complexity of modern systems, how do you keep up?

The right AI-driven observability strategy can help.

In this ebook, we'll show how you can take AI from trendy topic to a realistic roadmap, key considerations for a unified full stack observability platform, and how security and observability delivered through a single platform can create a powerful one-two punch for scale, speed, and protection.



1. Cost of data center outages, Ponemon Institute (2016)

What is observability?

First, a little history. The emergence of observability can be traced back to the early 2000s, pioneered in the form of application performance management (APM). During this era, data centers were the primary mode for delivering services to customers, and the new APM technology emerged to provide crucial visibility into application performance. Monitoring solutions were highly siloed: each team had its own tools for server monitoring, network performance monitoring, storage monitoring, and database monitoring. This fragmented approach led to a reactive problem resolution process, typically initiated by an on-call or pager system to address production issues as they arose.

In traditional monolithic systems, applications were built as a single, cohesive unit, making monitoring and problem resolution relatively straightforward, albeit siloed. As organizations transitioned to containers and microservices, the operational landscape became more distributed and complex. The emergence and adoption of Kubernetes has created dynamic and ephemeral services that are easy to run but difficult to diagnose. In addition, critical microservices often rely on serverless functions running on AWS Lambda, Google Cloud Functions, and Azure Functions.. Each microservice, functioning as an independent unit, introduced new challenges in monitoring and maintaining system performance and reliability.

As the digital landscape has evolved, so did the concept of monitoring. The journey from monolithic architectures to microservices has necessitated a more holistic and integrated approach — one that observability platforms are uniquely positioned to provide. These platforms have expanded their capabilities beyond APM to include digital experience monitoring (DEM), log and infrastructure monitoring, and support for cloud-native technologies. The core premise of APM

— that violations of error and latency thresholds degrade customer experience — remains relevant, with distributed tracing enabling teams to pinpoint issues within complex systems. However, challenges persist due to the cost, complexity, fragmented toolsets, and limited accessibility of comprehensive APM instrumentation, leaving many applications and third-party services unmonitored and problems elusive.

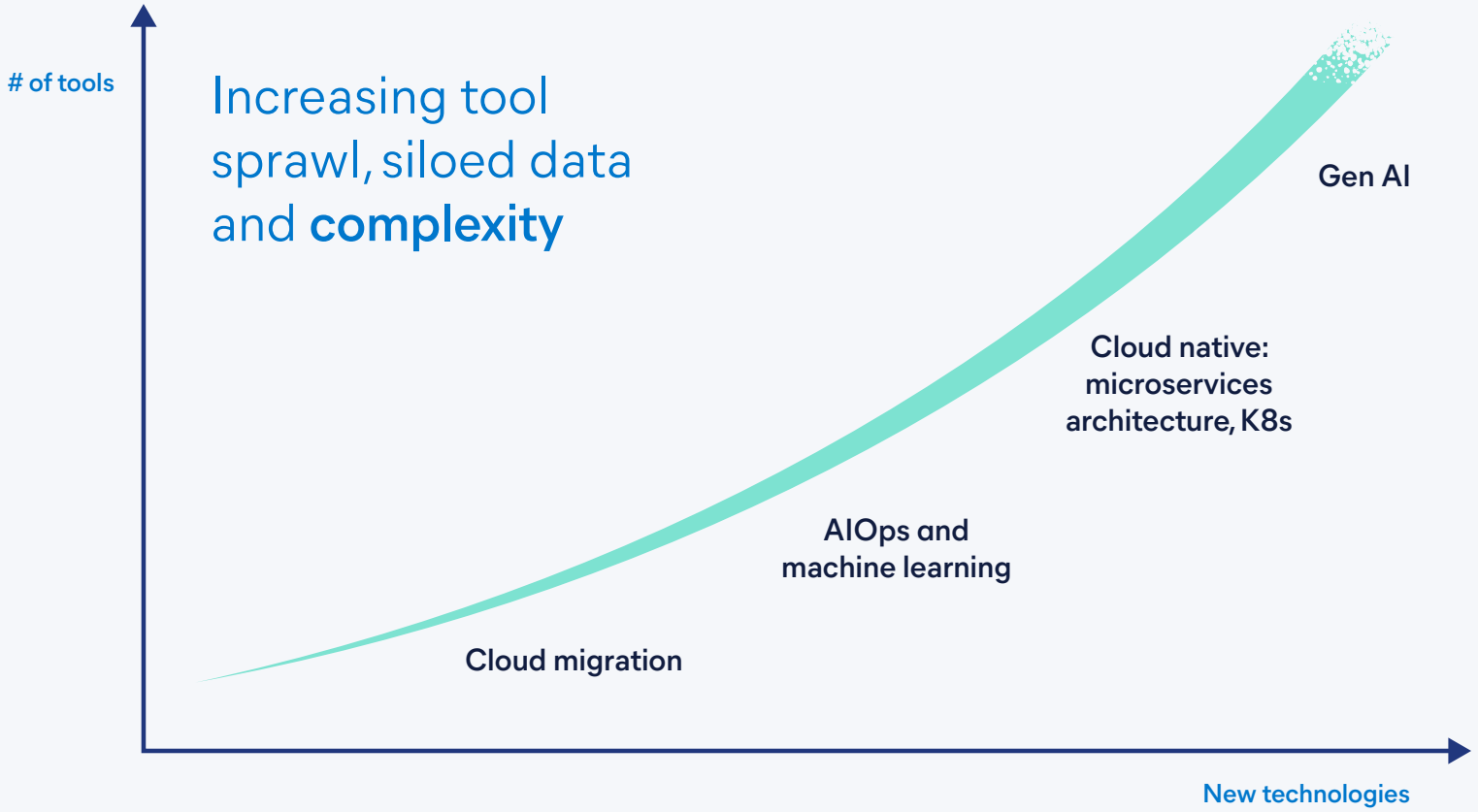
Observability has thus emerged with a broader mission: not only to provide a unified view of the operational landscape but also to proactively detect "unknown unknowns" across a wide variety of telemetry data types including metrics, logs, traces, and profiling data. Modern observability platforms ingest and correlate high-dimensionality and high-cardinality data across both operational and business data, and retain it in cost-effective and performant storage for future analysis and use. This wealth of data empowers teams to proactively identify root causes using advanced AI, analytics, and machine learning, addressing the complexity and scale of modern cloud environments.

Tech Tip

What is an unknown unknown?

Unknown unknowns are problems or conditions that are not familiar or anticipated and, therefore, do not have predefined monitoring rules or alerts. From an observability perspective, dealing with unknown unknowns requires a system that not only monitors known parameters and thresholds but also has the capability to detect and surface unexpected behaviors and patterns.

The obstacle:
digital transformation challenges



Why AI-driven observability matters

The evolution of software architecture from monolithic applications to distributed microservices, coupled with the rise of cloud technologies, has resulted in a demand for more sophisticated solutions.

- **Traditional monitoring tools fall short:** they operate in silos and struggle with the complexity of modern systems. The sheer scale of modern environments can overwhelm traditional tools with the amount of data generated, necessitating advanced data processing and scalable infrastructure.

- **Full-stack observability solutions** offer a comprehensive view, giving you full visibility into your application environment both real time and historically. Oftentimes, the volume of rich and detailed data you now have available, offers tremendous challenges and opportunities. How can you take advantage of this observability data beyond manual interactions and alerting?

- **AI-driven observability platforms** can overcome this hurdle by automatically surfacing insights and reducing labor-intensive manual data analysis. By leveraging machine learning and advanced analytics, AI can predict issues, identify patterns, and provide deeper insights into root causes, reducing toil for your SRE teams. Additionally, generative AI-based assistants can help interpret complex analyses, augmenting the expertise on the team with external domain expertise. The ultimate goal? Running AI assistants that leverage both external LLMs and internal knowledge bases to offer customized recommendations, automate routine monitoring tasks, and integrate with internal runbooks for remediation.

Evolution of observability

Bolstered by AI, observability not only unifies the disparate views of the past but also equips teams with the tools to navigate and manage the intricate web of modern cloud-native and microservices architectures.

By embracing AI-driven observability, organizations can gain deeper insights, enable data-driven decision-making, foster proactive problem resolution, and ultimately deliver better digital experiences in today's fast-paced, ever-evolving technological landscape.

How do you implement AI-driven observability? Let's dive into the building blocks for a future-proofed roadmap.

Considerations for an observability solution

By considering the following key aspects, enterprises can build a comprehensive and effective observability solution that enhances system reliability, performance, and business alignment.

The three pillars of observability: logs, metrics, traces (and profiling)

Effective observability hinges on the integration of three key types of telemetry data:

- **Logs:** logs are a primary data type generated by every type of system. They capture discrete events and provide contextual insights during incidents. The volume of logs can quickly become overwhelming (and costly to store) and logs often require enrichment and transformation to derive value from.
- **Metrics:** metrics track numerical data over time, such as CPU usage or request counts. They enable proactive detection of issues by monitoring key performance indicators (KPIs), as well as resource utilization trends, enabling capacity planning analysis.
- **Traces:** tracing forms the basis of service dependency maps that illustrate the flow of requests through distributed systems. Distributed traces help pinpoint performance bottlenecks and errors within a specific service, often down to the line of code.

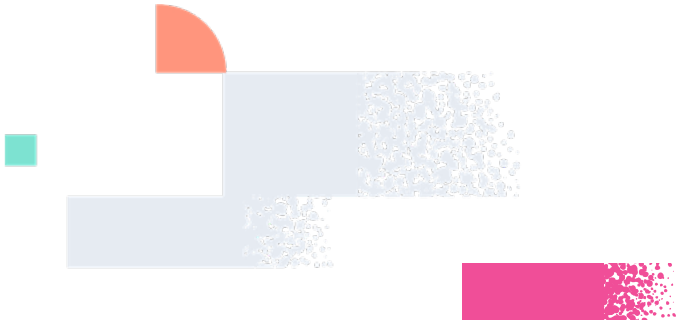
In addition, a fourth core data type has emerged in the form of profiling.

- **Profiling:** profiling provides granular insights into the performance characteristics of your applications, such as CPU and memory usage at the code level.

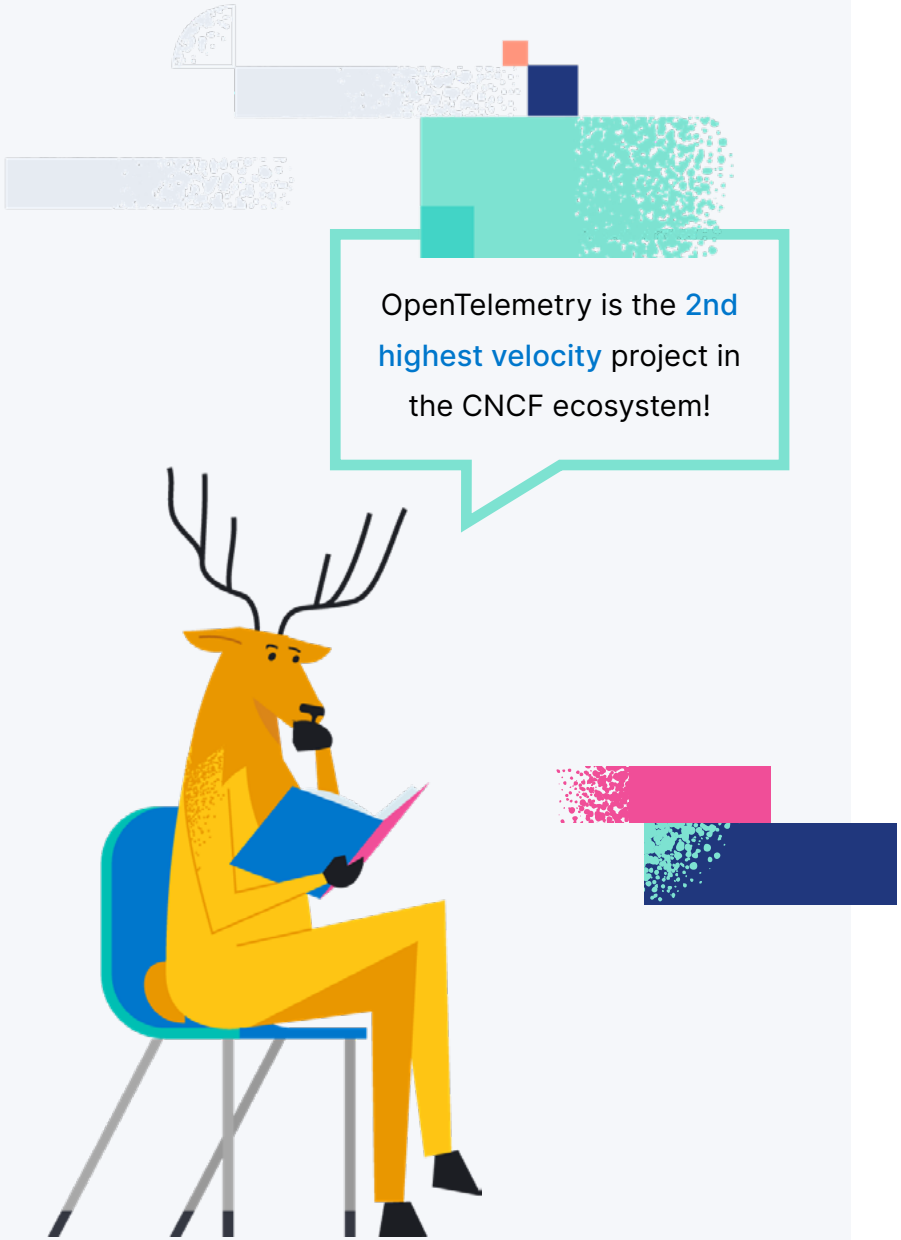
Importance of logs as a ubiquitous data source

Logs are a foundational data type generated by virtually every system, application, and device. They serve as a comprehensive record of discrete events and provide an internal view of what is happening within a system, as well as an audit trail for security monitoring by recording access attempts, configuration changes, and other potentially suspicious activities. This makes logs indispensable for both observability and security.

But logs do more than just record events: they can also provide contextual insights that are crucial during incidents. Each log entry typically includes a timestamp, an event description, and other relevant metadata such as severity level, source, and contextual information. This detailed record allows system administrators and engineers to track the sequence of events leading up to an issue, facilitating thorough root cause analysis. For instance, if an application crashes, log data can reveal the specific sequence of events that led to the failure, including error messages, configuration changes, or external factors. Analyzing error and event logs also reveals critical dependencies where failures in one service impact others.



Logs often contain enough information to extract insights typically obtained from metrics or traces. By correlating unique identifiers such as transaction IDs, session IDs, or user IDs, logs could be used to track the flow of a particular transaction or request across multiple services, and map out the sequence of service interactions and dependencies. Visualization tools and machine learning algorithms could further enhance this process by graphically representing service interactions and automatically detecting relationships. Despite their potential, logs are underutilized in many observability solutions. Unlike traces or metrics, which often have administrative or performance overhead, logs are continuously generated by virtually every system and application. By fully harnessing logs, observability teams can achieve a more complete understanding of their systems, leading to better detection, diagnosis, and remediation of issues.



Embracing open standards with OpenTelemetry

Open standards, such as the CNCF-supported OpenTelemetry project, will be crucial in future-proofing your observability strategy. OpenTelemetry provides a standardized framework for instrumenting, collecting, and exporting telemetry data from your applications.

OpenTelemetry also defines Semantic Conventions (SemConv): a common naming scheme for observability signals. SemConv helps teams generate a consistent telemetry format for all their observability data, ranging from events data, logs data, metrics, traces, and resources.

An open and common schema is essential for a robust observability platform. By standardizing the structure of telemetry data, a common schema simplifies data ingestion, transformation, and enrichment processes. It ensures consistency and accuracy across different data sources, enabling more efficient data analysis and correlation.

By embracing an open and extensible observability platform and by adopting OpenTelemetry, enterprises ensure interoperability across different tools and avoid vendor lock-in, facilitating a more flexible and scalable observability architecture for current and future use cases. This flexibility is essential for maintaining an agile and responsive observability strategy

Optimizing data ingest, costs, and storage

The scale and volume of telemetry data generated by modern applications necessitate efficient ingest and storage solutions. An enterprise observability platform must handle high-dimensional, high-cardinality, and high-volume data with low latency. A robust ingest pipeline, underpinned by a common schema, ensures data consistency and interoperability across different systems. Moreover, the transformation and enrichment of log data is crucial for preparing data for search, AI, and machine learning, extracting meaningful insights, and facilitating comprehensive analysis.

Performant and scalable storage solutions, combined with cost-effective data tiering and smart retention policies, ensure that observability data remains manageable and accessible. By optimizing the data ingest process and employing efficient storage strategies, enterprises can maintain the performance and reliability of their observability platforms while keeping operational costs in check.

Ownership of data and its lifecycle is another critical consideration. Organizations must ensure they have full control over their data from ingestion to deletion. This includes understanding where data is stored, how it is transformed, and how long it is retained. Effective data lifecycle management practices help in complying with regulatory requirements, protecting sensitive information, and optimizing storage costs. By maintaining ownership and control throughout the data lifecycle, you can better safeguard your data, ensure compliance, and derive maximum value from observability investments.

Advancing your observability capabilities with AI

As technology evolves and environments become more complex, distributed, and dynamic — particularly with the adoption of cloud (multi and hybrid) and cloud-native architectures (Kubernetes, serverless) — the need for a unified AI-driven observability platform becomes paramount.

By combining centralized data management with advanced AI capabilities, this approach improves the ability to correlate disparate data types, both business and operational, leading to enhanced system reliability and better business outcomes.

- **Centralized data management** through a unified data store and federated search supports the correlation of diverse signal types — such as logs, metrics, traces, and profiling data — preserving valuable context, while high dimensional indexed data stored in a vector database facilitates faster and more accurate root cause analysis. This not only enables seamless search and analytics, but also promotes collaboration among teams by providing access to the same comprehensive data set through a single pane of glass, leading to efficient triage and problem resolution.

- **AIOps (artificial intelligence for IT operations)** enhances observability by integrating machine learning and advanced analytics to automate anomaly detection, correlation, root cause analysis, pattern isolation, forecasting and trending. With unsupervised learning, AI can even help identify unknown unknowns in a system. Detecting an unknown unknown is different from other kinds of issues because it involves identifying problems that have no predefined patterns, metrics, or alerts, making them unpredictable and harder to detect. By taking a proactive approach with machine learning, teams can address issues before they impact end users, reducing downtime and improving system reliability in complex, dynamic environments.

- **Generative AI-based assistants** can greatly enhance observability by offering an interactive natural language based experience. These assistants can interpret complex data sets, offer actionable context-aware insights and recommendations, automate routine monitoring tasks, and even execute runbooks for remediation. They are especially valuable when integrated with an organization's internal knowledge base, ticketing systems, and runbooks through retrieval-augmented generation (RAG) technologies. By surfacing the most relevant and accurate insights from vast amounts of telemetry data, and taking automated corrective actions, they reduce the burden on IT teams and accelerate incident resolution.

By combining unified observability with advanced AI, teams can seamlessly perform search and analytics across the entire data estate. This enables the correlation of business and operational data, enhanced context and collaboration, expedited problem resolution, and effectively addresses the complexities of modern cloud environments.

BONUS

Glossary of AI terms for observability

Large language models (LLMs): A large language model (LLM) is a type of artificial intelligence model designed to understand and generate human-like text and understand complex language patterns.

Some popular LLMs used for observability include Open AI and Gemini. These models are often employed in chatbots and virtual assistants to help with tasks such as report creation, script generation, and even remediation.

Retrieval augmented generation (RAG): Retrieval augmented generation (RAG) is a technique that combines the strengths of retrieval-based models and generation-based models to produce more accurate and contextually relevant outputs.

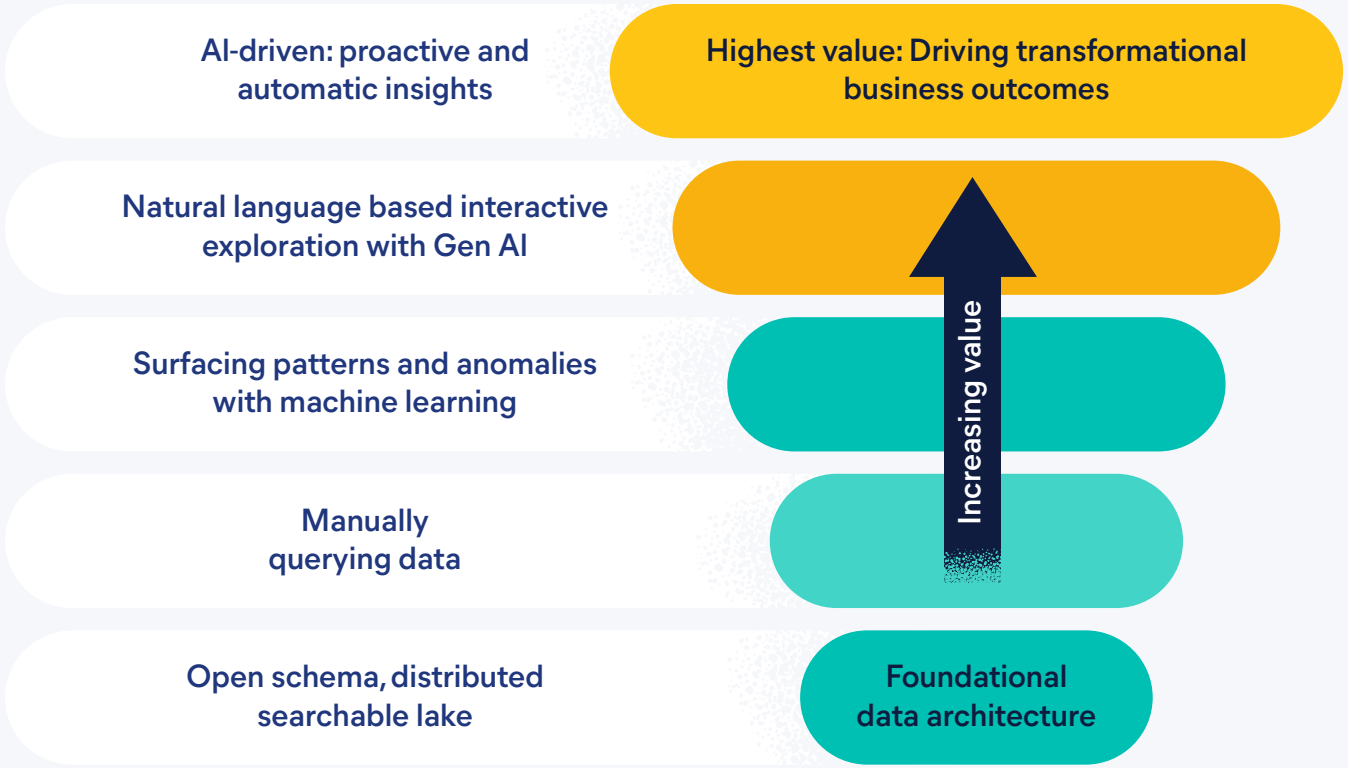
In RAG, an AI system first retrieves information from relevant organizational or other knowledge sources (retrieval phase) to generate a context-aware response (generation phase) from the LLM.

This combination can facilitate proactive issue detection, comprehensive incident analysis, personalized dashboards, and more.

Vector databases: A vector database is a database that stores, indexes, and queries vector embeddings. Vector databases can significantly accelerate AIOps by efficiently managing high-dimensional data and enabling fast retrieval and pattern recognition across the large datasets generated by observability platforms. They are crucial for real-time analytics and anomaly detection at scale.

Forecasting: In machine learning, forecasting involves using historical data to predict future values and trends. For instance, machine learning models can analyze patterns in resource usage, error rates, or response times to forecast when a system might become overloaded or when components are likely to fail. This predictive capability is crucial for identifying potential issues before they occur, allowing for proactive maintenance and optimization.

Predictive recognition: Predictive recognition involves identifying patterns, anomalies, or specific events in data, often in real-time, to predict outcomes or classifications. For example, predictive recognition might identify patterns in log data that predict a potential security breach, system malfunction, or degradation in performance, allowing for immediate intervention to prevent issues.



AI-driven analytics

Shift from manually chasing data to surfacing insights automatically with AI-driven analytics

Integrating SLOs with incident management workflows

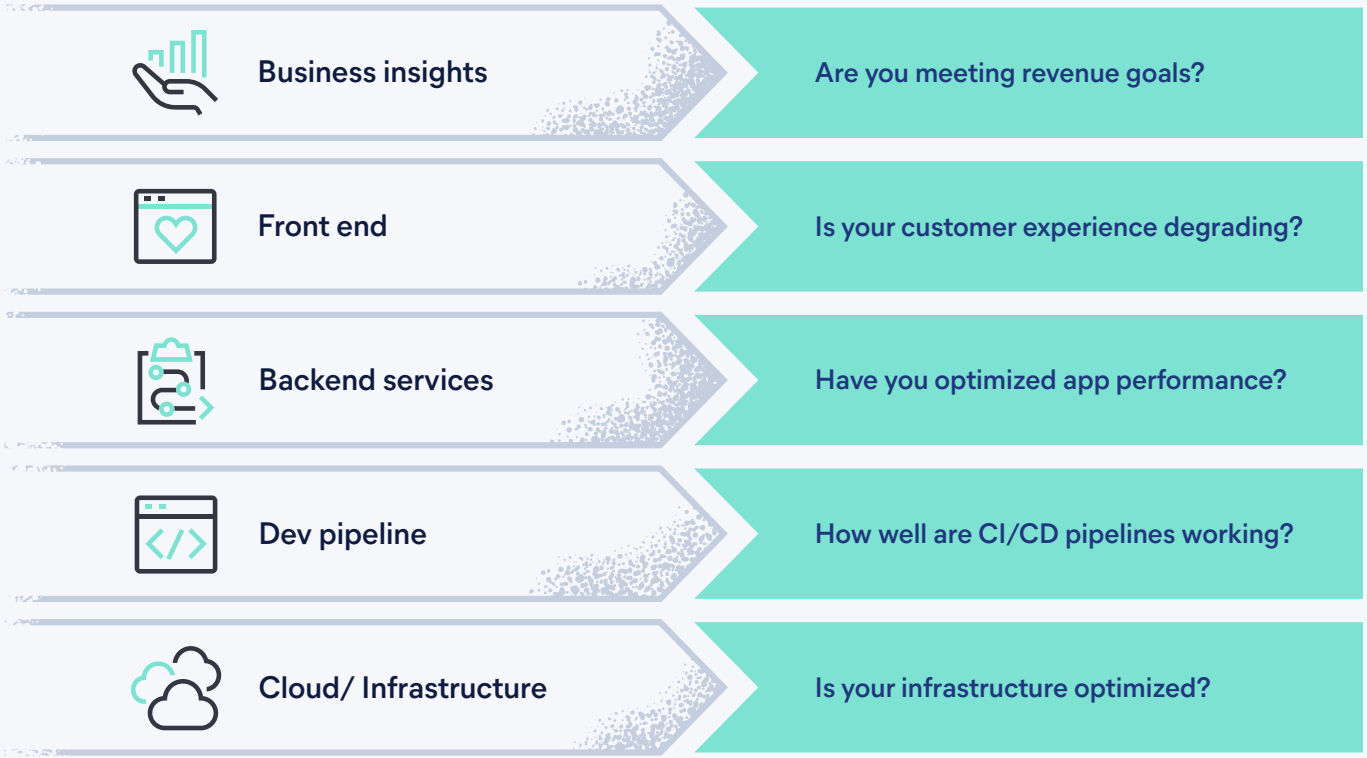
Service level objectives (SLOs) are essential for setting measurable targets that ensure your services meet desired performance and reliability standards. An enterprise observability solution should integrate seamlessly with incident management systems to enable automated workflows for SLO monitoring and alerting.

This integration ensures that incidents are managed efficiently and in alignment with business objectives. By transforming raw data into actionable insights, reducing alert noise, and prioritizing critical issues, teams can enhance overall operational efficiency. Integrated service desk workflows ensure that incidents are logged, tracked, and resolved systematically.

SREs can better manage business and operational Service Level Agreements (SLAs) by defining SLOs for service performance, error budgets, reliability, and business goals. Key features include:

- **Real-time monitoring:** Continuously track service performance against defined SLOs.
- **Error budget management:** Monitor and manage error budgets and burn rates to maintain service reliability.
- **Automated alerts:** Generate alerts based on SLO breaches and integrate them into incident management workflows.
- **Comprehensive reporting:** Provide detailed reports on service performance, reliability metrics, and business impacts to ensure alignment with business objectives.

By leveraging these capabilities, you can deliver superior customer experiences and ensure service reliability, aligning technical performance with business goals and improving overall service management.



Driving business and operational excellence with well defined SLOs and SLIs.

Are you meeting SLOs/SLIs?

Combining observability and security in a single platform

In the modern enterprise, the convergence of observability and security is crucial for maintaining robust and resilient IT environments. A single platform that integrates both observability and security capabilities offers several significant advantages:

- **Centralized data management:** By consolidating security information and event management (SIEM), endpoint security, and observability data into a single platform, organizations can achieve a comprehensive view of their IT landscape. This centralized approach enables more effective monitoring and analysis, facilitating quicker detection and resolution of issues.

- **Streamlined operations:** Integrating observability and security reduces the complexity of managing multiple disparate tools. This streamlined approach not only simplifies workflows but also minimizes the risk of data silos and blind spots, ensuring a holistic view of system health and security posture.

- **Improved collaboration and efficiency:** With a single source of truth, observability and security teams can collaborate more effectively, and streamline DevSecOps processes. Shared insights and unified dashboards enable SRE/SOC teams to work together seamlessly, improving efficiency and reducing the mean time to resolution (MTTR) for both security incidents and other operational issues.

By combining observability and security in one platform, organizations can ensure that their IT operations are not only efficient and reliable but also secure. This holistic approach supports the overarching goal of maintaining business continuity and protecting sensitive data in an increasingly complex and distributed IT environment.

Aligning IT and business goals with business observability

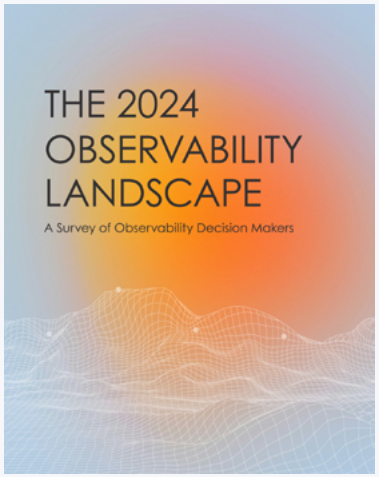
As data volumes and costs grow, there is a greater expectation for observability to become more explicitly business-oriented and help drive the data-driven enterprise. Decision-makers are focusing on cost, value, and outcomes, leading initiatives such as tools consolidation, establishing observability centers of excellence (CoE), implementing chargebacks/showbacks (FinOps), and adopting technologies that avoid vendor lock-in.

This is where business observability steps in. Business observability extends beyond technical metrics to include business-related data, such as transaction volumes, user engagement, and revenue impacts. By extracting valuable business information from logs and operational data, your company can gain contextual insights and create business dashboards that correlate technical performance with business outcomes. The end result: a comprehensive understanding of how system performance affects user experience and business success.

Tech Tip

The observability landscape

In a survey of 500 IT decision-makers, 71% of respondents have an observability center of excellence (COE) or have plans to create one[2].



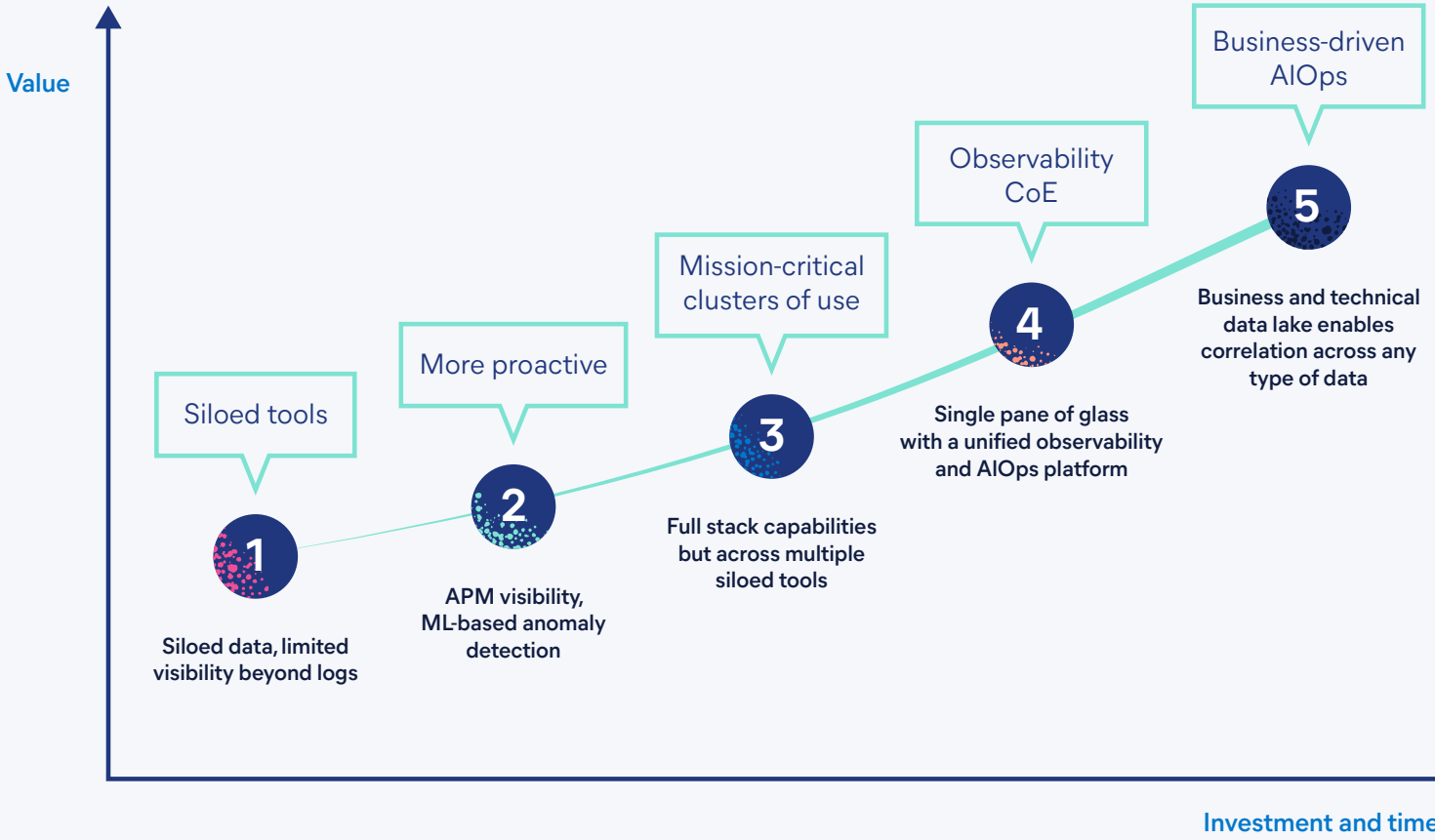
Get the report

2. The 2024 Observability landscape report, research conducted by Dimensional Research and sponsored by Elastic

How do you reach this level of maturity? Organizations are building an observability center of excellence (CoE) as a strategic investment to enhance their ability to manage and optimize complex IT systems. By bringing together cross-functional expertise from SREs, developers, operations, data analysts, and security experts, this collaborative approach ensures that observability practices are standardized and aligned with business objectives, fostering a culture of continuous improvement and rapid incident response.

By leveraging the latest tools and technologies, such as AI and machine learning, an observability CoE can transform raw data into predictive insights, enabling proactive management of system health and performance. Integrating both business and operational data into a data lake takes it one step further. By enabling comprehensive correlation and analysis across the entire data estate, this approach results in a data-driven AIOps-enabled observability strategy that further enhances business outcomes.

Accelerating your observability journey



Observability Maturity Curve

What can AI-driven observability do for me?

In the rapidly evolving landscape of modern IT, AI-driven observability with mature observability practices is crucial for maintaining robust, secure, and high-performing systems.

Unified observability offers you a single pane of glass approach, centralizing data from diverse sources into one cohesive view. This centralization not only enhances the ability to correlate and contextualize business and operational data but also fosters collaboration and speeds up problem resolution for SREs. In today's complex environments, where cloud-native architectures and multi-cloud strategies prevail, having a unified platform simplifies the management of high-dimensional, high-cardinality, and high-volume data.

AI takes unified observability to the next level:

- AI-driven observability leverages AIOps to automate anomaly detection, root cause analysis, and predictive trending so potential issues are identified and addressed before they impact end users – significantly reducing downtime and enhancing system reliability.
- Generative AI assistants provide an interactive, natural language-based experience, offering actionable insights and automating routine tasks, which eases the burden on IT teams and accelerates incident resolution. Ultimately, AI-driven observability not only supports operational efficiency and reliability but also aligns with broader business goals. It enables organizations to extract valuable business insights from operational data, driving informed decision-making and fostering a data-driven culture.
- Lastly, when you combine SIEM, endpoint security, and observability data into a single platform, you streamline threat detection and incident response, improve collaboration, and achieve a holistic view of system health and security.

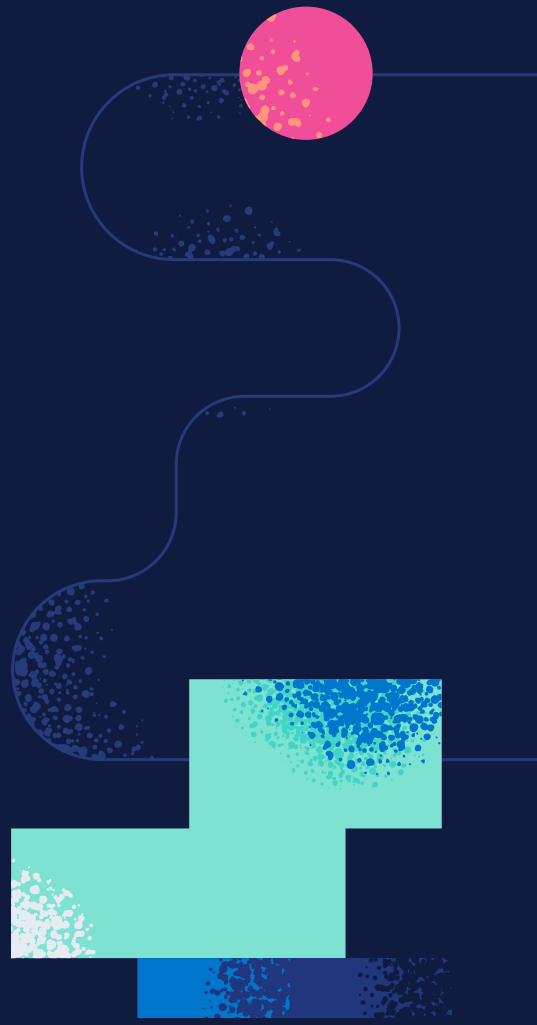
Meet Elastic Observability

Observability continues to evolve in response to ever-increasing data, complexity, and pace of change.

In the intricate, cloud-native landscape, just having any monitoring system isn't enough. Choosing the right observability platform is crucial in preparing for the future. With the right platform, organizations can stay ahead of the curve and leverage AI-driven observability to optimize their operations, gain valuable insights, and make data-driven decisions that drive growth and success.

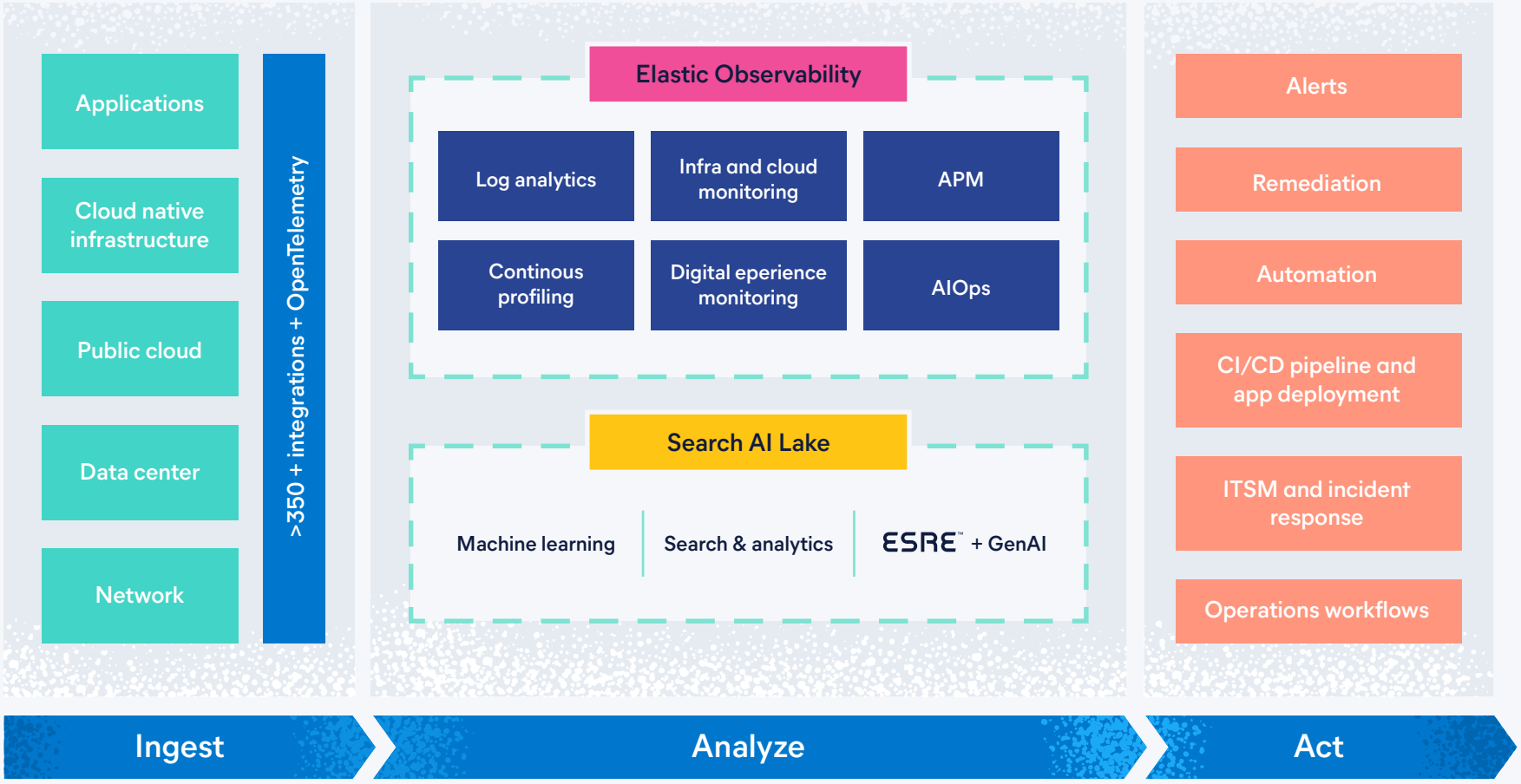
Is there a single platform that can help you use the power of AIOps and generative AI to transform the way you do business? Yes. Meet Elastic Observability.

Elastic Observability is a comprehensive, full-stack observability solution with a great foundation for AIOps and generative AI. Ingest all your data across metrics, logs, traces, and even business data — in a unified Search AI platform, built on the ELK Stack.



Elastic Observability

Powered by SearchAI



Bonus: observability solutions checklist

Use this checklist to ensure your observability platform is comprehensive, efficient, and secure.

Unified visibility

- [ ] Unified datastore for logs, metrics, traces, profiling, events, and custom data.
- [ ] Seamless search and analytics across all data clusters.
- [ ] Centralized data management.
- [ ] Workflows with context and correlation across business and operational data.
- [ ] Single platform for observability and security.

Data ingestion and storage

- [ ] Efficient handling of high-dimensional, high-cardinality, and high-volume data.
- [ ] Scalable storage solutions with flexible and performant data tiering.
- [ ] Cost-effective retention options.
- [ ] Support for log transformation and enrichment.
- [ ] Ownership of data and data lifecycle.

Telemetry

- [ ] Support for OpenTelemetry and other open standards.
- [ ] Common schema for all signal types.
- [ ] Real-time log collection and analysis.
- [ ] Monitoring of key performance indicators (KPIs).
- [ ] Distributed tracing for end-to-end service dependency mapping.
- [ ] Correlation of traces with logs and metrics.
- [ ] Detailed profiling for performance optimization.

AI and analytics

- [ ] Machine learning for anomaly detection and predictive analysis.
- [ ] Automated root cause analysis and pattern recognition.
- [ ] Proactive and actionable alerts based on AI.
- [ ] Natural language-based interactivity with Gen AI.
- [ ] AI assistant RAG integration with internal knowledgebases and runbooks.
- [ ] Automated insights and recommendations.

SLA and incident management

- [ ] Real-time tracking of SLAs, SLOs, and SLIs.
- [ ] Proactive alerting and error budget management.
- [ ] Comprehensive reporting and analysis.
- [ ] Integration with service desk and incident management systems.
- [ ] Reduced alert noise and prioritized issue resolution.

Business insights

- [ ] Monitoring of custom data such as transaction volumes, user engagement, and revenue impacts.
- [ ] Correlation of technical performance with business outcomes.
- [ ] Support for chargebacks/showbacks.

Integration and customization

- [ ] Open and extensible platform.
- [ ] Support for diverse multi-cloud and hybrid cloud environments (AWS, Azure, GCP).
- [ ] Integration with cloud-native technologies (Kubernetes, serverless).
- [ ] Seamless integration with existing tools and workflows.
- [ ] Customizable dashboards and alerts

Security

- [ ] Robust encryption, audit, and access controls.
- [ ] Compliance with industry standards and regulations.

Ready to begin your AI-driven observability journey?

First, let’s assess your organization’s observability maturity. Take our 15-minute assessment to understand how you score on observability readiness, adoption, and results. You’ll also receive a custom report and recommendations on key competencies like APM, AIOps, root cause analysis, and beyond. Let’s get started!

Get started

Thank you.

elastic | The Search AI Company