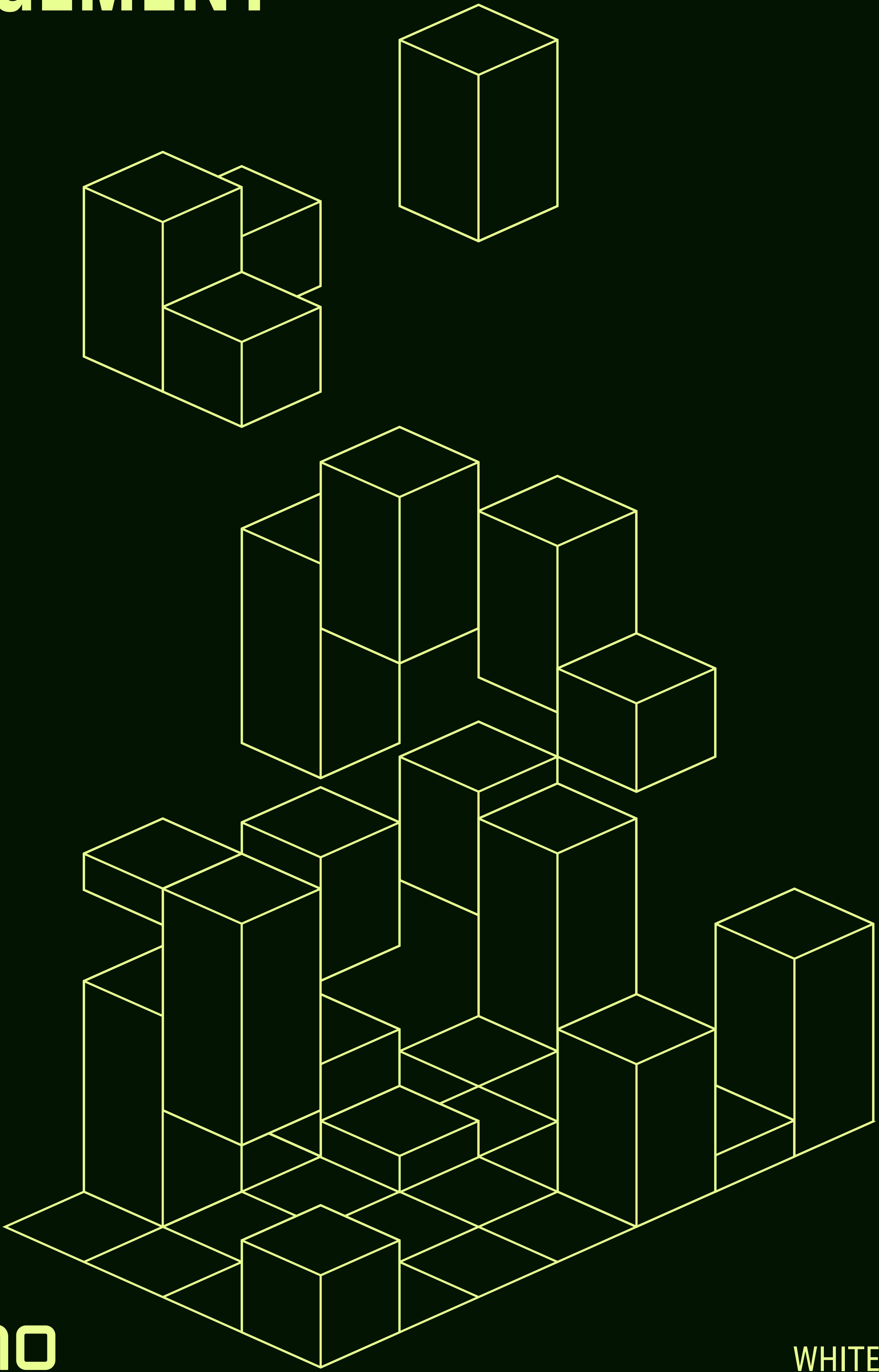
MEZMO'S METHODOLOGY FOR TELEMETRY DATA MANAGEMENT



mezmo

WHITEPAPER

ADDING VALUE TO TELEMETRY DATA

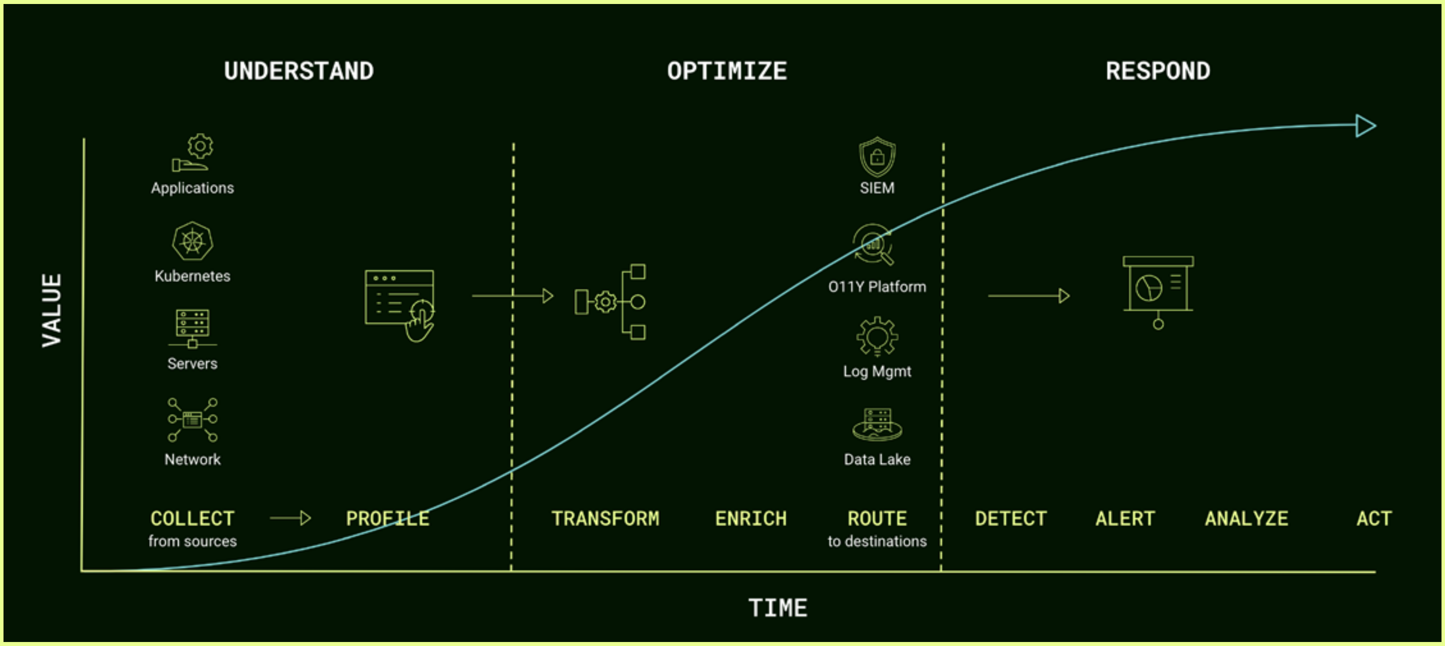
While there has long been an opinion that telemetry data is simply an exhaust byproduct generated by operational systems that has little value, organizations are increasingly realizing that, in this voluminous output, there is data that can be used to ensure business continuity, drive cost reduction efforts, understand buyer sentiment, ensure compliance to legal regulations, and identify new business opportunities.

However, in its raw output, telemetry data initially has low value. Like any raw resource, it must be processed, optimized, and distributed downstream for its value to be realized. The way in which this is achieved is through a telemetry pipeline that transforms your raw data, increasing its value through each incremental step.

While some may consider a telemetry pipeline simply as a means for getting log data from one point to another, we think of it as a way to add value to your data. When we talk about data optimization, we don't mean just re-formatting it for easy loading into storage or an observability tool, we mean transforming that data into useful information that adds value to your business, and adds value to your storage solution by making sure you only pay for data that you really need. When the optimized information is then distributed downstream, it becomes an enterprise asset that can be delivered to Engineering, DevOps, Security, AIOps, and the Business in a form that is most useful to them, rather than being locked into a "single pane" solution.

In this White Paper, you'll learn about Mezmo's methodology for managing telemetry data, and how our three-phase approach helps you understand your data, optimize it to provide useful information, and then respond to that information.

This chart illustrates the Mezmo perspective on how a telemetry pipeline increases the value of telemetry data through each step of the Pipeline. In the next sections, you'll learn more about how the three phases of Understand, Optimize, and Respond each contribute to the incremental value chain of telemetry data, and how this is reflected both in our product and in our methodology.



THE UOR APPROACH TO TELEMETRY DATA MANAGEMENT

Designing a telemetry pipeline that adds value to your data involves three phases:

1. Understanding your data.
2. Optimizing your data.
3. Responding to changing data and business needs.

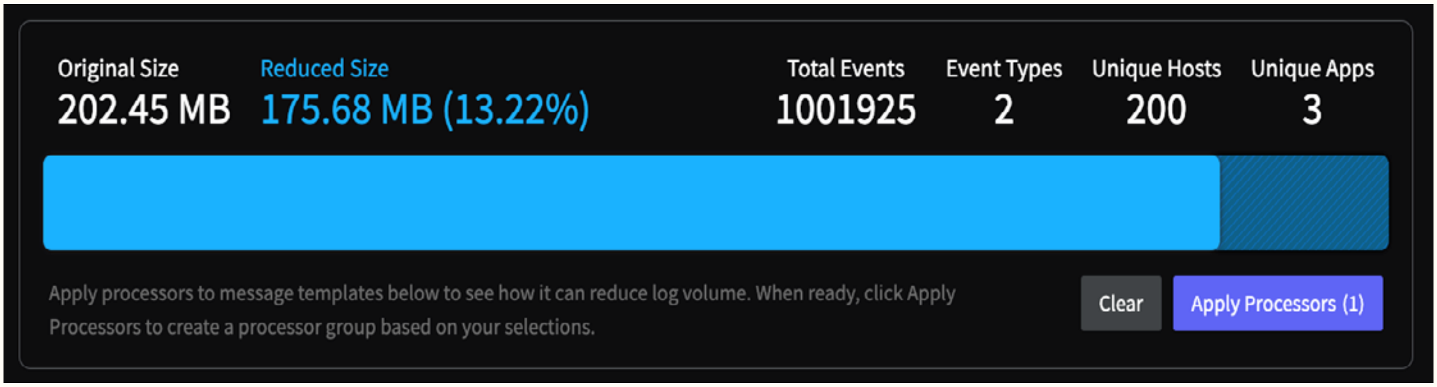
This approach to managing telemetry data, which we call UOR (Understand, Optimize, and Respond), is grounded in data engineering principles that have been informed by our experiences in working with customers to uncover and understand their business needs. In each of the following sections, this paper will describe those best practices.

PHASE 1: UNDERSTAND THE DATA

During our initial research into our customer's needs for telemetry data management, we consistently encountered a significant barrier to realizing the value of the data: customers didn't know what was in their data, much less how to optimize or manage it. Or they knew there was data they wanted, but couldn't get to because of how it was formatted or being transmitted. These issues were compounded in environments where multiple systems and applications were sending log data, making the configuration of observability tools extremely difficult for targeted information.

As the first step to add value to telemetry data, we recommend profiling the data to understand it. This means identifying patterns in telemetry data for repeatable, redundant, and usable information, which in turn will inform the design of the routing and processor chains within the telemetry pipeline. The Mezmo Platform includes the Data Profiler that helps to understand the data before developing a Pipeline to manage and optimize it.

The Data Profiler in the Mezmo Platform enables rapid identification and understanding of the most common event types found in the source data. The Log Composition analysis component of the Data Profiler provides an overview of the total number of events and their total size, the number of unique event types, unique hosts, and unique applications for the log types identified through auto-parsing.



Original Size 202.45 MB

Reduced Size 175.68 MB (13.22%)

Total Events 1001925

Event Types 2

Unique Hosts 200

Unique Apps 3

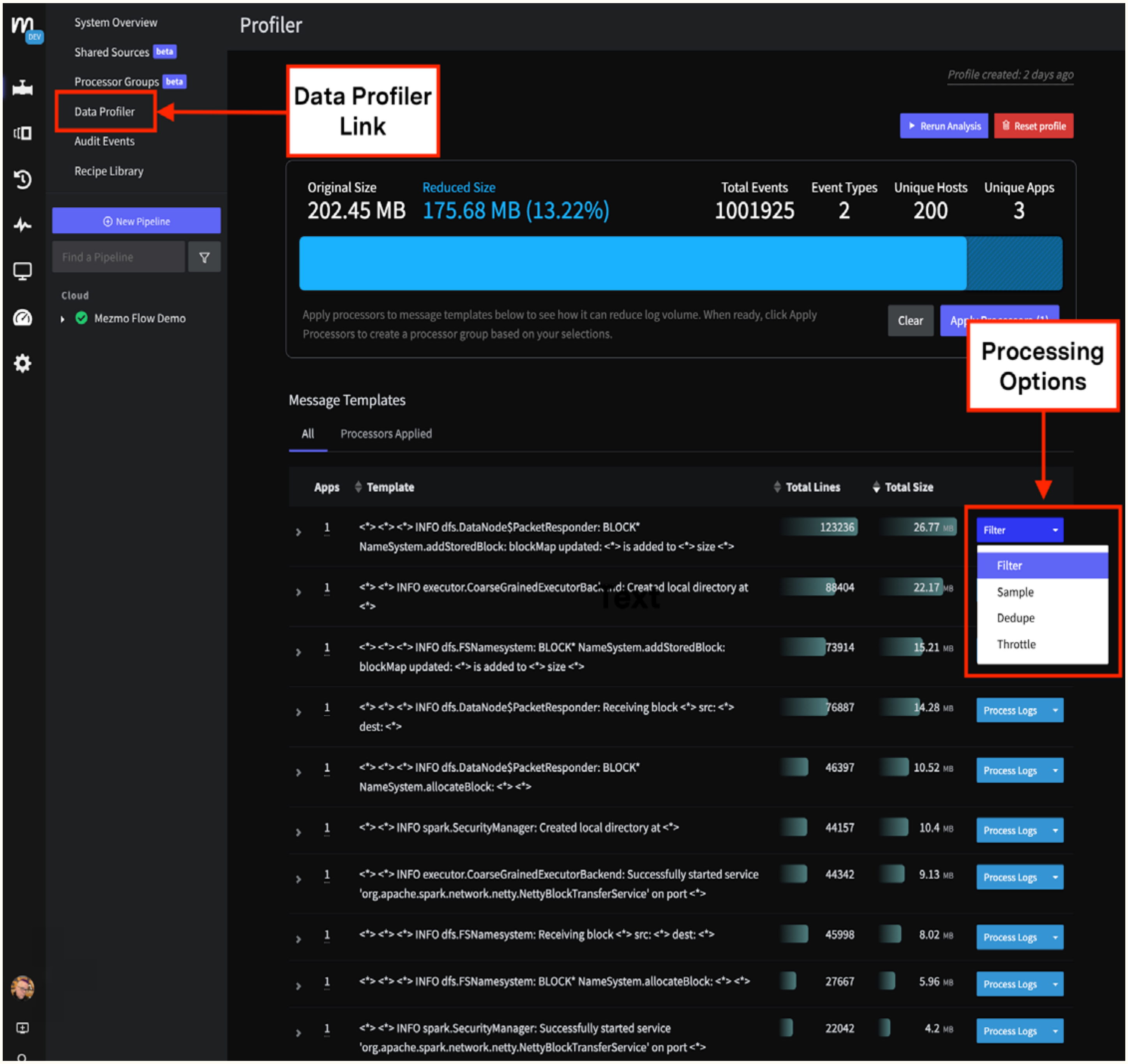
Apply processors to message templates below to see how it can reduce log volume. When ready, click Apply Processors to create a processor group based on your selections.

The output of the Data Profiler showing the total number of events, hosts, event types, and applications contained in the telemetry data, along with the log volume reduction after applying the Filter Processor.

Within the Data Profiler, the Templates section displays the patterns within the logs. These include the standard format for each pattern, as well as the total number of lines that contain each pattern, and the total size of those log lines within the overall log volume.

With the identification of the patterns within the log data, the Data Profiler can then offer recommendations on which actions to take for those messages, and which Processors are best suited for managing them. As you select processors, you can see the effect that applying a processor will have on overall log volume, and your selected processors will be added to the pipeline as a processor group.

You can create a data profile as part of the Mezmo Flow onboarding process, or by setting up a Data Profiling Processor at the beginning of your pipeline. Once your source data is analyzed, you can access the complete profile through the Data Profiles link in the Mezmo Web App, and re-set or re-run the analysis at any time.



Data Profiler Link

Processing Options

Profile created: 2 days ago

Original Size	Reduced Size	Total Events	Event Types	Unique Hosts	Unique Apps
202.45 MB	175.68 MB (13.22%)	1001925	2	200	3

Apply processors to message templates below to see how it can reduce log volume. When ready, click Apply Processors to create a processor group based on your selections.

Message Templates

All | Processors Applied

Apps	Template	Total Lines	Total Size	
1	<*> <*> <*> INFO dfs.DataNode$PacketResponder: BLOCK* NameSystem.addStoredBlock: blockMap updated: <*> is added to <*> size <*>	123236	26.77 MB	Filter
1	<*> <*> INFO executor.CoarseGrainedExecutorBackend: Created local directory at <*>	88404	22.17 MB	
1	<*> <*> <*> INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: <*> is added to <*> size <*>	73914	15.21 MB	
1	<*> <*> <*> INFO dfs.DataNode$PacketResponder: Receiving block <*> src: <*> dest: <*>	76887	14.28 MB	Process Logs
1	<*> <*> <*> INFO dfs.DataNode$PacketResponder: BLOCK* NameSystem.allocateBlock: <*> <*>	46397	10.52 MB	Process Logs
1	<*> <*> INFO spark.SecurityManager: Created local directory at <*>	44157	10.4 MB	Process Logs
1	<*> <*> INFO executor.CoarseGrainedExecutorBackend: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port <*>	44342	9.13 MB	Process Logs
1	<*> <*> <*> INFO dfs.FSNamesystem: Receiving block <*> src: <*> dest: <*>	45998	8.02 MB	Process Logs
1	<*> <*> <*> INFO dfs.FSNamesystem: BLOCK* NameSystem.allocateBlock: <*> <*>	27667	5.96 MB	Process Logs
1	<*> <*> INFO spark.SecurityManager: Successfully started service 'org.apache.spark.network.netty.NettyBlockTransferService' on port <*>	22042	4.2 MB	Process Logs

An example Data Profile showing the most common message patterns found in the data source, as well as processing options for specific message templates.

PHASE 2: OPTIMIZE THE DATA

Having gained an understanding of the telemetry data that is sent by all your sources, you can now focus on optimizing it to derive value from it. In many cases there are standard processing chains, or Processor Groups, that you can use for specific data types.

For example, if your goal is to reduce the volume of data you send to security or an observability platform, you can implement the recommended five steps described in the guide "Optimize Your Telemetry in Five Steps":

1. Filter

2. Route

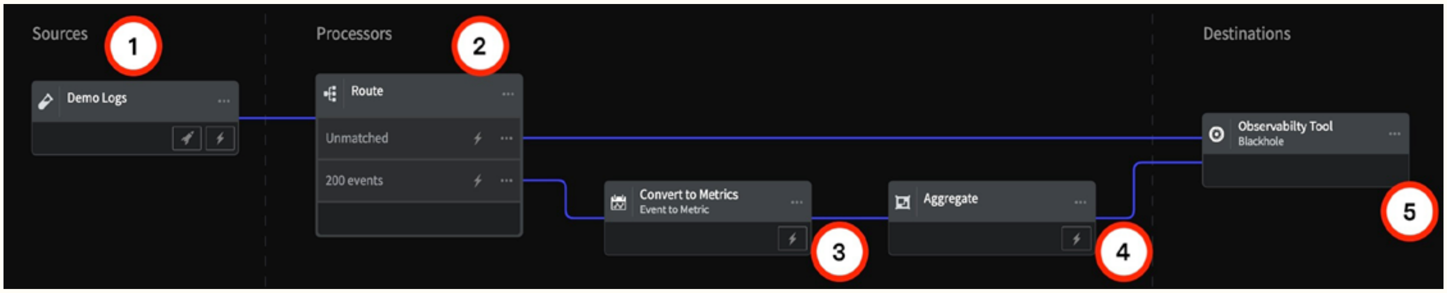
3. Trim & Transform

4. Merge

5. Condense Events to Metrics

Based on our research, applying these five steps can result in an over 70% reduction in data volume from the pipeline source to the destination. For example, converting events to metrics can reduce the volume of HTTP log data by up to 50% on its own.

This same technique is also useful for generating business insights by separating telemetry and business events for processing through different routes, and then converting the business events to meaningful metrics. For example, transaction events in themselves don’t contain much information, but changes in transaction event volume over time can be correlated with system events, marketing campaigns, and other business data to understand the relationship between the data points.



Typical configuration of an Event-to-Metric processor group for data volume reduction. Starting with the Source (1), this group includes a Route Processor (2), an Event-to-Metric Processor (3), and an Aggregate Metrics Processor (4), with an observability tool or other Destination (5) terminating the Pipeline.

PHASE 3: RESPOND TO INFORMATION

Your business and data are dynamic and change with time. Changes in source data or incidents detected in the data destination may change your data requirements, and your telemetry pipeline should be able to respond to that. Optimized data becomes information when it provides you with the ability to act and make decisions, either manually or programmatically. For example, in the normal functioning of a system, 200-OK responses contain little actionable information. It's only when they are converted to metrics, and you are able to detect sudden changes in the number of those responses over a time period, that the metric becomes meaningful because you now have potentially actionable information about the health of the system under observation. Because being able to respond to information is a critical component of telemetry data management, the Mezmo Platform incorporates both Alerts that can be triggered based on defined changes to the data in stream, and Responsive Pipelines that can change their functioning based on those alerts or even based on incidents detected in observability or SIEM platforms.

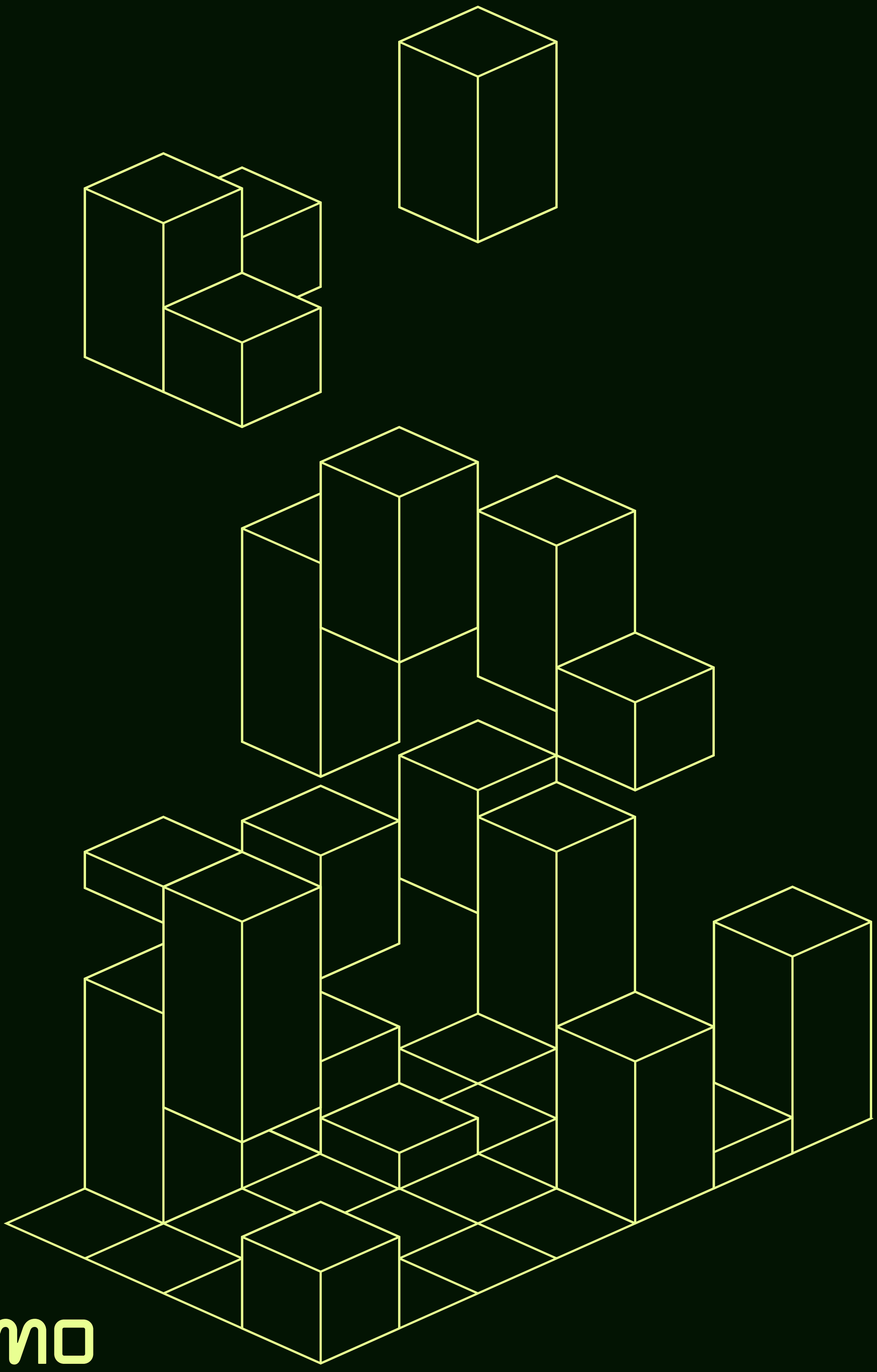
In the example of 200-OK responses, you can use the Mezmo Platform to understand the number of responses to expect during a defined internal under normal conditions, and then set a threshold that would trigger an Alert to a monitoring tool if the number falls below a percentage of that number. That alert could, in turn, trigger the Pipeline to shift from Normal to Incident mode, in which full-fidelity versions of the logs with the individual 200-OK events are sent to a tool like Mezmo Log Management, rather than being converted to metrics. You may also have a situation where you use a telemetry pipeline to reduce your log volumes before sending them to Splunk to manage your spending. You have determined that 40 percent of the data is good enough in a steady state. But suddenly, one day, your SIEM system detects an incident, and you need all the data to resolve that issue. In such a situation, pipelines must reroute the data and send full-fidelity information to SIEM until the issue is resolved."

CONCLUSION

A telemetry pipeline is more than just a centralized place to manage and route telemetry data, it is also a tool for you to understand, optimize, and respond to your data. Without the right context, telemetry data's value is minimal, noisy, and requires a lot of mental toil to interpret. By first understanding what's in your telemetry data, you will start to notice common patterns, identify redundant information, and discover ways to take action. With this understanding, you can optimize your data so it becomes easier to generate business insights, and implement systems that will enable you to respond to the dynamic and constantly changing nature of your data.

Want to get better insights on your own data? Schedule a consultation now.

SALES CONTACT:	OUTREACH@MEZMO.COM
SUPPORT CONTACT:	SUPPORT@MEZMO.COM
MEDIA INQUIRIES:	PRESS@MEZMO.COM



mezmo