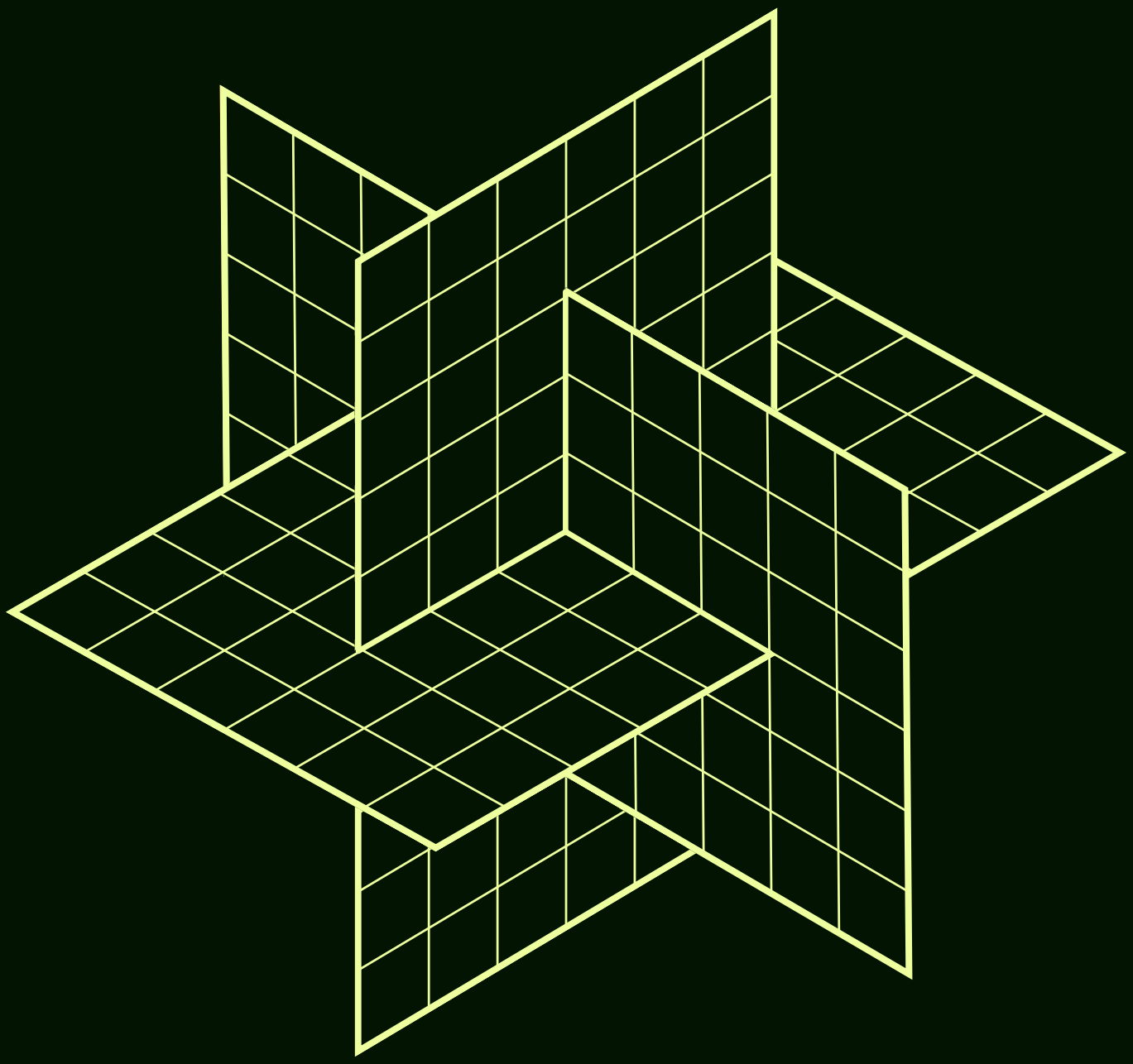
THE TELEMETRY BLUEPRINT: TURNING VAST DATA INTO BUSINESS INSIGHTS



mezmo

WHITE PAPER

SUMMARY

50% Over 50% reduction in data volume without compromising observability quality.

63% Filter and reduce 63% of data volume from standard web logs like Apache and Nginx.

61% Segregate 61% of Kubernetes logs. Keep only the essential, archive the rest.

50% 50% reduction in Kafka logs by extracting common message data.

94% Reduce 94% of Firewall Log Volume.

90% Convert logs and metrics and achieve up to 90% reduction and streamline data into actionable insights.

INTRODUCTION

In today's application and infrastructure management landscape, the influx of telemetry data is both a boon and a challenge. As the volume multiplies, extracting insights from this sea of information gets harder. The typical approach — directly routing data – often results in ballooning costs. Unchecked telemetry volume can drive compounding tolls and overages for observability platforms, cloud egress costs, and storage.

Telemetry pipelines are emerging as an effective technology to solve this worsening data problem. They streamline, process, transform, and route data, enabling organizations to navigate mountains of telemetry data. The intent is simple: turn expansive data sets into concise business vision without losing important information.

At Mezmo, we've embraced telemetry data and maximized it with an intelligent pipeline designed for the modern observability architecture. Mezmo's Telemetry Pipeline goes beyond standard practices, enabling our unique 5-step approach that has proven successful at maximizing the value of observability data:

1. Noise Filtering: Sifting through data to spotlight the essentials
2. Long-Term Data Retention: Safeguarding valuable data for future use
3. Event Trimming: Tailoring data for optimal analytics
4. Data Condensation: Translating voluminous logs into focused metrics
5. Operational Efficiency Boosting: Amplifying operating speed and reliability

INTRODUCTION CONT.

The implementation of these steps in real-world tests has shown remarkable outcomes, such as:

- A measured reduction of data volume by over 70% without compromising data integrity
- Enhanced analytics capabilities and business value by reducing data noise
- Achieving up to a 90% reduction in event data volume by converting logs to metrics, delivering business insights to a broader audience.

This white paper explores the 5-step process, highlighting our findings across many data sources and rigorous testing of our telemetry pipeline. As we detail our methodology, we'll also suggest different use cases. The goal is to showcase our journey and equip you with the tools and understanding to optimize your telemetry data.

Continue reading to learn how to transform your telemetry data into a blueprint for maximizing your observability investment.

RESEARCH AND METHODOLOGY

Our engineering and product teams used the following techniques to substantiate data reduction opportunities:

- We collected data from internal Mezmo sources, to emulate real-world scenarios. We also sourced data from Kaggle.com and other open-source locations, such as GitHub.

- We created telemetry pipelines with a standard account tied to the individual source types. We injected data into each pipeline for each sample through an HTTP source. To measure the results, we compared the before and after byte counts for a number of samples.

THE FIVE STEPS IN PRACTICE

STEP 1: NOISE FILTERING - TURNING VOLUME DOWN, AMPING VALUE UP

Organizations face the challenge of sifting through excessive noise within their data.

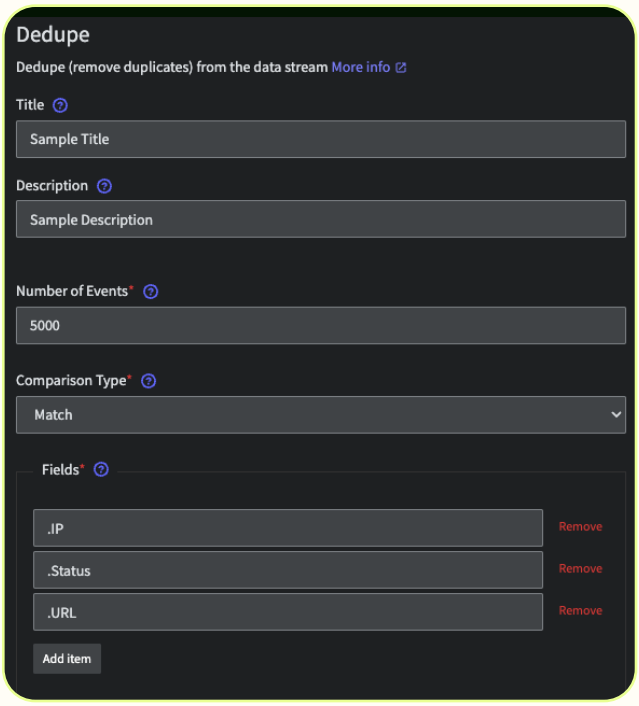
Take Apache Weblogs, for example- saturated with repetitive get requests and often redundant logs. Pertinent system anomalies like 4XX and 5XX status codes remain buried under routine 2XX codes. This issue isn't exclusive to weblogs but pervasive across many data types.

Organizations need a solution that helps them spotlight crucial information amid vast and noisy data streams.

The remedy to this is a pipeline adept at filtering noise, including events like positive confirmations, recurring notifications, and redundant "status=200" messages that inflate weblogs.

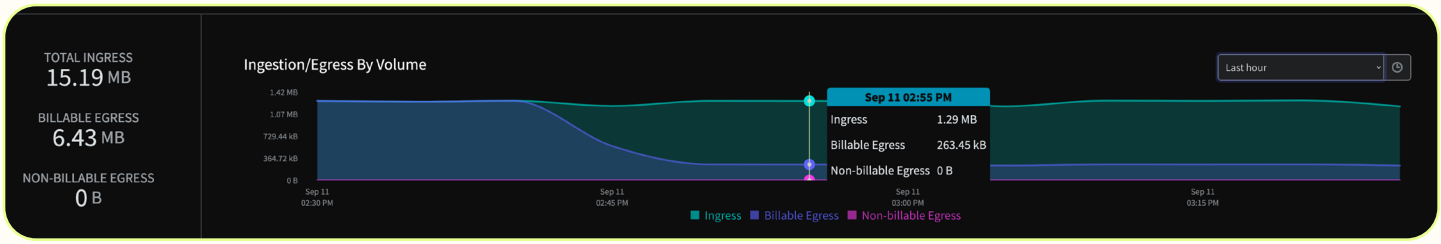
Ultimately, the telemetry pipeline would scan data, zero in on specific details, and omit superfluous information (based on user criteria) while registering their frequency for future use.

Within the Mezmo Telemetry Pipeline, the solution to this issue is our Dedupe Processor.



Dedupe Processor configuration form

Our Dedupe Processor dissects data within specific windows, keeping unique client requests and eliminating excessive repetitions. When employed, this process has shown its value, enabling a 63% volume reduction. Additionally, when combined, our Route and Dedupe processors can efficiently exclude the timestamp or focus on chosen fields, offering precision in deduplication.



Monitoring screen showing ingress and egress traffic to destination

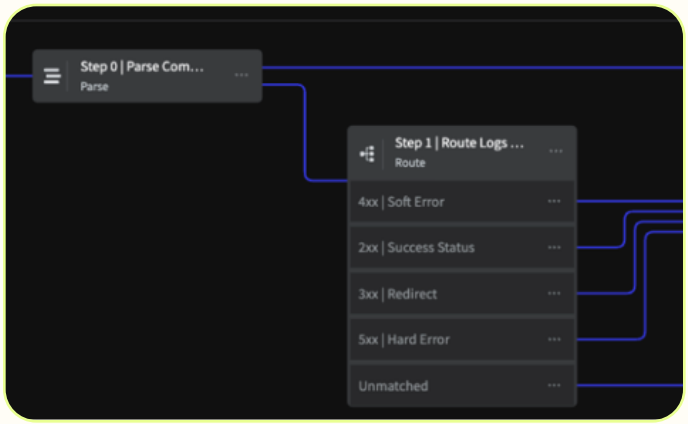
Precise, concise data paves the way for sharper insights and provides SREs with unambiguous perspectives, leading to fast troubleshooting and data-driven decisions. You can turn expansive datasets into brief, actionable information through intelligent data management and filtering.

STEP 2: LONG-TERM DATA RETENTION – PRESERVING INSIGHTS FOR TOMORROW

After deduplicating data and eliminating redundancy, the next challenge is determining how to retain valuable data cost-effectively. While keeping logs streamlined for immediate analysis is essential, specific datasets are indispensable for future needs, such as compliance audits, or security forensics.

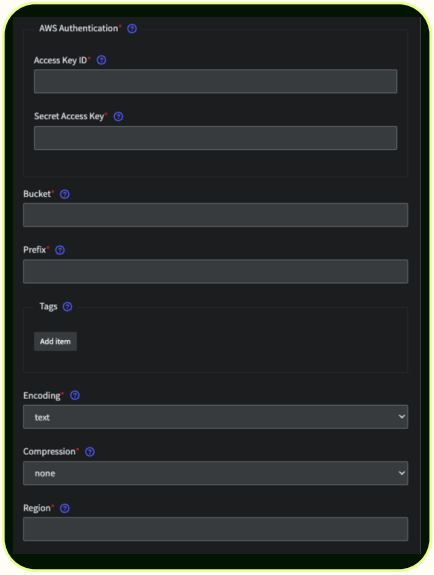
Organizations need a solution that stores a full-fidelity copy for easy future retrieval while allowing real-time data refinement.

Mezmo's Telemetry Pipeline offers a long-term data retention capability using the Route Processor. Not only does the pipeline streamline current data flows, but the Route Processor ensures that you can retain data long-term. Imagine a busy crossroad in a city, with data being the traffic that converges from various sources. Mezmo's Route Processor would be the traffic light at the intersection.



The Route Processor moving data to S3 and Datadog

It adopts a dual strategy. One lane would direct data straight to long-term storage (like AWS S3 vaults), ensuring it remains available for future use. The other lane would route the needed data to the observability platform destination.



A destination configuration form within the Route Processor for AWS S3

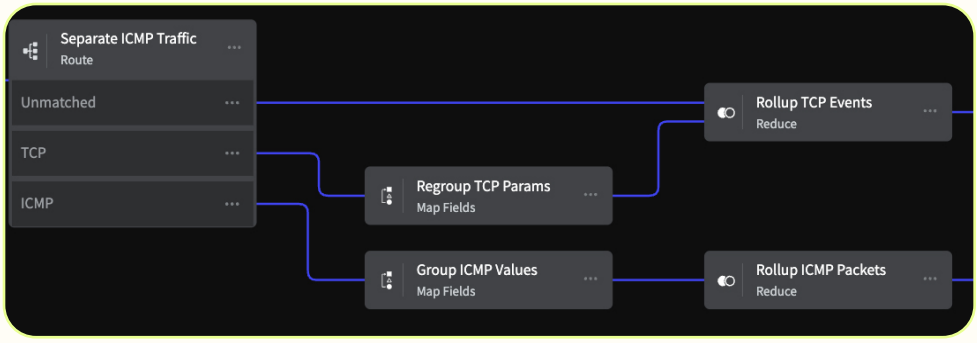
Additionally, you can add filtering capabilities that selectively channel specific logs to designated storage or processing paths, ensuring optimal resource utilization and tailored data accessibility. Routing data to long-term retention while gaining the benefits of data reduction by using the telemetry pipeline ensures cost reduction with no loss of fidelity.

Tip: If you're interested in deep-diving into routing specifics, check out our "Drop, Encrypt, and Route Data to Storage" guide in our documentation. Routing data to long-term retention while gaining the benefits of data reduction by using the telemetry pipeline ensures cost reduction with no loss of fidelity.

STEP 3: EVENT TRIMMING – CRAFTING DATA FOR DEEPER ANALYSIS

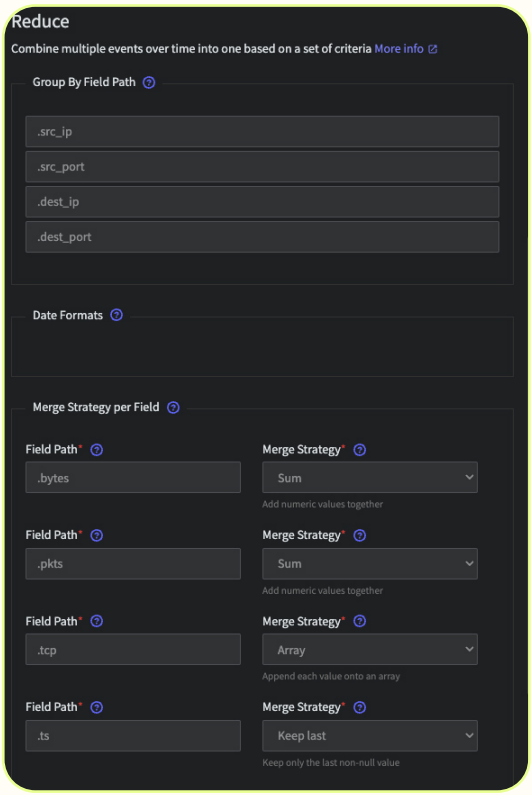
Another source of telemetry data excesses are redundant or null fields in log data. For example, security devices, notably the AWS Firewall, constantly record potential threats, resulting in intricate logs. These logs are critical for security (and thus retained in most cases long-term) but are also saturated with redundant or null fields.

Organizations need a way to transform these granular logs into more digestible and actionable datasets. More specifically, they need a way to strike a balance between retaining the vital components of logs while managing and reducing the overall data volume for efficiency.



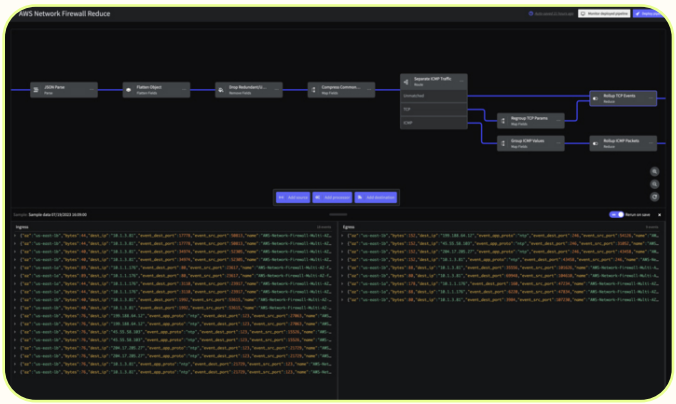
Routing within Mezmo Telemetry Pipeline based on protocol type

Mezmo includes the concept of Traffic Flow Segregation as part of its advanced data management. By segregating traffic– for instance, categorizing it as TCP or ICMP– Mezmo enables you to group logs by shared attributes. The resultant data, though streamlined, retains its crucial elements. As an extension, our Reduce Processor focuses on strategies like consolidating recurring data patterns in TCP sessions while summarizing ICMP packets, ensuring each data subset serves its specific audience–SREs or network engineers. By distinguishing traffic based on its nature and consolidating logs around common attributes using our Merge technique, we achieved a remarkable 94% data reduction within our pipeline. This substantial reduction ensures that observability platforms receive well-structured messages while preserving crucial information and enhancing performance.



The Reduce Processor for TCP and ICMP Packets

When compressing large quantities of log data into individual lines (typically for debugging purposes), our Parse Processor proves invaluable. The Parse processor enables you to extract essential elements from these massive data logs and trim the excess, offering a more concise yet informative view of the data.



A before (left) and after (right) comparison of data gathered through our integrated simulation feature

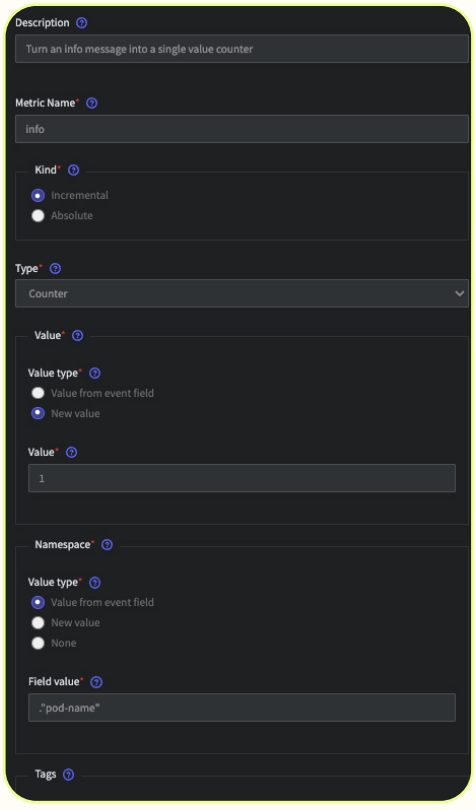
To illustrate the transformation and ensure user clarity, we also incorporate an integrated **simulation function** within our pipeline. This feature provides a before-and-after snapshot that users can use to understand their data flows, transformations, and make informed decisions about their data management.

Optimal data management goes beyond mere storage. The ability to craft and transform raw logs into actionable, insightful datasets is crucial. Such a transformation facilitates deeper analysis, ensuring organizations can glean genuine value from their extensive data pools.

STEP 4: DATA CONDENSATION – REDEFINING VAST DATA STREAMS

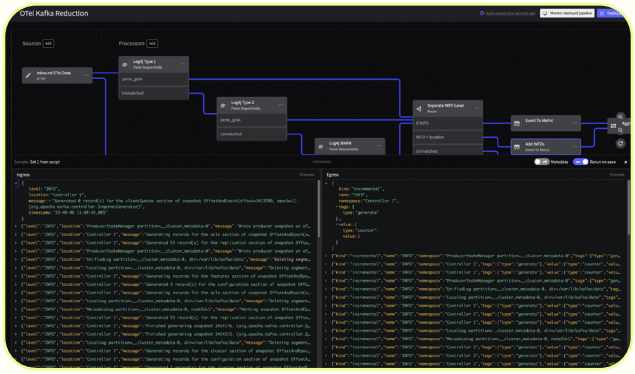
Building on the concept of refinement, a fresh set of challenges arises as systems expand and embrace newer platforms for operational excellence. While integral to stream processing, platforms like Kafka churn out massive volumes of data that can seem overwhelming to manage. At the same time, while renowned for their precision, monitoring solutions like Prometheus produce outputs that, if not streamlined, can become another source of data deluge.

Organizations need a way to manage this deluge of data and refine it, making the data from these streams both manageable and meaningful.



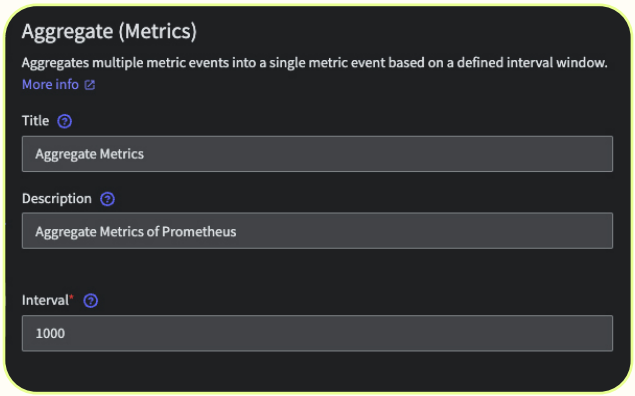
Event-to-Metric Processor configuration form

At Mezmo, we've tailored our Event-to-Metric Processor to address this challenge. By adeptly translating large log data from Kafka into succinct summary metrics, we successfully reduced Kafka logs by over 90%. Refinement like this alleviates data size and event count burdens, optimizing storage requirements and enhancing query speeds.



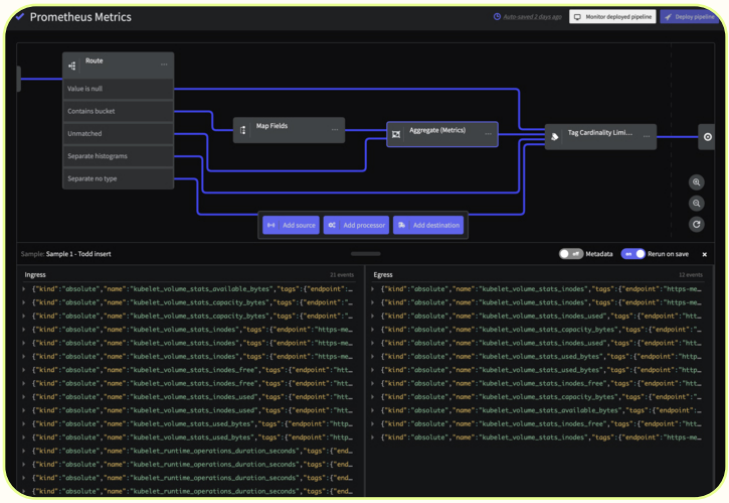
A before/after comparison of data in the events-to-metrics processor using our integrated simulation function

The value extends beyond mere volume reduction, though. The pipeline transforms the logs into clear metrics primed for visualization on operational dashboards. This approach reduces data processing overhead and equips SRE teams with actionable insights.



The Aggregate Processor configuration form

Regarding Prometheus, our Aggregate Processor seamlessly integrates and streamlines its metric output. Resulting in a volume reduction of over 90%. The outcomes are twofold: significant conservation of storage space and a focused emphasis on the most pertinent metrics, ensuring extraneous data don't bog down teams.



Ingress/Egress test results from the Aggregate process using the simulation function

Effective data management is no longer a luxury but a critical necessity. Addressing the challenge of noisy and overwhelming data streams streamlines operations and directly influences decision-making and organizational agility. Organizations can use the Mezmo Telemetry Pipeline to harness this potential, turning data challenges into unparalleled insights and strategic advantage.

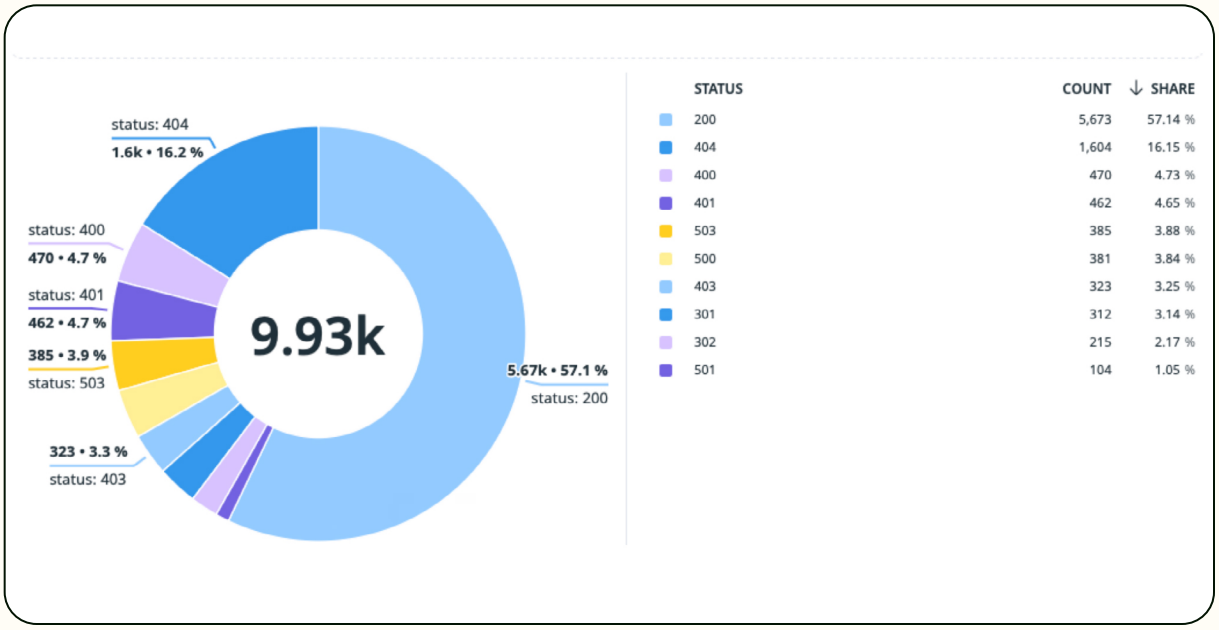
STEP 5: OPERATIONAL EFFICIENCY BOOSTING – AMPLIFYING BACKEND SPEED AND RELIABILITY

There's another dimension that organizations have to deal with when it comes to effective telemetry data management: the increasing complexity of operations. As digital infrastructures evolve and expand, the challenge isn't just about managing the data and ensuring the speed and reliability of the operations that process this data.

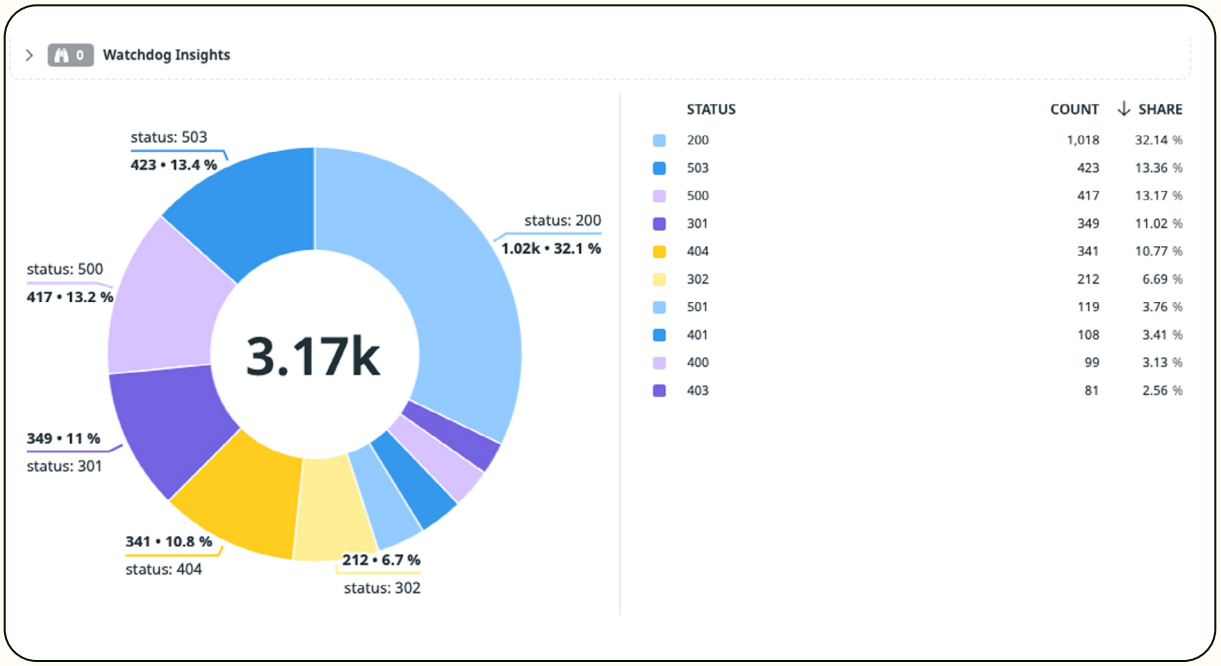
With increasing ingestion rates, there's a pressing need to derive real-time value from data without dealing with inefficiencies. Vendor-lock-in is a genuine concern, and the surging volumes of unstructured data place undue strain on budgets and deployment strategies. Businesses cannot afford to be slowed down by ballooning costs and vendor lock-in.

Grafana is used to analyze reduced observability costs through data efficiency gains

Organizations need agile, robust solutions that streamline operations, maximizing the value derived from their data while making costs manageable.



Datadog Status Server Code, Original Event Distribution



Datadog Status Server Code, Optimized Event Distribution.
Note the reduction of over 70% in overall event volume.

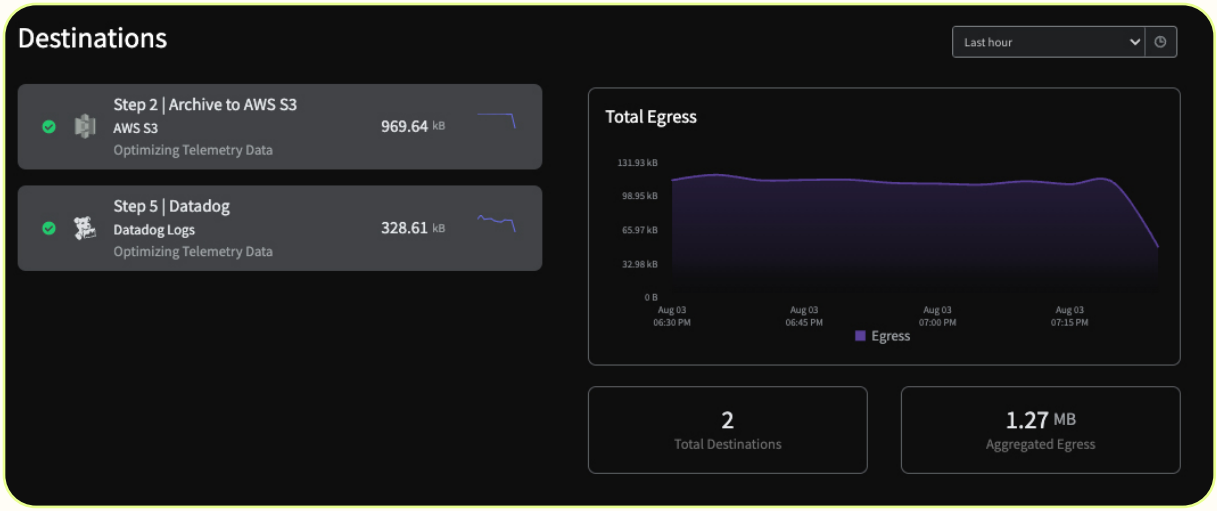
Our platform optimizes telemetry data, turning vast amounts of unstructured data into actionable insights. At the heart of the platform is its data refinement prowess:

- Efficiently eliminating redundant data
- Harnessing cost-effective storage solutions
- Aptly morphing events into aggregated metrics

By integrating the strength of diverse tools into one cohesive platform, Mezmo simplifies data management and further trims operational costs. We can dramatically reduce the data volume to platforms like Datadog, Splunk, Dynatrace, and many more, resulting in minimized data and meticulous tailoring for actionable insights.

To further understand the monumental impact of our approach, consider the volume based on payload size. In subsequent evaluations, the dataset dispatched to the target observability platform saw an impressive 70% reduction.

Data accumulation without purposeful action can lead to stagnation. Efficient backend operations are a critical imperative for streamlining and refining data. Organizations that harness the potential of their data, manage it diligently, and process it all with speed and reliability position themselves at the forefront of their industries with a competitive edge that ensures quicker, more informed decision-making, reduced overheads, and a more straightforward path to innovation.



A destination recap of the egress data sent to Datadog from Mezmo Telemetry Pipeline

KEY FINDINGS RECAP

Through our deep dive into telemetry data management, the following measured improvements have been demonstrated, each carrying profound implications for businesses:

HOLISTIC REDUCTION

Embracing the strategies we proposed can yield over a 50% reduction in telemetry data volume without compromising observability quality. In fact, it allows for maximizing the observable surface area while controlling cost.

This not only enhances system efficiency but significantly curtails operational costs.

WEB LOG REFINEMENT

The Filter technique combined with deduplication delivered a 63% volume reduction from standard web logs like Apache and ginx. This ensures that teams focus on critical events, reducing analysis time and making anomaly detection faster.

KUBERNETES LOG ROUTING

Businesses can optimize storage costs by employing the Route technique to segregate 61% of Kubernetes logs, keeping only essential logs in readily accessible storage while archiving the rest.

KAFKA LOG TRANSFORMATION

A 50% reduction in Kafka logs by extracting common message data ensures a more refined data stream. This precision increases system reliability and avoids overwhelming analysts with noise.

FIREWALL LOG EFFICIENCY

The 94% reduction in Firewall Log volume via the Merge technique saves storage and bolsters security measures by offering a clearer view of potential threats.

LOG-TO-METRIC CONVERSION

By converting logs into metrics, we achieve up to a 90% volume reduction, streamlining data into actionable insights.

These findings highlight that streamlined data management is not just a technical need but a business imperative. Reducing noise and redundancy drives cost savings and empowers teams to make faster and more data-informed decisions.

THE PIVOTAL ROLE OF STRATEGIC TELEMETRY DATA MANAGEMENT

The developers, SREs, and platform owners thrive in the balance between innovation and reliability, with telemetry data serving as their lifeblood. However, this flood of data can lead to a dichotomy: while it promises richer insights, it threatens observability silos.

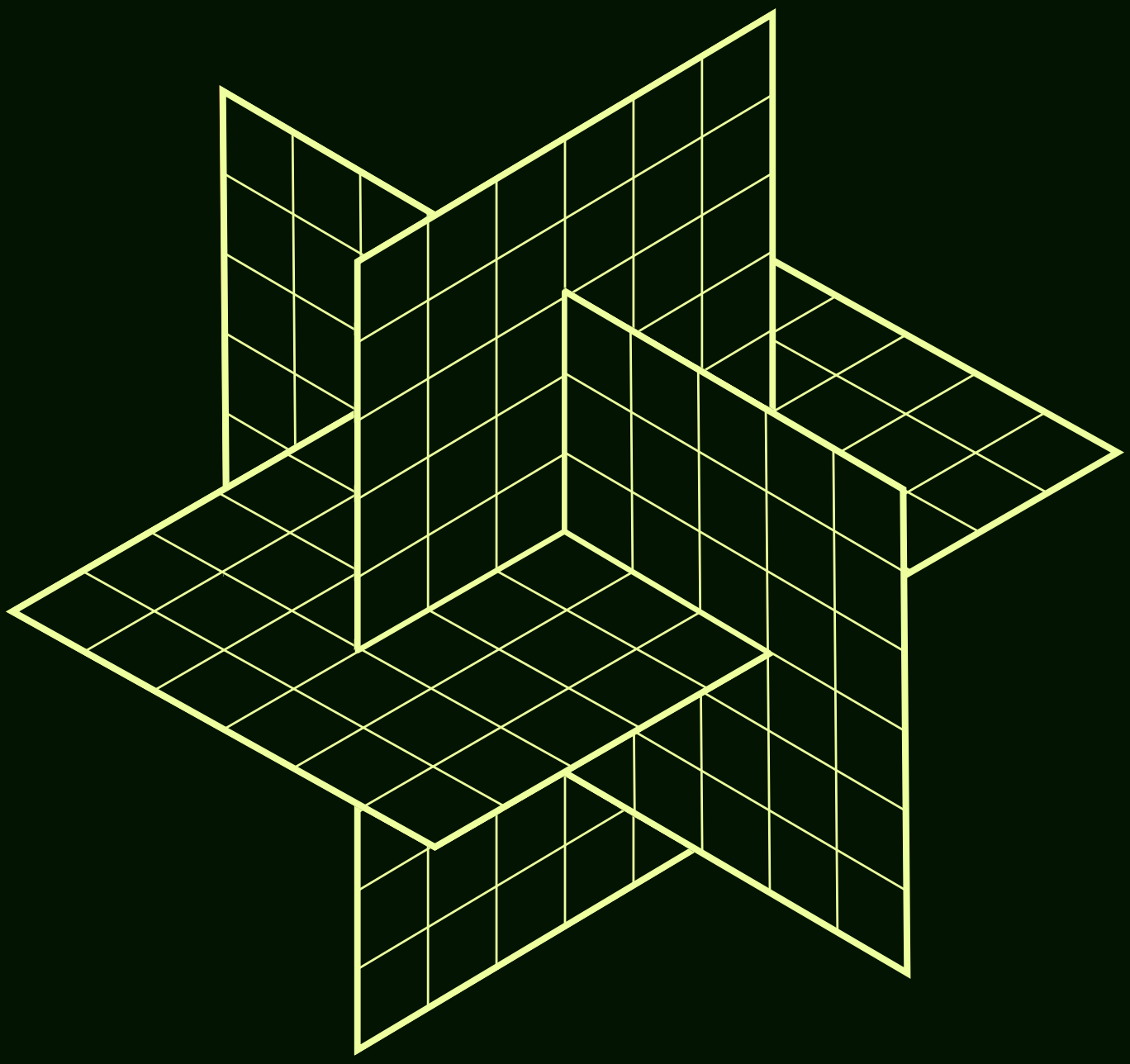
For business decision-makers, refined telemetry data ensures faster, data-driven choices, leveraging timely insights for competitive advantage. Financial decision-makers can appreciate the cost savings and operational efficiencies that streamlined data brings, optimizing budgets and allocating resources effectively. Practitioners can focus on mission-critical tasks with reduced noise and actionable signals, ensuring that digital platforms remain robust and agile.

Aiding in this process is the Mezmo Telemetry Pipeline, which serves as an enterprise platform and a beacon. Our telemetry pipeline sets a new benchmark in telemetry data management, including the 5-step method described above. The strategies outlined in this paper resonate deeply within our pipeline's capabilities, providing unparalleled clarity and focus.

EXPERIENCE DATA OPTIMIZATION WITH MEZMO

Get started for free with Mezmo Telemetry Pipeline today to see its transformative power for yourself, crafted around our unique 5-step process. Telemetry data management should be simple, and your telemetry data shouldn't be a challenge but a catalyst for precise and actionable insights. With Mezmo, developers, platform owners, SREs, and security engineers are mastering telemetry data.

SALES CONTACT: outreach@mezmo.com
SUPPORT CONTACT: support@mezmo.com
MEDIA INQUIRIES: press@mezmo.com



mezmo