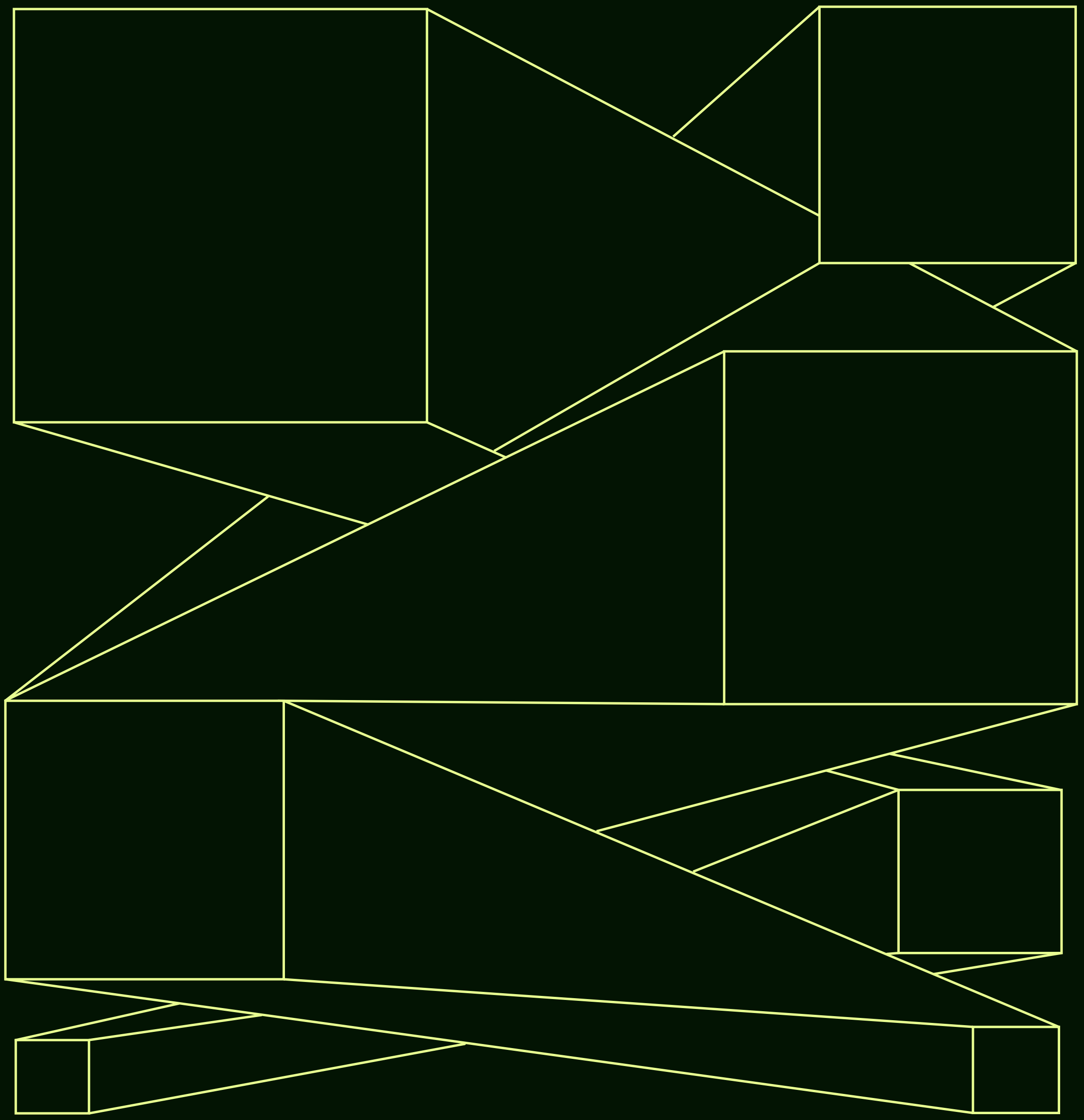
MANAGE TELEMETRY DATA WITH CONFIDENCE



mezmo

WHITE PAPER

INTRODUCTION

In the era of digitization, there is a surge in telemetry data comprising logs, metrics, and traces generated from the many systems, machines, applications, and services that power digital operations. Telemetry data provides valuable business insights through a constant stream of information, such as performance metrics, user sentiment, signals for workflow bottlenecks, and digital signatures of bad actors that impact operational efficiency.

However, organizations face a fundamental challenge as the volume of data grows. The cost and complexity of handling this data increase disproportionately to the value it delivers. The unique nature of telemetry data compounds this challenge because it is dynamic, continuously growing, and constantly changing with sporadic spikes.

Due to the dynamic nature of business and its data, enterprises need help delivering the right data to the right user over time. The data content can change rapidly with any adjustment in infrastructure, application code, or user behavior. Application developers frequently impact the data generated by applications as they rapidly deploy new features and capabilities, and vendors leveraged to provide and enable composite services or architectures are constantly evolving the schema and composition of the data. The criticality of any given data type or field can shift dramatically based on context changes driven by external events, and the delivery of a complete data set often can not be guaranteed.

Because of the volume of data and the multitude of sources it emanates from, the overall content of the data is uncertain, leading to concerns about its completeness and potentially sending sensitive PII information in data streams.

These factors reduce trust in the data collected, causing users to be skeptical about the analysis performed in downstream SIEM or APM systems. At the same time, needs and access requirements change based on evolving circumstances. What might seem insignificant during normal operations becomes crucial during a security breach or system performance incident. This unpredictability necessitates a paradigm shift in how organizations manage and extract value from telemetry data.

WHY IS IT NECESSARY TO HAVE CONFIDENCE IN YOUR TELEMETRY DATA?

To effectively grapple with the complexities of telemetry data, organizations must ensure that they have confidence in the data collected, the enrichments and transformations it requires, and distribution to the proper parties. Teams waste considerable time searching for information. They can only separate signal from noise if they believe that the data they are receiving is correct and of the right quality.

If teams lack confidence in data processing, they will instead turn to collecting every piece of data and pushing it to expensive analytics systems, incurring high costs for computing, data storage, and data egress. More importantly, if teams do not have confidence in the data, they avoid making data-driven decisions and revert to gut feeling or guesswork that they relied on in the past. This lack of confidence in data ultimately impacts service level objectives, MTTx, and, most importantly, customer confidence.

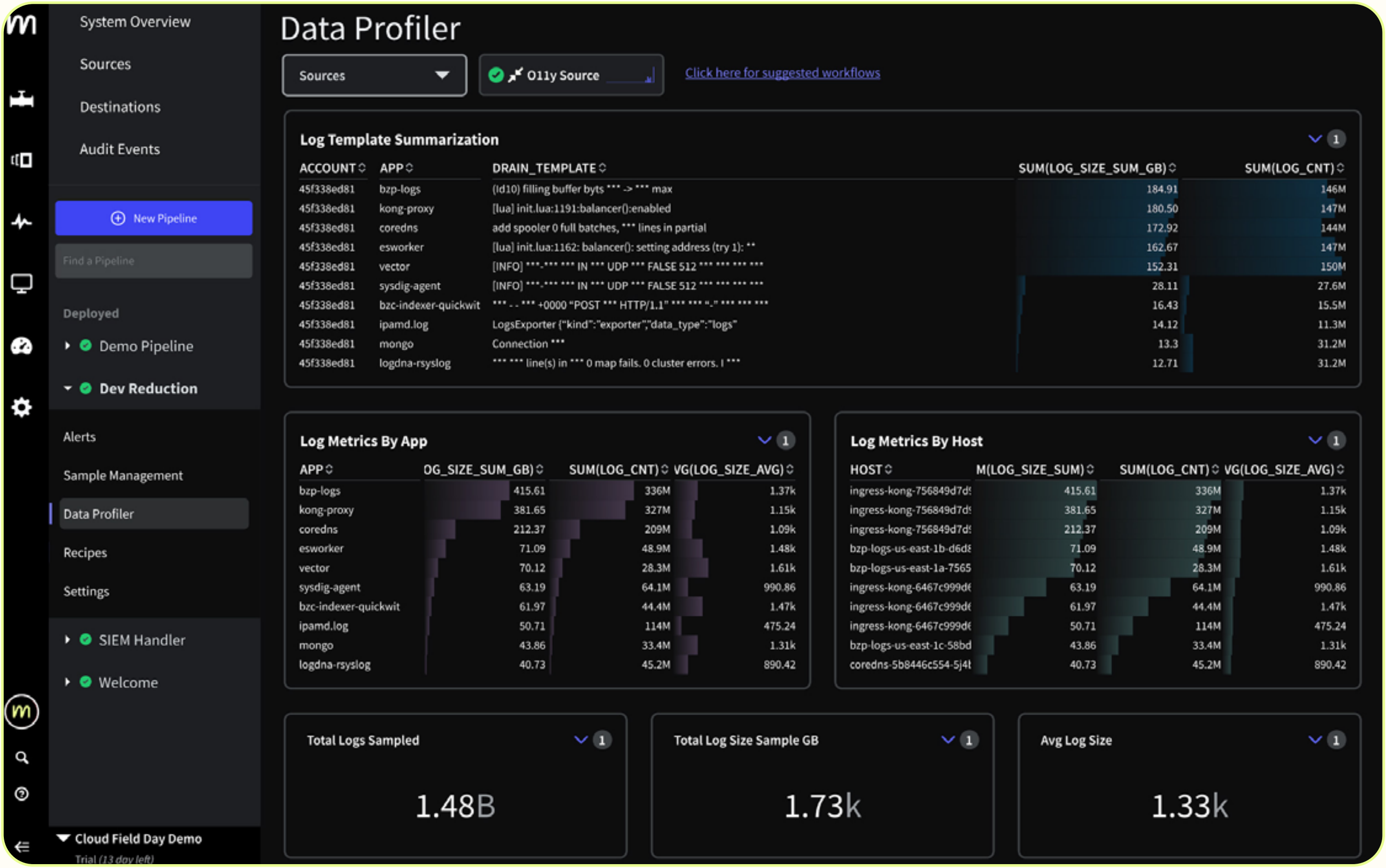
The key reason for the lack of confidence in the data is a lack of understanding and control. The lack of understanding is created by a confluence of realities present at various levels of severity in every organization. However, all result in the same problem: Enterprises simply do not understand their telemetry data. The foundational and unaddressed problem is that those responsible for managing the data, the cost of the data, and the tools that allow the consumers of the data to access the data need to be in control of the data.

CONFIDENCE IN DATA STARTS WITH DATA UNDERSTANDING

To address this challenge, we recommend taking a Data Engineering approach. While the nature of telemetry data is unique, and the workflows of the owners of telemetry data are different from traditional data engineering, the principles that drive data engineering all apply.

Applying data engineering techniques to telemetry data means monitoring, tracking, and analyzing the data in real time to ensure its reliability, accuracy, and compliance with required standards. Understanding the data involves capturing and analyzing its structure, content, lineage, repeated patterns, required transformations, and other dependencies.

Understanding data starts with Data Profiling, the first step towards determining how well data meets the expectations and requirements for analytics. It involves assessing various aspects of your data, such as log patterns, frequency, completeness, validity, timeliness, consistency, and relevance. This assessment is essential for ensuring your data is trustworthy, as it affects your data analysis, reporting, and decision-making. To undertake the examination that will help you understand and trust your data, you need a data profiling tool to parse data, detect patterns, and regularly identify the frequency of redundant and repeated patterns in your data.



Data Profiler

Log Template Summarization

ACCOUNT	APP	DRAIN_TEMPLATE	SUM(LOG_SIZE_SUM_GB)	SUM(LOG_CNT)
45f338ed81	bzp-logs	(Id10) filling buffer byts *** -> *** max	184.91	146M
45f338ed81	kong-proxy	[lua] init.lua:1191:balancer():enabled	180.50	147M
45f338ed81	coredns	add spooler 0 full batches, *** lines in partial	172.92	144M
45f338ed81	esworker	[lua] init.lua:1162: balancer(): setting address (try 1): **	162.67	147M
45f338ed81	vector	[INFO] ***-*** *** IN *** UDP *** FALSE 512 *** *** *** ***	152.31	150M
45f338ed81	sysdig-agent	[INFO] ***-*** *** IN *** UDP *** FALSE 512 *** *** *** ***	28.11	27.6M
45f338ed81	bzc-indexer-quickwit	*** - - *** +0000 "POST *** HTTP/1.1" *** *** "-" *** *** ***	16.43	15.5M
45f338ed81	ipamd.log	LogsExporter {"kind":"exporter","data_type":"logs"	14.12	11.3M
45f338ed81	mongo	Connection ***	13.3	31.2M
45f338ed81	logdna-rsyslog	*** *** line(s) in *** 0 map fails. 0 cluster errors. I ***	12.71	31.2M

Log Metrics By App

APP	OG_SIZE_SUM_GB)	SUM(LOG_CNT)	VG(LOG_SIZE_AVG)
bzp-logs	415.61	336M	1.37k
kong-proxy	381.65	327M	1.15k
coredns	212.37	209M	1.09k
esworker	71.09	48.9M	1.48k
vector	70.12	28.3M	1.61k
sysdig-agent	63.19	64.1M	990.86
bzc-indexer-quickwit	61.97	44.4M	1.47k
ipamd.log	50.71	114M	475.24
mongo	43.86	33.4M	1.31k
logdna-rsyslog	40.73	45.2M	890.42

Log Metrics By Host

HOST	M(LOG_SIZE_SUM)	SUM(LOG_CNT)	VG(LOG_SIZE_AVG)
ingress-kong-756849d7d	415.61	336M	1.37k
ingress-kong-756849d7d	381.65	327M	1.15k
ingress-kong-756849d7d	212.37	209M	1.09k
bzp-logs-us-east-1b-d6d	71.09	48.9M	1.48k
bzp-logs-us-east-1a-7565	70.12	28.3M	1.61k
ingress-kong-6467c999d	63.19	64.1M	990.86
ingress-kong-6467c999d	61.97	44.4M	1.47k
ingress-kong-6467c999d	50.71	114M	475.24
bzp-logs-us-east-1c-58bd	43.86	33.4M	1.31k
coredns-5b8446c554-5j4	40.73	45.2M	890.42

Total Logs Sampled

1.48B

Total Log Size Sample GB

1.73k

Avg Log Size

1.33k

The Mezmo Data Profiler

Once you understand your data, the next step is to Optimize it so users can have confidence in it for their downstream applications. The solution is to set up Telemetry Pipelines to collect, understand, and optimize data so that users can confidently utilize that data in downstream applications. Pipelines process the data, transform and format it, and check for sensitive information to ensure that the right data, in the right format, is delivered to the right applications and teams in a timely and compliant manner.

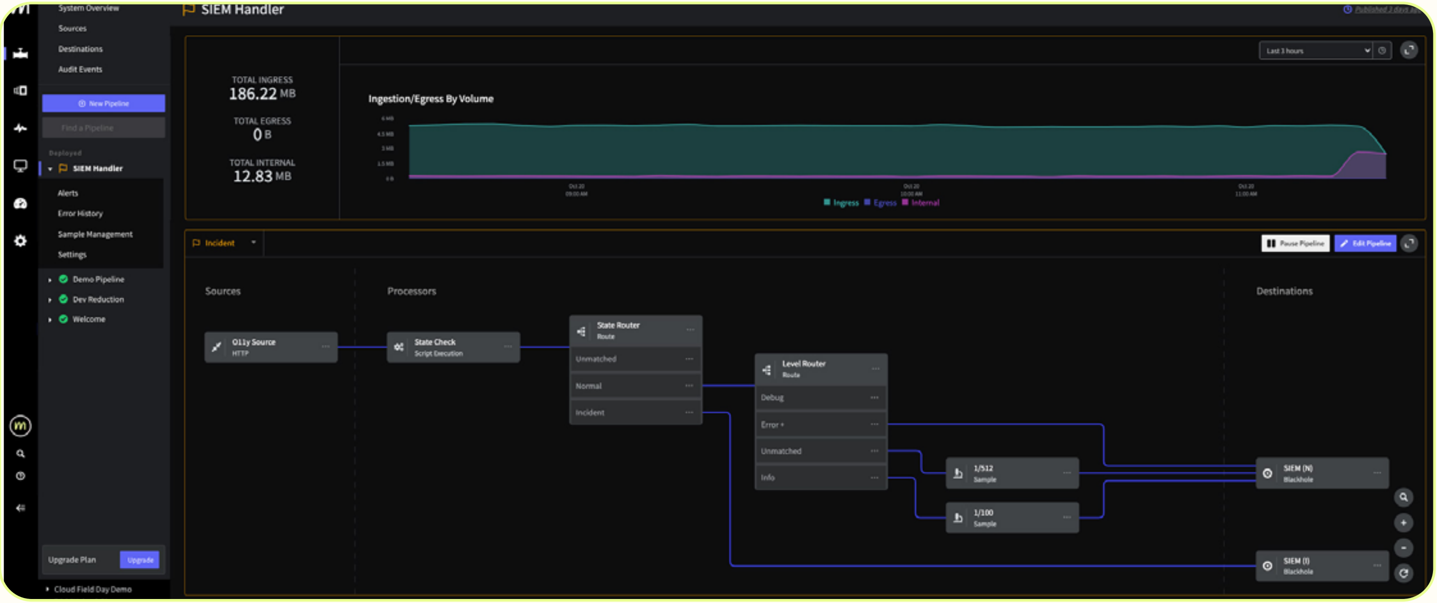
Here are seven ways that pipelines can help build confidence in your data:

CONFIDENCE IN THE DATA OPTIMIZATIONS

Telemetry pipelines offer capabilities to optimize data for cost reduction and to extract business insights. You can reduce, filter, sample, and aggregate data to manage the volume of data reaching the expensive analytics systems. You can also create metrics from logs and events or extract the metrics from them, reducing the volume of data by up to 90%.

While you are reducing the data, teams need to trust that the data egressing the pipeline is valid and consists of the correct information. At any time in the pipeline, you can tap and see the data flow passing through the data processor. You can run simulations to ensure the pipeline is transforming, formatting, and reducing the data as you expect.

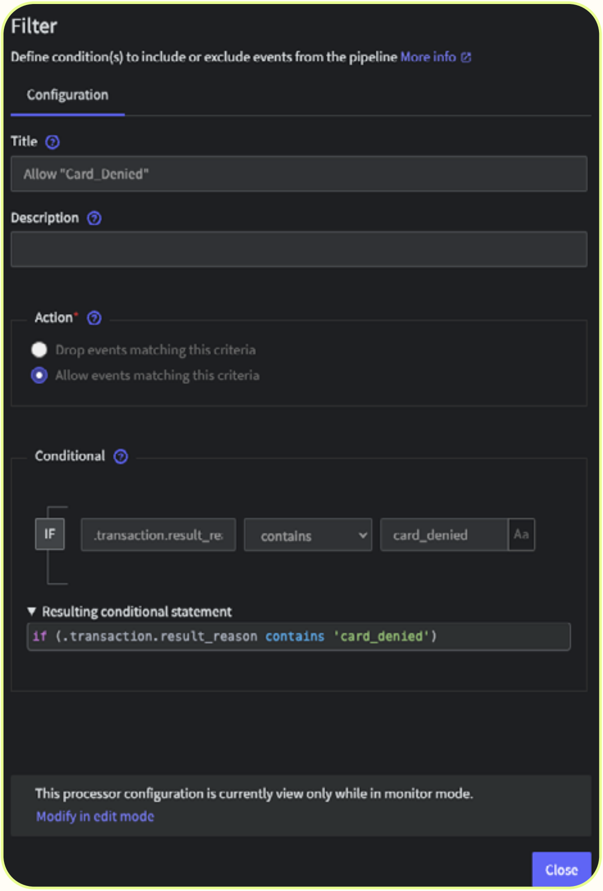
We also know that as business changes, the data will change. Any change in application code can open the floodgates of unanticipated logs that can consume your month's observability budget—in just days or hours. Users can configure the telemetry pipeline to detect such data drifts and protect their organization from unexpected costs. Timely alerts for any unexpected data spikes help you confidently consume the data.



A Mezmo Telemetry Pipeline for Data Optimization

CONFIDENCE IN DATA STORAGE

Due to the high cost of data storage in SIEM or analytics systems such as Splunk or DataDog, most organizations filter or sample the data before sending it to these systems. However, your compliance requirements, or need for data for future incident debugging, may require you to store the complete datasets for a period, generally seven days and, in some cases, up to a month. In such situations, the telemetry pipeline can send only a sample of data to analytics platforms and divert the rest to low-cost storage such as AWS S3. Teams can be confident in case of any compliance audits, or if they need data to investigate a security breach, the data can be rehydrated through the pipeline.



Filter

Define condition(s) to include or exclude events from the pipeline More info

Configuration

Title: Allow "Card_Denied"

Description:

Action*
○ Drop events matching this criteria
● Allow events matching this criteria

Conditional

IF .transaction.result_re contains card_denied

▼ Resulting conditional statement

```
if (.transaction.result_reason contains 'card_denied')
```

This processor configuration is currently view only while in monitor mode.
Modify in edit mode

Configuration of the Filter Processor in a Mezmo Telemetry Pipeline

CONFIDENCE IN DATA COMPLIANCE

Organizations must adhere to many privacy laws, including GDPR, CCPA, and HIPAA. Telemetry data such as logs, events, and traces may contain PII information that, if not appropriately scrubbed, may lead to unintended distribution of personal data and exposure to regulatory fines. The Telemetry pipeline offers features like redact, mask, encrypt, and decrypt to manage the PII data.

As the data moves through the pipeline, users can ensure that the platform will share personal information only if it is required by systems such as SIEM. For other analytical systems, it can be redacted or aggregated into metrics.

CONFIDENCE IN DATA ACCESS AND ROUTING

Data access and breaking departmental silos have been a long-standing challenge for DevOps, Security, and SRE teams. Data is often sent to a system, locked in, and rendered unavailable to other systems and teams.

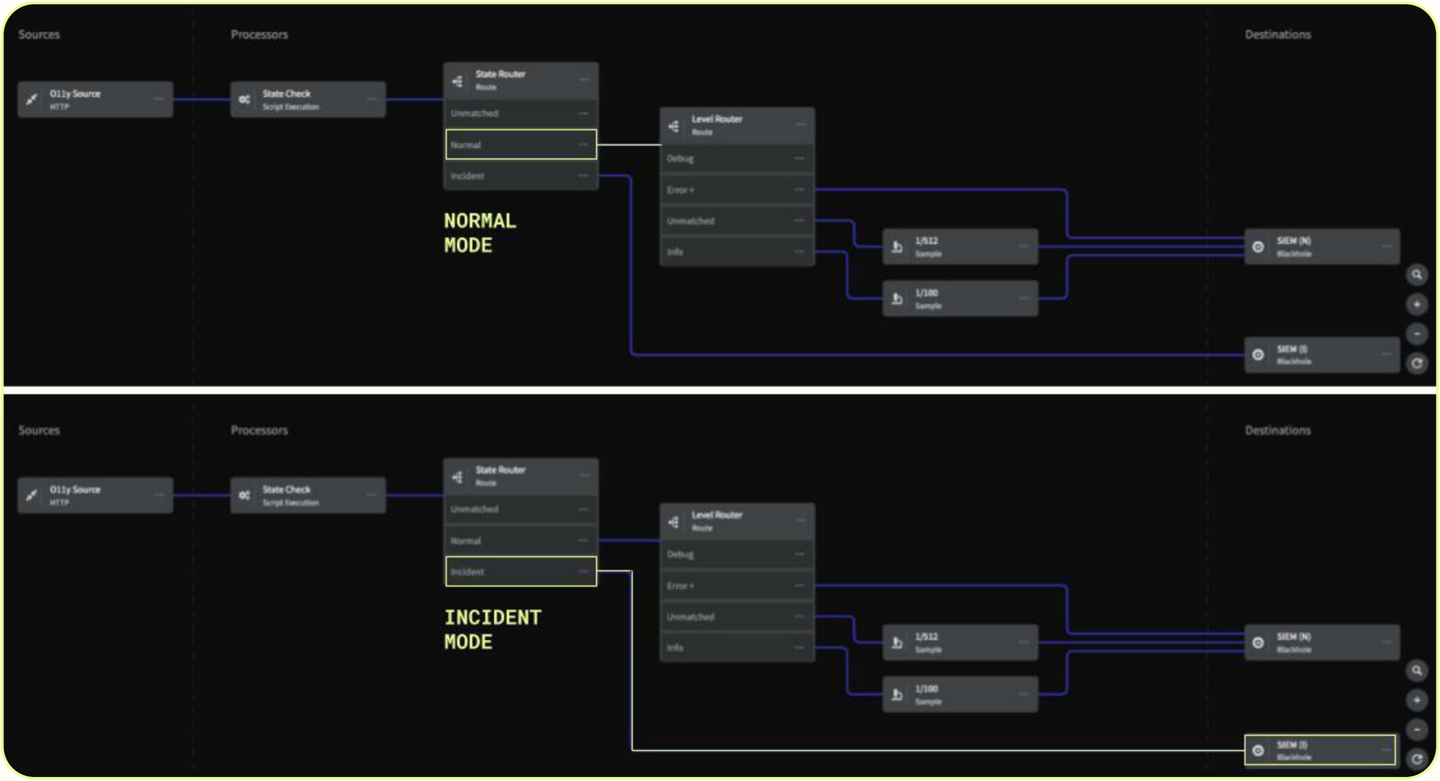
With the telemetry pipeline acting as the central collector and distributor of the data, enterprise teams can ensure that the correct data is available to teams so they can perform their jobs effectively. It also ensures that the users get only the data they need and have authorization to access it. Incorporating data engineering principles of data cataloging, governance, and policy enforcement helps with access control.

CONFIDENCE IN INCIDENT RESOLUTION

DevOps and security teams rely on telemetry data to resolve incidents. These can range from performance issues to security breaches, Challenging teams with competing objectives of reducing MTTx and observability budgets. There is always a concern that they may not capture enough of the required data in case of an incident. The result is that they tend to capture and store all the data from a selected set of systems while leaving a vast amount of observability surface area uncovered.

With telemetry pipelines, you can cost-effectively capture all the data you need and send samples to the high-cost analytics systems. If an incident happens, a responsive pipeline can switch to an incident mode, sending the full-fidelity data to systems, e.g., SIEM, when an incident is detected. Once the incident is resolved, the pipeline can revert to the normal sampling mode.

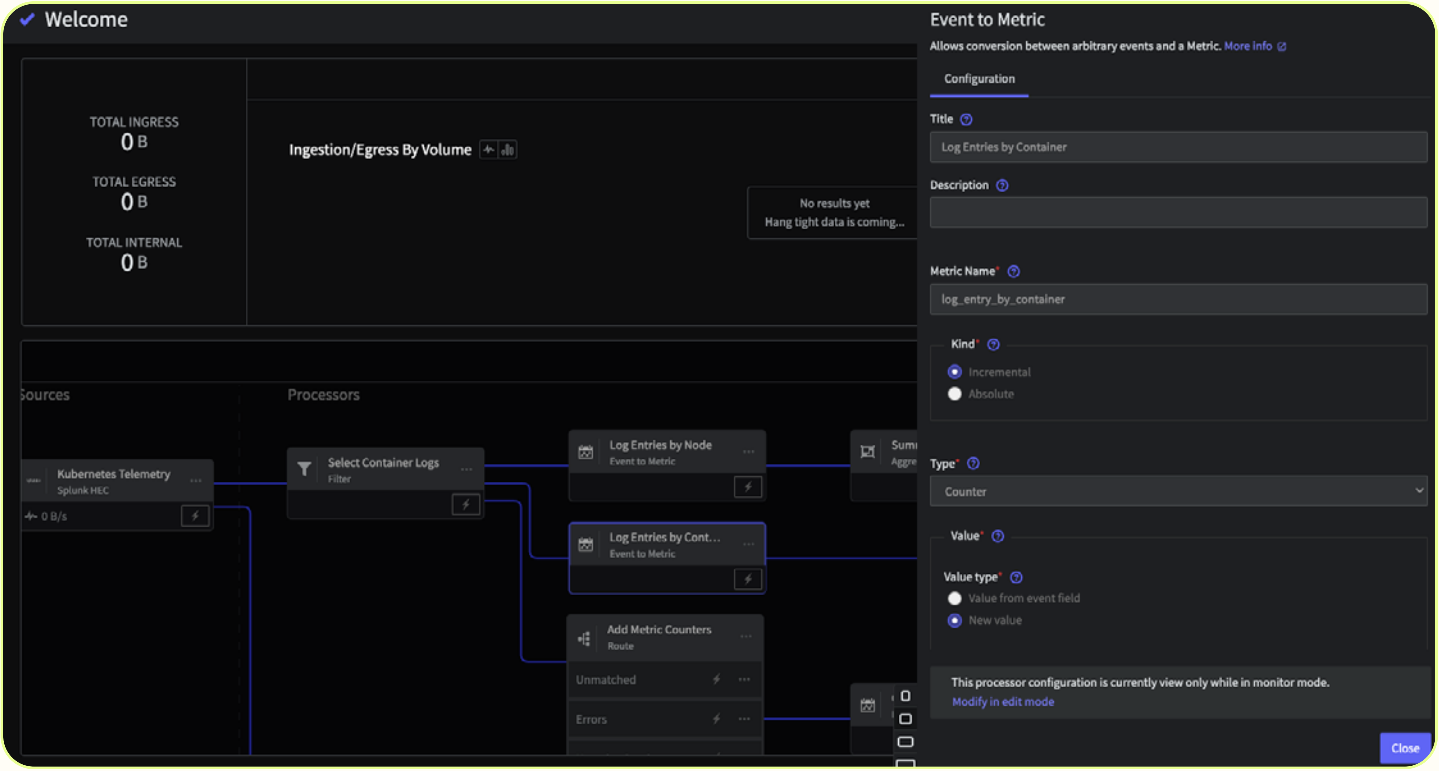
Therefore, with the pipeline, teams can rest assured that they will always have access to the correct data when needed.



A Responsive Mezmo Telemetry Pipeline activated during a detected incident

CONFIDENCE IN BUSINESS INSIGHTS

A tremendous amount of information in the telemetry data is essential not only for debugging and incident resolution but also for extracting business insights. If you are not using the telemetry data for those, you are not getting the total value from the data. Telemetry data can give you business insights in real time. You can extract e-commerce insights through metrics such as product orders, cart checkouts, and performance from transaction events and logs. Pipelines can help you derive those metrics or create new metrics from the events and logs in real time.



Event to Metric

Allows conversion between arbitrary events and a Metric. More info

Title: Log Entries by Container

Description:

Metric Name*: log_entry_by_container

Kind*: ● Incremental ○ Absolute

Type*: Counter

Value*

Value type*: ○ Value from event field ● New value

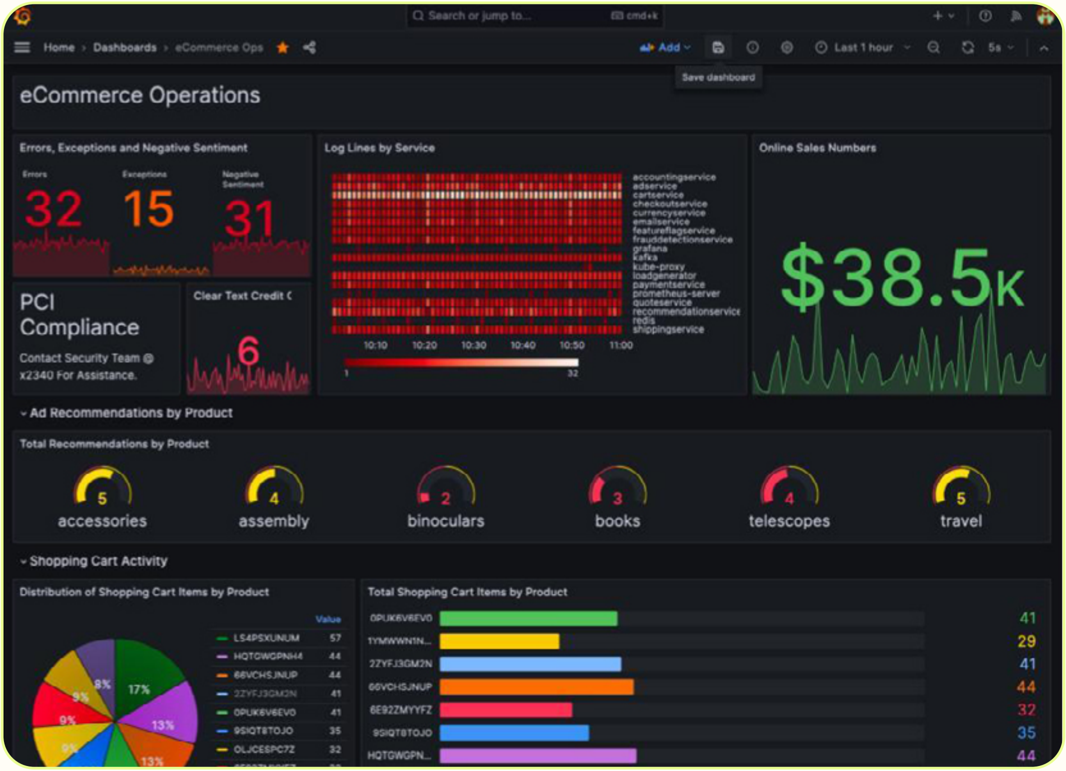
This processor configuration is currently view only while in monitor mode.
Modify in edit mode

Capturing event counts and converting them into metrics for dashboarding

With the pipeline, you can confidently view and analyze your reports as the data is aggregated, enriched, and delivered in the proper formats for easy visualization in tools like Grafana. In addition, you can create a loop pipeline between your data lake as a source and the destination to ensure that the data is accurate or needs updates resulting from changes in the data from some other source.

CONFIDENCE THAT THE DATA IS CURRENT

As the data sources and content keep changing, data users require that the data they use for incident resolution or decision-making is current. With a telemetry pipeline, you can quickly onboard new data sources, format the data, and ready it for use immediately. In addition, you can refresh the data in data lakes with additional information. As the data gets stored in data lakes, you may need to update or add additional information. In such cases, a "loop pipeline" can take the data from the lake, run it through it, and put it back into the data lake with all current information.

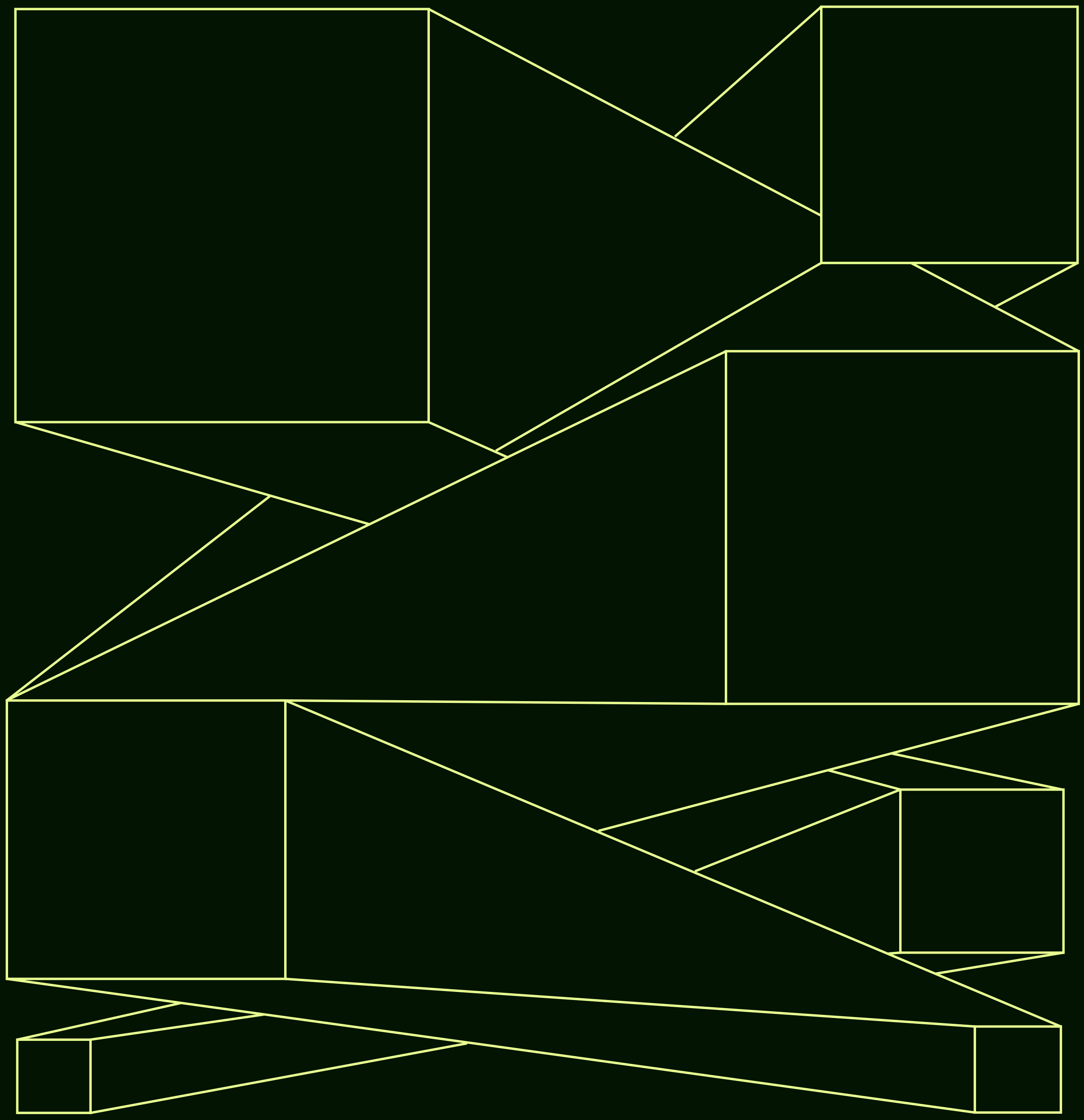


Dashboard created directly from metrics captured and generated from logs

CONCLUSION

With current complex technology ecosystems, the need for observability and appetite for telemetry data will keep growing. Telemetry data must also be treated like business data, as an enterprise asset. Organizations must have confidence in its content and quality to extract the complete value from the data, proportional to the investment into compute and storage. Companies increasingly rely on central telemetry pipelines to build that trust across the enterprise.

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



mezmo