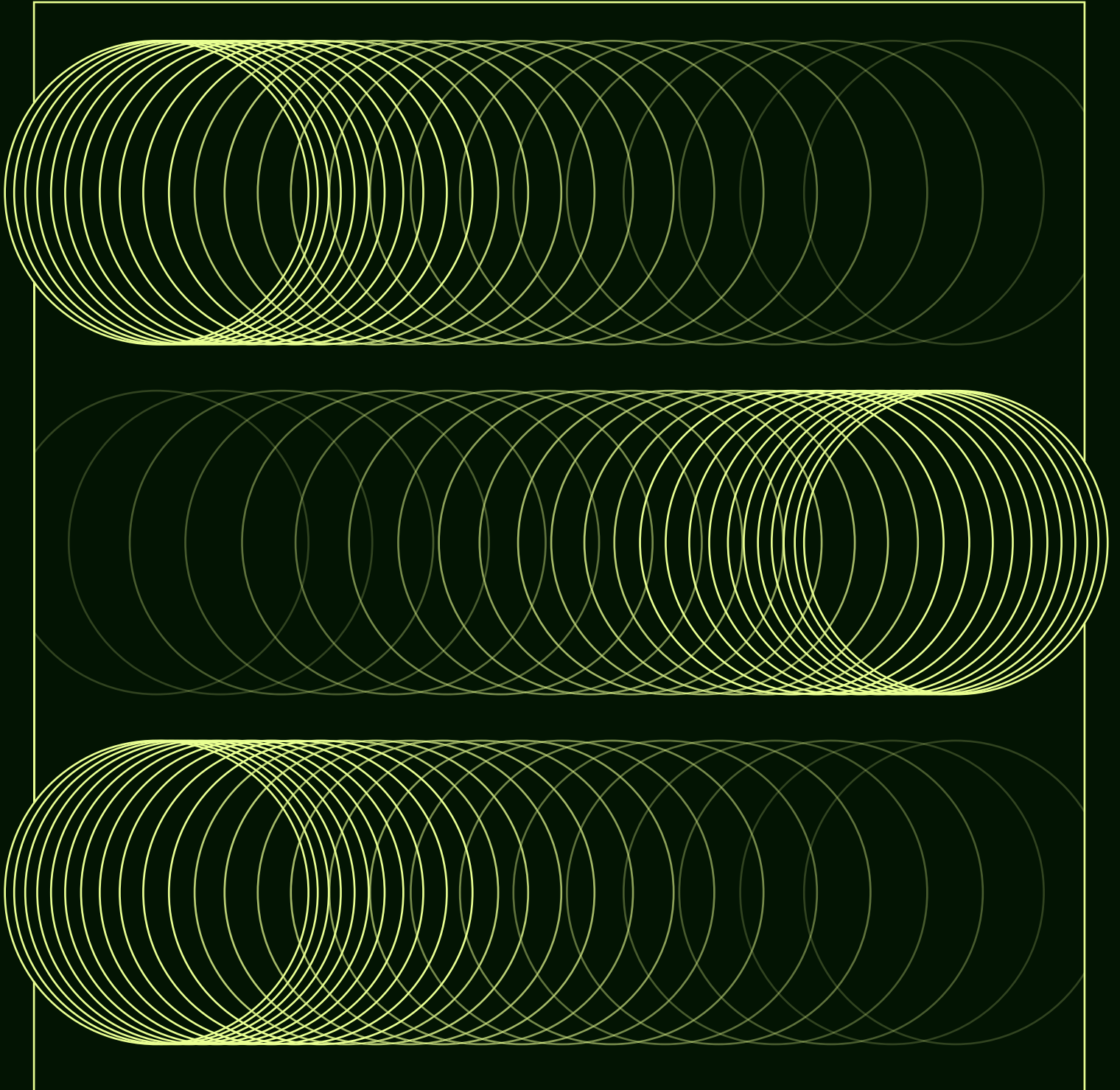# OBSERVABILITY PIPELINE PRIMER

INTRODUCTION
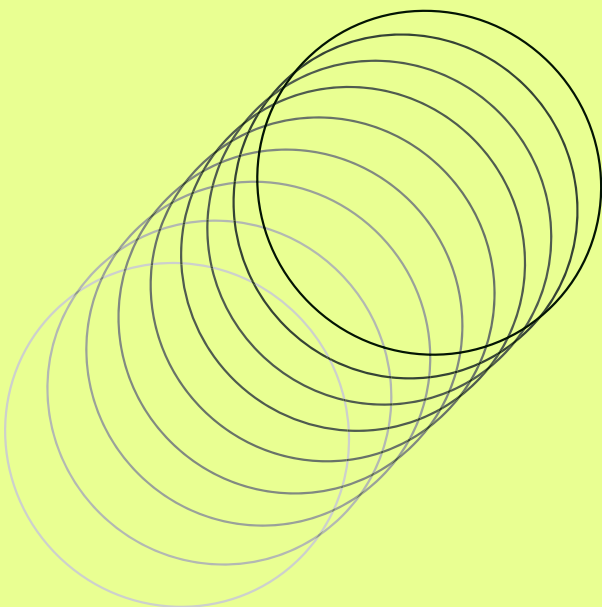
DATA HAS BEEN REFERRED TO AS "THE OIL OF THE TWENTY-FIRST CENTURY." BUSINESSES DEPEND ON DATA TO POWER VIRTUALLY EVERY ASPECT OF THEIR OPERATIONS – FROM SOFTWARE DEVELOPMENT AND MANAGEMENT, TO SALES AND MARKETING, STRATEGIC PLANNING, AND BEYOND.

An isolated oil well in the desert would not have delivered much value in the twentieth century if it didn't connect to a network that could process the fuel and provide it to consumers. Modern business data, like that oil well, is of little use if it is difficult to ingest, transform and interpret. Data must flow seamlessly across the organization to drive business success. It's only by making data easy to ingest, transform, correlate and route to relevant destinations that businesses can leverage data to maximum effect and with maximum efficiency.

This is why forward-thinking businesses are building observability pipelines. Like 20th and 21st century companies rely on oil pipelines to quickly move fuel from production sites to consumers, today's digital businesses increasingly rely on observability pipelines to guarantee the fast and efficient movement of information of all types to drive agility, efficiency, security, and compliance.

This white paper explains how observability pipelines work, as well as how they deliver business agility and a competitive advantage by providing functionality that individual tools lack. In addition, it discusses what to look for when building or purchasing an observability pipeline solution. It also provides examples of how businesses have successfully leveraged observability pipelines to derive maximum value from data.

WHAT IS AN OBSERVABILITY PIPELINE?

An observability pipeline is a tool or process that centralizes observability data (logs, metrics, and traces) from multiple sources, enriches it, and sends it to various destinations. Observability pipelines are a centralized means of interacting with data from across the business and ensuring that it is made available to consumers in whichever format they need to work with the information efficiently to drive crucial decisions.

Observability pipelines are essential because conventional observability tools alone don't guarantee that data can be ingested, processed, analyzed, or consumed efficiently. Observability tools function as the end destination for observability data. They may also be able to perform some data transformations and correlations. But because each observability tool works in a unique way, teams often struggle to leverage each one of them effectively. The fact that observability tools typically operate as siloes, with little or no ability to integrate data with other observability platforms, exacerbates the challenge

of relying on observability tools alone to drive insights. So does the high cost of funneling data to and storing it in observability tools.

Observability pipelines solve these challenges by making it much easier to move data from multiple sources across the organization, while also performing analytics on the data while it is in motion. To return to the oil analogy, observability pipelines provide an efficient means of moving, processing, and delivering data that might otherwise sit isolated in one hard-to-access location – just as oil pipelines streamline the process of moving oil from the remote oil wells where it originates into networks of refineries and consumers.

In addition to helping data to flow seamlessly across the organization, observability pipelines help to reduce costs, even as data constantly grows in volume. Storing data is expensive, and the ability to analyze it in motion using an observability pipeline reduces the amount of information that businesses need to store persistently.

HOW DO OBSERVABILITY PIPELINES WORK?

OBSERVABILITY PIPELINES CONSIST OF FOUR KEY COMPONENTS:

INGESTION:

Ingestion is the process that collects data from various sources and aggregates it into a central pipeline. Ingestion processes also define which types of data to collect, how often to collect it, and how to secure it once it is collected.

TRANSFORMATION:

Transformation enriches data to make it more actionable. For example, data might be filtered or deduplicated to remove redundant information. Or, you could sample it to reduce the volume of data within the pipeline without compromising teams' abilities to analyze the data.

CORRELATION:

Correlation relies on intelligent algorithms to bring various data types together and create actionable signals based on them. An observability pipeline allows correlation to occur while data is in motion so that decisions can be made on the fly and in real time – as opposed to comparing data sets manually or relying on batch processing to derive insights. so that decisions can be made on the fly and in real time – as opposed to comparing data sets manually or relying on batch processing to derive insights.

ROUTING:

Routing ensures that data is continuously streamed to the various stakeholders who need to consume the data. Because each consumer may require a different type of data format to support varying use cases, the routing process also ensures that data arrives in a way that makes it readily usable.

Again, observability tools alone don't provide all the functionality described above. They mainly serve as endpoints for the destination of data. And while it may be possible to integrate observability solutions manually with other tools for transforming, correlating, and routing data, the purpose of an observability pipeline is to provide a comprehensive, easy-to-deploy platform that delivers all of the functionality needed to work with data out-of-the-box.

THE BUSINESS VALUE OF OBSERVABILITY PIPELINES

From a business perspective, the value of observability pipelines is simple. By ensuring that data of all types can be collected, processed, and routed efficiently and effectively, observability pipelines help businesses derive maximum value from the available data. In turn, observability pipelines deliver advantages that include:

FASTER DECISIONS:

With an observability pipeline, stakeholders from across the business can gain insights from data as quickly as the data is produced. Instead of reviewing data periodically or waiting on the next manual report or audit to make decisions, they can identify trends, risks, and opportunities in real time. In this way, observability pipelines help reduce the Mean Time to Respond (MTTR) to technical problems, for example, and enable more rapid threat detection in cybersecurity.

IMPROVED EFFICIENCY:

By reducing the time and effort required to work with data, observability pipelines allow businesses to redirect staff resources toward other, more productive work. For instance, rather than manually parsing application and infrastructure logs to understand the root cause of a performance issue, IT engineers can use an observability pipeline to correlate the various logs automatically, freeing the engineers to focus on optimizing application performance instead of sorting through data by hand.

MAXIMUM CONTEXT:

Because observability pipelines combine data from many sources into a centralized location where it can be transformed and interpreted collectively, they maximize the context available about data. They make it possible, for example, to compare technical data about application performance alongside business data related to customer satisfaction, which makes it easier to understand how software performance impacts the customer experience.

COLLABORATION ACROSS THE BUSINESS:

Along similar lines, observability pipelines help technical and non-technical stakeholders to work collectively, breaking down the silos that have traditionally separated development and IT teams from the rest of the business.
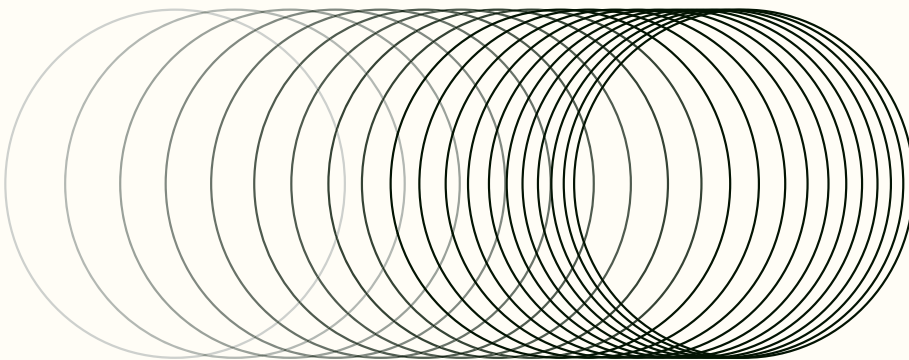
CENTRALIZED CONTROL:

Observability pipelines provide fine-grained controls that allow businesses to define how they collect, store, manage and secure information. Implementing these policies across all data from a central location saves time and effort while reducing the risk of oversights.

COST SAVINGS:

Observability pipelines can help businesses avoid wasting money on data they generate but don't use. Instead, they empower businesses to route compliance data to low-cost storage, while ensuring that critical data types still end up in their chosen tools.They also optimize the efficiency of data analytics and storage through processes like filtering and deduplication.

In each of these ways, observability pipelines help businesses to gain a competitive advantage over organizations that lack an efficient means of processing and analyzing data in real time. Companies equipped with observability pipelines can make decisions faster, more accurately, and at a lower total cost than those that rely on manual processes or ad hoc integrations to move and process data.

Observability pipelines deliver key business benefits today, but their value is poised only to grow over the coming years as businesses generate more data and the importance of data analytics to decision-making increases. The sooner your organization implements a data pipeline, the faster it can get a handle on the vast volumes of data it needs to transform into value and prepare for a future that promises to be even more data-centric than the present.

OBSERVABILITY PIPELINES IN PRACTICE: TWO EXAMPLES

To illustrate how observability pipelines deliver value for businesses in the real world, consider the following examples of observability pipelines at work.

1.

DELIVERING VISIBILITY INTO DISTRIBUTED APPLICATIONS

An eCommerce business deploys an application with a microservices architecture to power its retail website. Each microservice in the application generates its own observability data. Detecting and troubleshooting performance issues requires the ability to aggregate and correlate data from across the entire application as well as the infrastructure that hosts it. Analyzing all of the data centrally is quite costly due to the volumes being generated. Addressing issues in near real time requires the analytics process to begin on data as closely as possible to the time it was generated.

An observability pipeline addresses this need by centralizing data from multiple microservices as it's generated. The pipeline must include the ability to detect and alert on anomalies in real time. The ability to perform cross-domain correlations is also critical because it allows operations teams to drill into root causes across the distributed system without the need for manual reconciliation.

By leveraging an observability pipeline infrastructure, the eCommerce business is able to handle increasingly larger data volumes while remaining agile. This ensures faster troubleshooting and less revenue loss from potential technical issues.
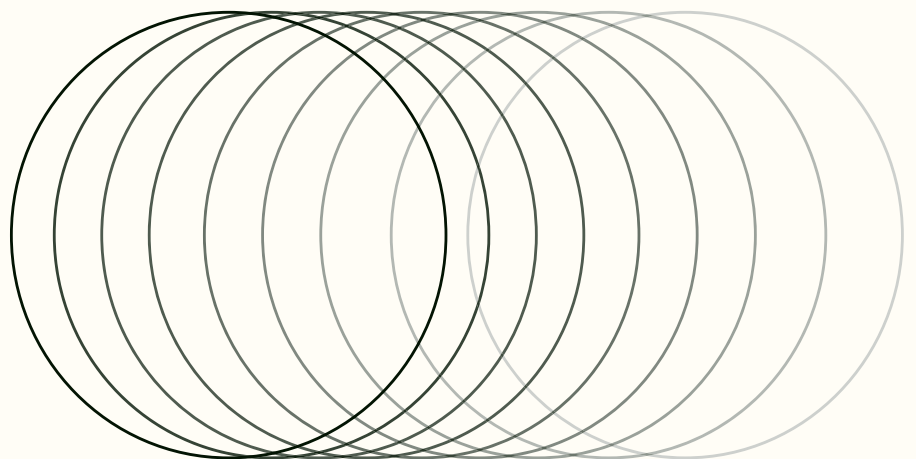
2.

MAXIMIZING INSIGHTS INTO SECURITY THREATS

A fintech company has a Security Incident and Event Management (SIEM) platform in place to analyze data for security anomalies and generate alerts. However, increasing storage costs and practical limitations with the sheer number of systems in the organization are making it more challenging to do everything in the SIEM.

Implementing an observability pipeline reduces the burden on the SIEM by moving parts of the security apparatus upstream before the data is stored. Certain alerts are moved into the pipeline along with data enrichment while it is in motion, optimizing and reducing the total volume sent to the SIEM. This allows the security team to focus on threat hunting and correlation in the SIEM in a more cost effective manner but still retain important alerts.

The result is the ability to detect and respond to threats with greater accuracy and effectiveness.

OBSERVABILITY PIPELINE BEST PRACTICES

It's critical to ensure that your observability pipeline provides the following features and best practices to maximize the benefit that it delivers to your business:

FLEXIBLE DATA INGESTION:

The best observability pipelines can ingest virtually any type of data in any format. Open platforms that support all common data types deliver the flexibility to support not only the data you need to analyze today, but also new types of data your business may encounter in the future.

FINE-GRAINED CONTROLS:

Your observability pipeline should allow you to control how you manage data at each stage of the pipeline process in a highly granular fashion. From ingestion and transformation through correlation and routing, you should be able to define how each data asset or unit is processed, stored, and secured. Doing so means that you can control costs and minimize risks.

IMPROVED DATA EFFICIENCY:

Observability pipelines should enable organizations to optimize their data spend with higher data efficiency. Pipelines should provide the ability to choose what to store where while streamlining the data contents to reduce total volume where possible.

AUTOMATED CORRELATION:

Making the most of data analytics requires the ability to correlate disparate data sources in order to gain the broadest context possible about them. Observability pipelines should therefore make it easy to link multiple data sources to drive contextual understanding. And they must be able to do this at scale, in order to keep pace with growing data volumes – ideally via a built-in data store that is capable of housing data alongside the pipeline.

RAPID TIME-TO-VALUE:

Observability pipelines should make it fast and easy to turn any type of data into value. They do this through simple deployment processes, support for a wide variety of use cases and workflows and the ability to search through, alert upon and generate visualizations for data while it is moving through the pipeline.

By choosing an observability pipeline solution that provides these advantages, your business will benefit from a highly efficient and cost-effective approach to handling increasing volumes of data while retaining the flexibility to evolve with changing data analytics needs.
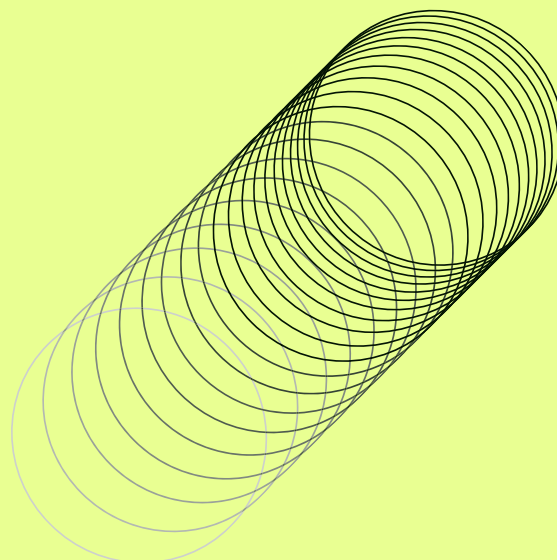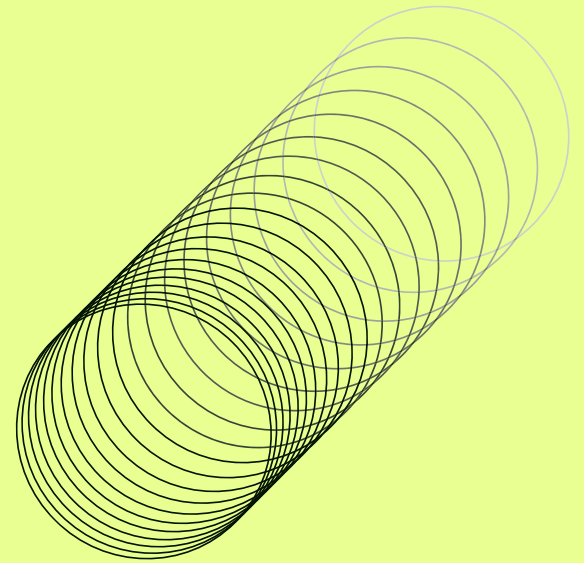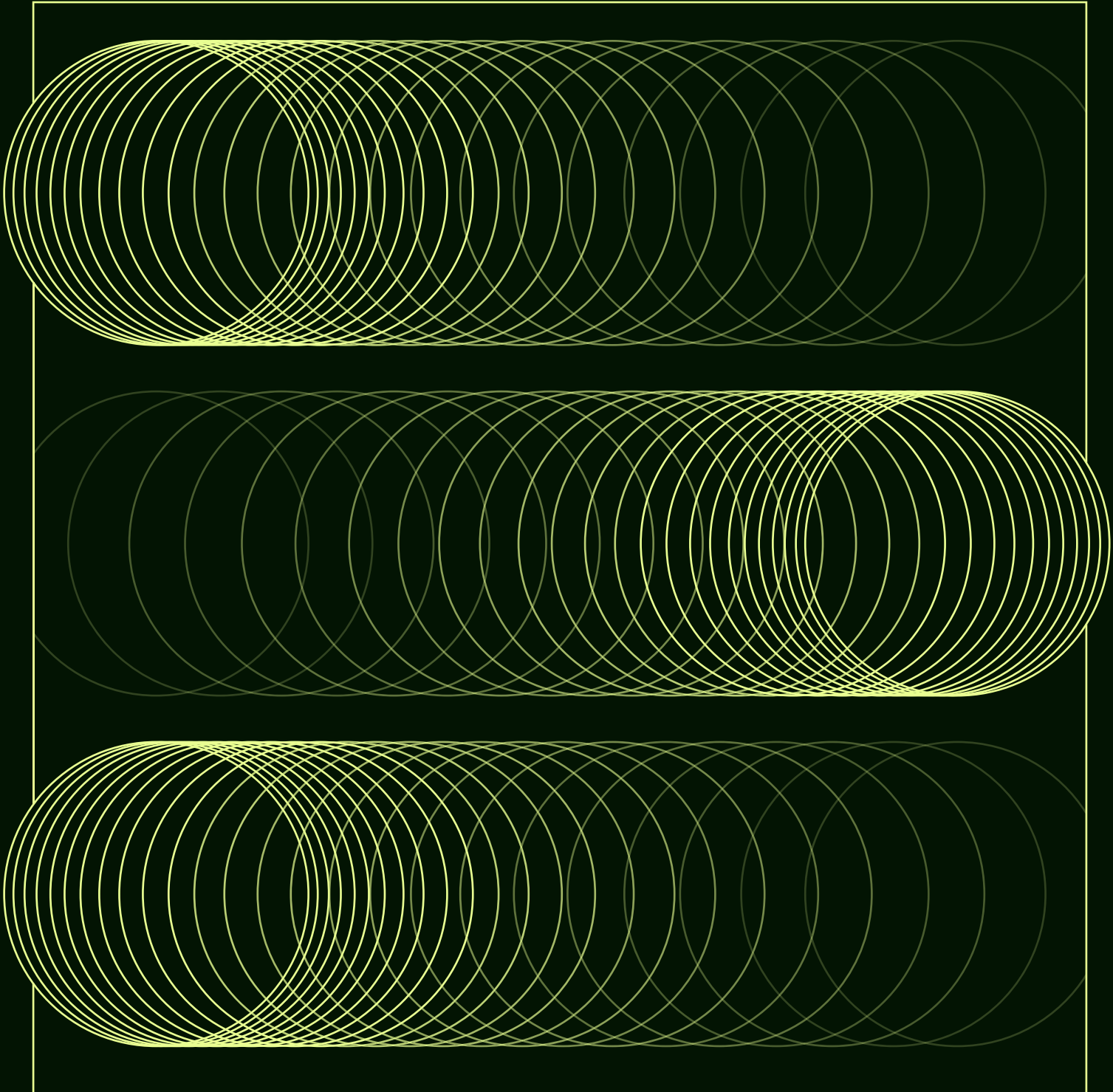
CONCLUSION

The number, volume, and complexity of the data sources businesses rely on to make accurate decisions have increased significantly over the past several years, and they will continue to do so for the foreseeable future.

Observability pipelines play a central role in helping businesses to put data to use, no matter how much data they have to work with or how diverse the data is in terms of type and format. By streamlining the processes of ingesting, transforming, correlating, and routing data, observability pipelines ensure that actionable data is always available to the decision-makers who need it – no matter where they are in the business.

Mezmo was designed from the ground up to provide a holistic observability pipeline that can accommodate any data. By making it possible to collect data from all common sources, then search, alert on and visualize it in motion using analysis features that are native to the pipeline, Mezmo automatically drives real-time intelligence that can support any business unit or workflow. Learn more by requesting a demo.

mezmo