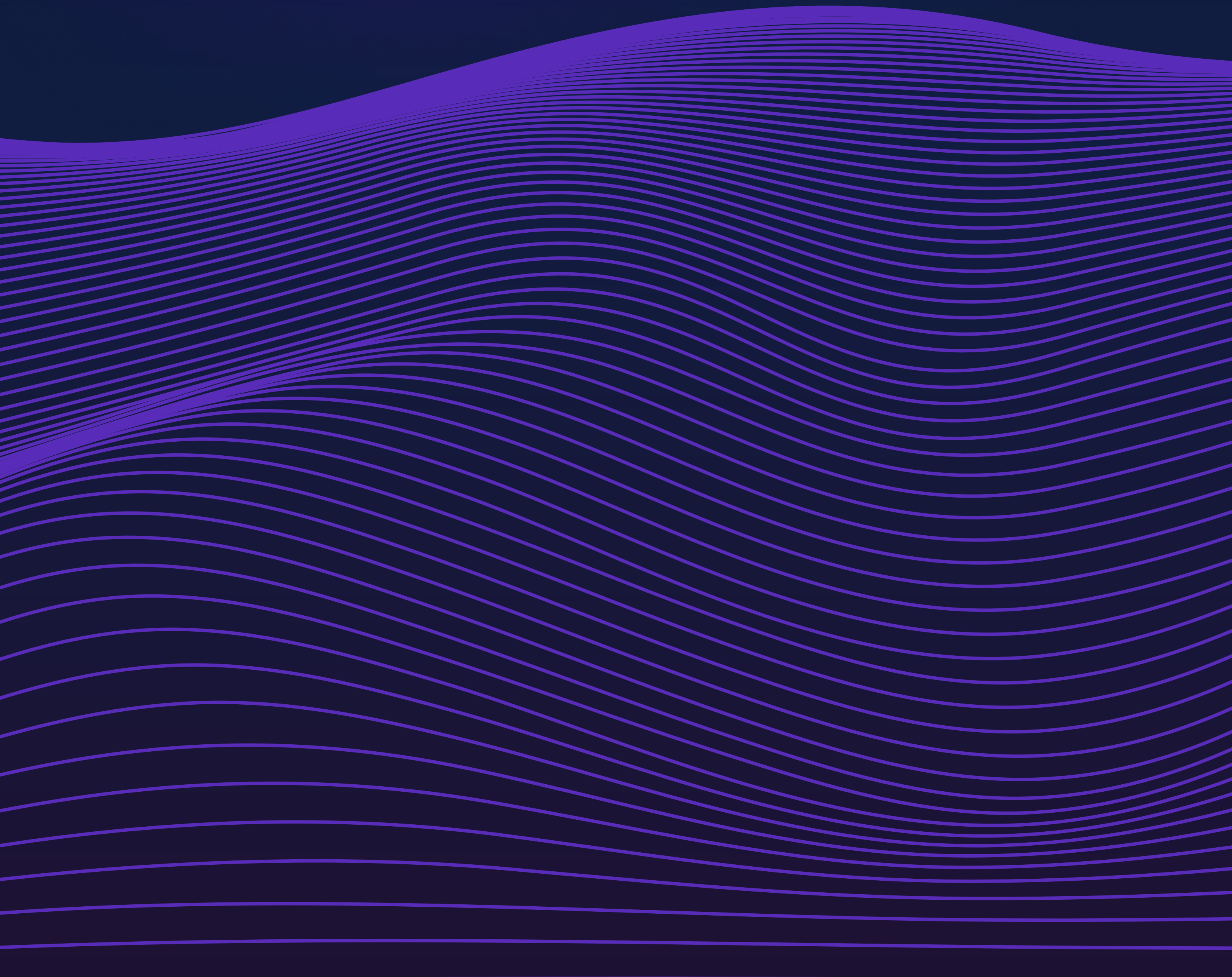
observo.ai

Whitepaper

Elevating Observability: Intelligent AI-Powered Pipelines

Create a dynamic, self-learning observability pipeline powered by artificial intelligence.



INTRODUCTION

Observability is the practice of asking questions about the inner workings of a system or application based on the data it produces. It involves collecting, monitoring, and analyzing various data sources to gain a comprehensive view of how the system behaves, its performance, and potential security threats.

Strong observability practices require you to collect enough data to answer these questions without knowing what those questions are today or will be tomorrow. The cost of collecting, storing, indexing, and analyzing all of this data can be exorbitant. Adding to the budget challenge, this data is growing at 35% a year for most organizations. That means that every three years the amount of data ingested into Observability, SIEM, or Log Analytics tools grows by 140%. Adding more strain to enterprise budgets, the corresponding infrastructure costs may be growing even faster with this explosion of data. Budget concerns force tough decisions about log management – what data types can be analyzed, how heavily sampling is used, and how long to retain data for ad hoc investigations. These decisions introduce risk, because the actionable data that is deemed too expensive to analyze may be the key to preventing security breaches, downtime, or an unsatisfactory customer experience which could mean loss of revenues, exorbitant fines, and eroded trust.

35%

Data Growth per year for most organizations

The nature of this data is also constantly changing. This reflects the evolution of enterprise infrastructures - new applications, changes in customer behavior, and interactions with third parties are represented in the data. It may also reflect the changing nature of threats to security, performance, and infrastructure stability over time. If observability efforts are optimized for what the environment is conveying today, tomorrow may be another story. Observability is about asking questions about the security, stability, and performance of applications and systems, but the questions should always evolve to recognize new threats and challenges. To answer these new questions, Security and DevOps teams must collect the right data and have tools that are flexible to adjust to changes in their infrastructure.

In this paper, we'll discuss the challenges of observability. We'll introduce Observo AI's approach to observability - specifically showing how adding Artificial Intelligence (AI) to an observability pipeline can dramatically optimize data reduction, reduce the time to identify and resolve incidents, and allow anyone to get insights from IT and security data without having to be a data scientist by using AI-powered natural language queries.

MODERN OBSERVABILITY CHALLENGES

Data Overload

2x

Data volume doubles as often as every 18 months

The exponential growth of observability data from distributed systems presents a substantial hurdle in managing, analyzing, and extracting insights from this deluge of information. According to recent data from our research, data volume doubles as often as every 18 months. As systems expand and evolve, the volume of logs, metrics, and traces explodes, making it challenging to manage and interpret these vast data streams effectively. This data growth strains storage capacities and complicates the real-time analysis, hindering the ability to promptly detect and respond to issues. The rapid pace at which data accumulates demands more sophisticated data processing and analytical capabilities to derive actionable intelligence amidst this torrent of information.

Legacy Architectures

Traditional architectures, relying on static, rule-based methods and indexing for data processing and querying, struggle to adapt to the soaring data volumes and dynamic nature of modern systems. As the scale of data expands, these static methods fail to provide the agility and speed needed for real-time analysis and troubleshooting. Furthermore, as log data changes with new releases and services, it introduces complexities that disrupt these static systems, requiring continuous updates and substantial efforts to maintain effective observability. The rigidity of static systems impedes the seamless extraction of insights from the data.

Rising Costs

30%+

Annual increase in observability costs

As data volumes escalate, the need for scaling up storage and computational resources becomes unavoidable. A study conducted by our team indicates that organizations experience a 30%+ increase in observability costs annually due to the growth in Observability data. Most organizations consider the escalating license costs of SIEM and log analytics systems for the growth of IT and security data, but the cost to store and process this data can be as much or more than licensing alone.

These soaring expenses force organizations to make tradeoffs on what data they can afford to analyze, how much of it and how long they can retain this data for compliance, and which tools they can afford to keep their environment secure, stable, and performing up to expectations. Manual log management practices consume resources and daily to yield the desired cost controls.

Compliance and Security Risks

The proliferation of sensitive data within observability systems raises complex compliance and security concerns. Adhering to data privacy regulations becomes increasingly challenging as the volume and diversity of data expand. Maintaining stringent security protocols in the face of growing threats and evolving regulatory landscapes is crucial. Our analysis of industry reports reveals that data breaches originating from Observability systems have increased by 48% in the last two years, posing a significant risk to organizations and underscoring the need for comprehensive security and compliance strategies to safeguard sensitive data.

Noisy Data Overwhelming Useful Signal

80%

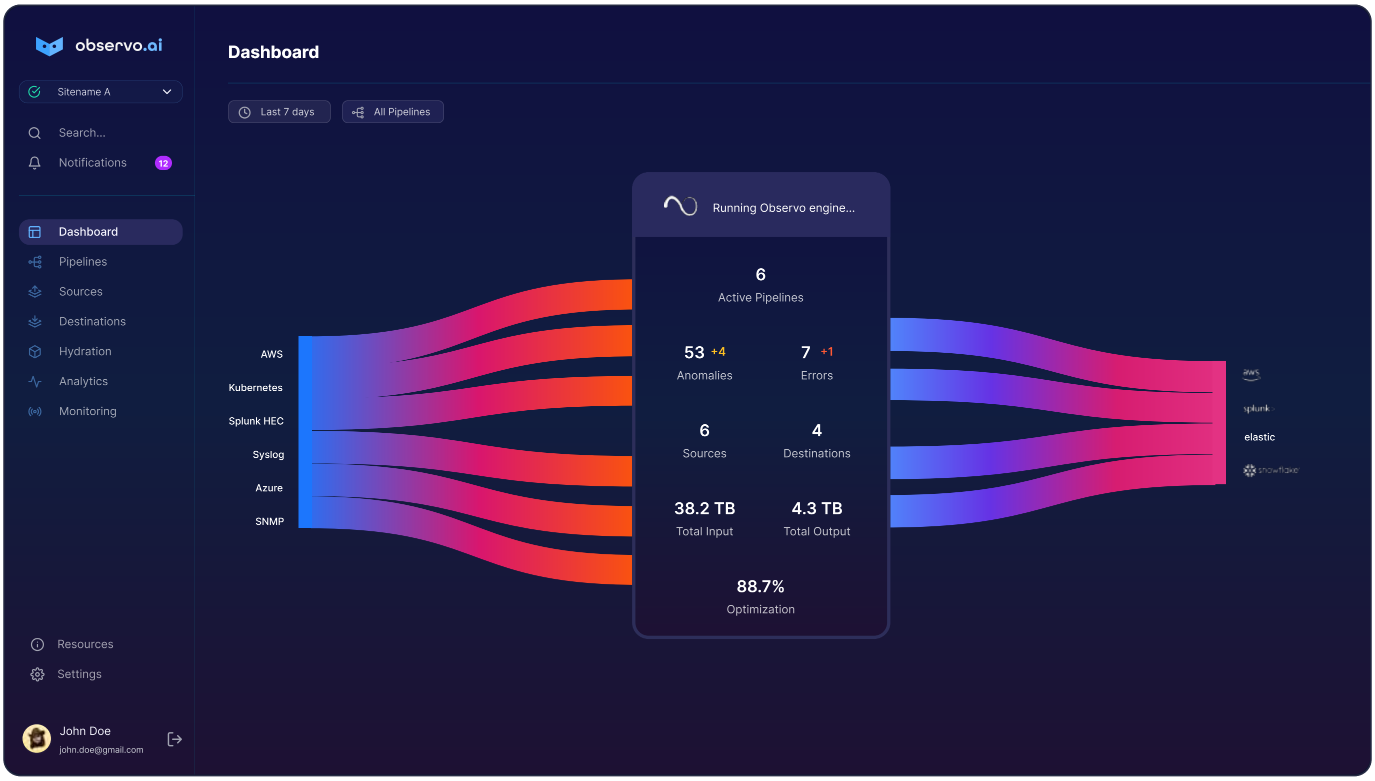
of log data has zero analytical value

In a study of enterprise log data, our team concluded that as much as 80% of log data has zero analytical value. Sending all of that unusable data to your analytics systems makes it that much harder to find the signal among all of the noise. Consider the needle in the haystack problem. Searching for just one needle is a lot harder as the haystack gets bigger and bigger. That's what happens when the indexes of analytics systems get bogged down with wasted data. You may be paying to analyze and store that data but it won't tell you anything about the health, security, and readiness of your environment. The signal/noise dilemma is also present among alerts generated by your analytics tools. Meaningful alerts are buried beneath a myriad of benign alerts, watering down their effectiveness and requiring manual review of each alert that your system generates. Wasting time on noise makes it much more likely that security breaches, system downtime, and real threats to your business will sneak through.

Lack of Dedicated Resources

To respond to all of the challenges we've listed above, companies have doubtlessly deployed teams to manually manage their data, attempt to maintain a pipeline that helps them address these challenges, and occasionally check in to see what changes in the data require further tweaks and new rules and filters to be applied. These homegrown pipelines are typically a motley collection of closely coupled scripts and tools that fail to solve the underlying problems. This approach requires dedicated resources with specific expertise in the underlying tools and data structures that they've been tasked to manage and optimize. When your pipeline experts leave the team, they take with them the secrets to keep everything up and running. This reliance on a few key employees strains your observability efforts and makes them less effective.

THE OBSERVO AI SOLUTION



Running Observo engine...

6 Active Pipelines

53 +4 Anomalies

7 +1 Errors

6 Sources

4 Destinations

38.2 TB Total Input

4.3 TB Total Output

88.7% Optimization

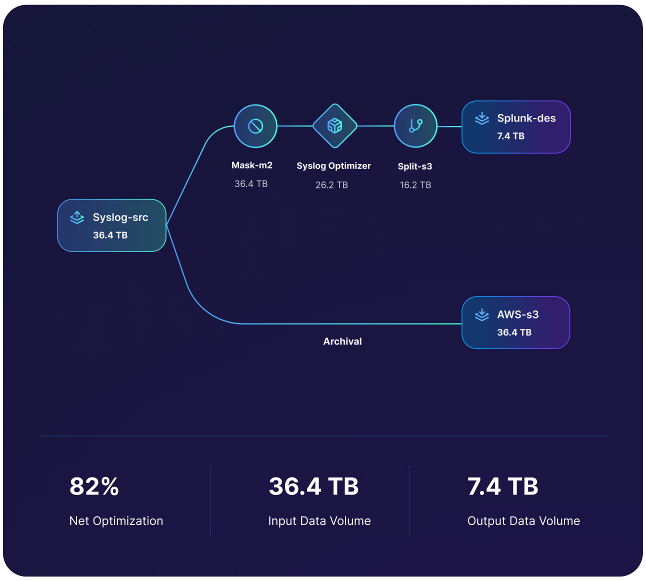
Observo AI was created to help DevOps and Security teams solve their biggest telemetry data challenges. In fact, our founders faced their own budget crisis when they were tasked with coming up with alternatives to curb a huge increase in an impending analytics tool contract. They leveraged their deep expertise in AI to optimize their data stream and the idea for Observo AI was born. By adding AI to observability, customers can achieve dramatically reduced time to solve incidents while optimizing the data in their stream to control costs and surface better, meaningful, and actionable insights. AI helps automate data optimization. We've created unique algorithms specific to each data type to eliminate as much noise as possible. This is a revolutionary approach to observability pipelines and replaces traditional pipelines with their static, rule-based methods that don't change to keep pace with evolutions in your IT and Security data.

50%+

Observo.ai can help reduce costs by more than half

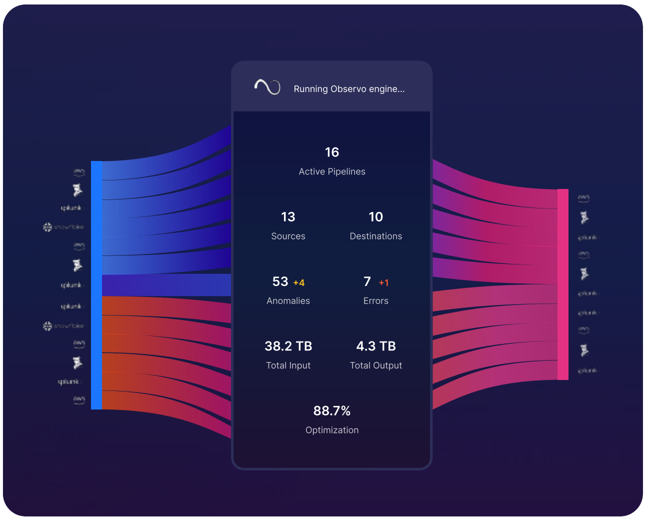
Observo AI can help reduce Observability costs by 50% or more based on the data types you are analyzing and the analytics tools you are using. By enriching data with AI-generated sentiment analysis, our customers have cut the time to identify and resolve incidents by more than 40%. Observo also automates security compliance by detecting and masking private and sensitive data that other tools may miss.

Data Optimization and Reduction



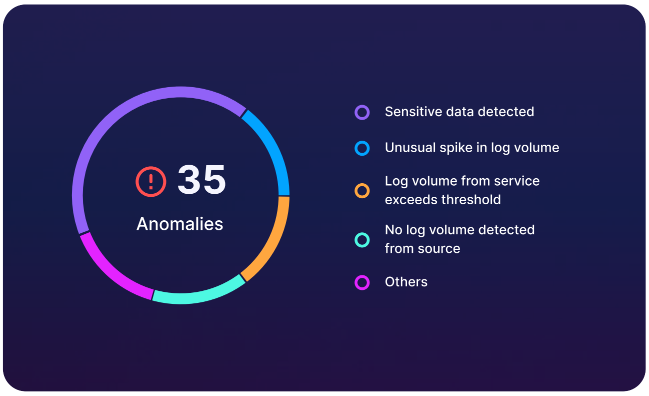
Observo uses AI and machine learning techniques to right-size data without having to set static rules that require the user to be an expert on their data and what is useful or not. Observo AI optimizes data by creating intelligent data groupings that can reduce noise by 80%. The smart summarizer has optimization transforms specifically built for each data type you want to analyze including VPC Flow logs, Firewall logs, OTEL, OS, CDN, and Application logs. These deep learning techniques are constantly learning and looking for improvements for further optimization. As your data changes, so do the optimizations.

Smart Routing



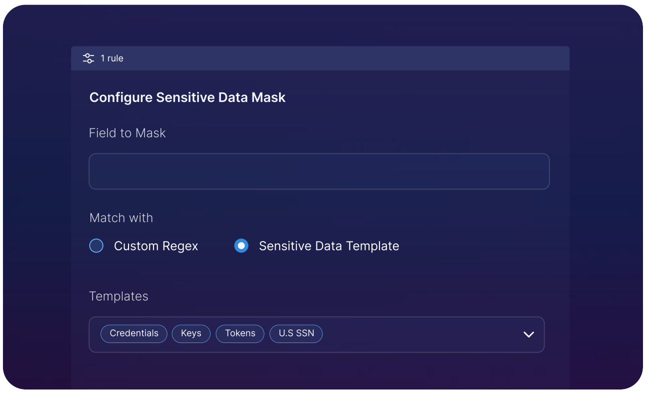
Observo AI can also transform data from any source to any destination - allowing you to choose the right tool or many tools to optimize what types of data need to be analyzed by the most expensive tools and which can be routed to a more cost-effective tool. Gone are the days of collecting data in different formats for every tool. With Observo AI, you can collect data once and route it to the right tool or storage destination in whatever format is required. Our AI-based models automate this so you don't need an expert to establish a long list of rules and you can be optimized and running in hours.

Anomaly Detection



The Observo AI pipeline "learns" what is normal for any given service and routes it accordingly to archives if it is normal or to an analytics tool for further investigation if it is considered an anomaly. The Observo AI Sentiment Engine enriches log data by tagging it before it's indexed and correlated by the analytics tool. This can be integrated with common alert/ticketing systems like ServiceNow, PagerDuty, and Jira for real-time alerting. By adding positive or negative sentiments to your data, you can much more efficiently route data to where it has the most impact. Take action on data that is outside the norm right away and either archive or deprioritize the rest.

Sensitive Data Discovery



1 rule

Configure Sensitive Data Mask

Field to Mask

Match with

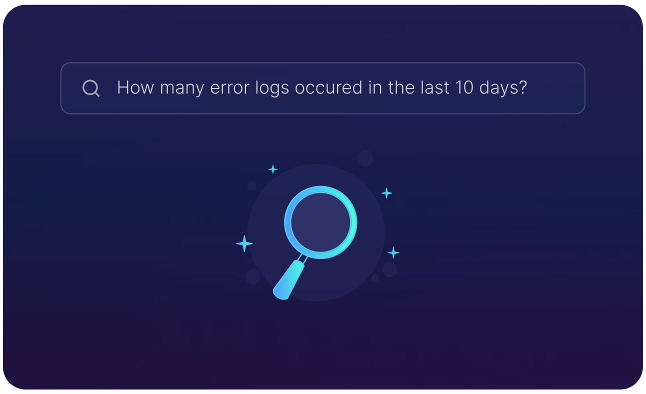
○ Custom Regex ● Sensitive Data Template

Templates

Credentials Keys Tokens U.S SSN

The Observo AI Pipeline detects sensitive data allowing you to secure it through obfuscation or hashing. Unlike static tools that set rules for what is sensitive data, Observo's AI uses pattern recognition to discover all sensitive data, even if it's in an unexpected field or metric. Observo AI automates compliance with privacy regulations like GDPR, CCPA, and PCI. Observo AI helps you keep all sensitive data safe and protected.

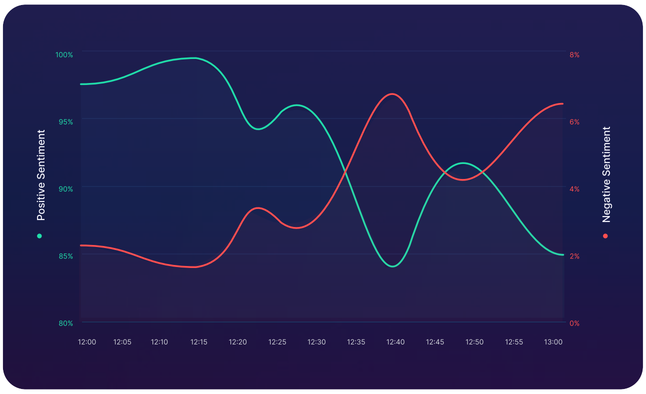
Searchable, Full-Fidelity Data Lakes in Low-cost Cloud Storage



How many error logs occured in the last 10 days?

Observo AI recommends taking a copy of full-fidelity observability data and routing it to inexpensive cloud object storage like AWS S3, Azure Blob, or Google Storage. Observo AI transforms log data into Parquet format, a highly compressible data format that allows search using natural language queries through tools like Athena. The cost of storing data in Parquet in cloud object storage can be as little as 1-2% of the cost of storing the same data in block storage attached to your analytics systems. This allows customers to keep more data for longer periods, which bolsters their ability to investigate incidents like breaches, which often occur months, if not years before being discovered. It also helps them comply with log retention standards and regulations which for some industries require storing logs for up to seven years. Observo AI can "rehydrate" this data at any time and route back to your analytics system of choice should you need to investigate this data on demand.

Data Enrichment



Observo AI enriches data to add context. Observo's AI models detect anomalies and assign "sentiment" based on pattern recognition. Sentiment dashboards add valuable insights and help reduce alert fatigue by helping Security and DevOps teams discern meaningful alerts from run-of-the-mill items that don't require immediate attention. Observo AI can also enrich logs with 3rd party data like Geo-IP and threat intel to make data more actionable. Adding the right data can significantly speed up queries in downstream tools and reduce the compute toll on indexing engines.

OBSERVO'S APPROACH TO USING AI

The telemetry data analyzed by Security, IT and DevOps teams is very heterogeneous and always changing. Static optimizations become hard to maintain and lose effectiveness. Leveraging machine learning techniques to address this data produces much deeper optimizations and continuously improves as the data itself changes.

Pattern Extraction & Anomaly Detection

The Observo AI data pipeline has a built-in smart module that extracts pattern clusters by processing streaming data at wire-speed. Pattern mining in the context of log data typically refers to the process of identifying recurring patterns or structures within log files generated by software systems, networks, or applications. These patterns are invaluable for understanding system behavior, diagnosing issues, and improving overall system reliability. We have created memory efficient pattern mining algorithms that can process Petabyte scale log data very efficiently. Observo AI keeps full fidelity in the data lake for compliance and archival purposes - this data can be rehydrated on demand and routed to any analytics tool with the click of a button.

This robust pattern extraction has the following advantages:

Dimensionality Reduction: Patterns often represent recurring sequences of events or messages in log data. By identifying and extracting these patterns, one can condense multiple instances of similar events into a single representation.

Grouping Similar Log Entries: Patterns help in grouping similar log entries together. Instead of dealing with each log entry independently, you can work with a set of patterns that capture the commonalities among multiple entries. This grouping simplifies the representation of the data.

Facilitating Anomaly Detection: Pattern clusters help create a baseline of normal behavior across log data - this baseline serves as a foundation for robust anomaly detection. Deviations from the "normal range" of existing patterns are dynamically detected and flagged for downstream analytics tools.

Abstraction of Information & SImplified Analysis: Extracted patterns provide a level of abstraction, allowing you to focus on high-level trends and behaviors rather than individual log entries. This abstraction is valuable for understanding system behavior and identifying issues without being overwhelmed by the sheer volume of raw log data. Extracted patterns simplify the analysis process - analysts can focus on understanding the behavior captured by the patterns.

Natural Language Query Engine

Observo AI democratizes access to data in Observability Data Lake by allowing users to interact with and query the data using natural language instead of traditional query languages. Observo AI exposes a user-friendly interface that is accessible to individuals and does not require any knowledge of the underlying schema. Our Natural Language Processing (NLP) engine understands the user's intent and exposes queries used to extract relevant information. Other tools rely on the operator having deep knowledge of the native query language in order to pull insights from the data.

Sentiment Enrichment

The Observo AI pipeline supports context aware Sentiment Enrichment transforms. Observo AI supports smart reduction of events as they flow through the pipeline. This reduction can be used to group messages belonging to a transaction or logical operation together. After all the log messages for a logical operation are aggregated using the smart reduction transform, it can be enriched by the sentiment analysis transform. This sentiment enrichment on aggregate events that represent the entire transaction context are very accurate as the entire transaction context is presented as a single event to the sentiment analyzer transform. This sentiment enrichment can be instrumental in improving the speed of downstream analysis as users can now focus on enriched events that have negative sentiment for troubleshooting.

Stream-Type Based Optimization Modules

Observo AI has native support for smart optimization modules that implement individualized algorithms for source data type enterprises commonly use, greatly improving the reduction of data within each stream type. We've created specific optimization modules for VPC Flow logs, Firewall, OS, CDN, and Application logs, to name a few. These data-type specific smart optimization modules are key building blocks that simplify adoption of optimization best practices by users.

Observo AI's use of smart AI algorithms ensures that your pipelines are always optimized. They are constantly learning what is normal, what should be flagged as an anomaly, and what is flat-out useless and should be trimmed altogether.

Our models keep looking for the new normal and more importantly, the new threats entering your environment. Static approaches can't do that without a lot of manual intervention and that's only if a human can see how the data is evolving. AI-powered observability is more comprehensive, flexible, and vigilant than traditional approaches.

CONCLUSION

The explosion of telemetry data for Security and DevOps teams has made observability efforts more complicated and much more expensive. By the time you negotiate your next license agreement with your tools vendor, those costs can more than double. The tools of today attempt to control this flood of data, but they ultimately fail to keep pace with data that's not only growing but changing from one day to the next. Static, rules-based methods require your team to choose what data to reduce, how to transform it, and where to send it. This demands a deeply skilled team. This approach also only works for the world that exists today. Your data is changing because your applications and infrastructure are constantly evolving. So are the types of threats from external parties. Observo uses AI rooted in deep data science to solve these problems. Our pipeline learns from your data and helps you optimize your stream by reducing 80% or more of your observability data. As your environment changes in real time, so do the optimizations generated by Observo AI. We apply sentiment analysis to get the most urgent data in the right place to take action now, before larger problems spin out of control. This helps our customers reduce the time to identify and resolve incidents by more than 40%. That means fewer breaches and outages leading to lost revenue and customer trust.

If you think we can help elevate your observability efforts with our industry-leading AI-powered observability pipeline, contact us at **info@observo.ai**