| WHITE PAPER |

Accelerate the Preparation and Data Requirements for AIOps

Is Your Business Prepared To Benefit From AI From a Data Perspective?

Author: Jason Chaffee, AVP, Global Strategic Alliances.

TABLE OF CONTENTS

Introduction

Ensuring your organization is AI-ready requires complete immersion into your AIOps and observability data collection practices to revisit priorities and assumptions.

Data management for AIOps must move beyond the cleansing, de-duping, and control of discrete, additive data sets into the scalable, AIOps use case-directed curation of Conversational Wire Data filtered specifically for infrastructure platforms.

Consider complementary data alternatives that provide clarity and context across components and systems so that AI models can more readily recognize anomalies across standard behavioral patterns.

What is AIOPs?

AIOps is a practice that combines human and technological applications of AI/ML, advanced analytics, and operational practices to business and operations data.

AIOps enhances human judgment, provides proactive alerts on known scenarios, predicts likely events, recommends corrective actions, and enables automation.

It is fueled by coalescing and transforming sensory data into AI-enriched actionable information. A retrospective analysis and governance structure fueled foundational improvements and trust.*

* Forrester AIOps Reference Architecture

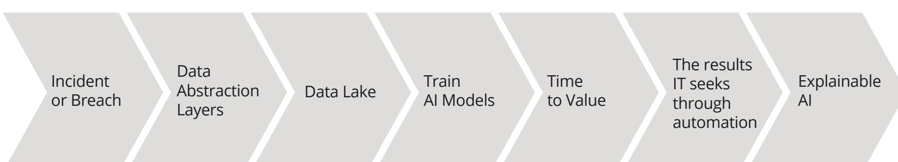
All data are steps removed from what they are intended to measure. Humans have generally accepted this fact in science and information technology (IT) domains. But must we accept it in our AIOps (AI for IT Operations) data strategies?

Let's Take a Closer Look

Measuring and observing occurrences within systems or applications in the IT world takes time. And with time comes layers of additive machine data sets like events, metrics, traces, and logs. Events, for example, may alert technologists that something occurred in a system so that logs can help find what is causing the problem. The data stacking game begins. While necessary, these data sets represent data abstraction layers in a broad sense, perhaps blurring important information needed to train AI. In the age of widespread AI adoption, this issue of data abstraction (e.g., adaptation steps removed from the data source with time and additive machine data) has become magnified in importance and relevance. The reason is that those layers of time and additive tech data from an incident or performance degradation or breach in the IT and cyber parlance can affect AI model output trust, as we must train AI models over prolonged periods. AI models are trained on the data we have today. So now our commitment to data and data abstraction is much longer and pronounced as AI monetary investments soar in the near term. In short, AI for IT Operations (more on this later) has a data problem. When training AI models, at least one data type in your data storage solution, like a data lake, must represent a north star to show interactions across components independently.

Data and AI are inextricably linked. One is as strong as the other. However, it is not just data or generalized 'data quality' but issues such as when it was collected. How many layers of change and abstraction occurred which could lead to data distortion? Can we trust AI model outputs when trained on multiple layers of abstracted, foundational observability data alone (i.e., metrics, logs, and traces)? Not only does it take time to collect meaningful data, but in the IT world, it is often necessary to combine data types to understand how individual components and systems interact and depend upon one another during run time and production environments. However, what AI models can thrive upon is data that has inherently captured the dependencies and interactions across components. Data 'closer to the source' requires fewer layers of other data sets to interpret the results. Is there another complementary data set for AI in IT Operations today?

The path to results with AI/ML in IT Operations:



The stakes have never been higher for large, complex organizations considering Artificial Intelligence (AI) as a certainty for solving problems and gaining competitive advantage across all sectors. At the core of this high-stakes doubling down on AI adoption are the data companies are already using today to steer IT investments, digital transformations, cloud-native application development, cybersecurity posturing, etc. Over time, AI can improve the data, but mainstream adoption of AI today must include current data to feed data lakes, train AI models, etc. Like all forensic problem-solving endeavors, it is vital to have data 'as close' to the occurrence after a breach, performance degradation, or partial outage. Seeing the issues sooner than later is advantageous for incident resolution in the broadest sense. But what if the lion's share of data feeding AI today were produced through multiple layers of data abstraction far removed from the actual 'incident'? Would it be beneficial to complement current data with another machine data alternative that can provide clarity and context across components and systems so that AI models can more readily recognize anomalies across standard behavioral patterns?

Data in Deeper Focus

In an increasingly complex, data-saturated, and digital business landscape, IT leaders acknowledge the advantages of Artificial Intelligence. For example, Chief Information Officers (CIOs) and Chief Information Security Officers (CISOs) can leverage AI models to inform and justify decisions regarding the successful deployment of IT resources and guide digital transformations. These leaders explain to CEOs and executive boards that AI models provide data-driven insights, enabling more accurate forecasting and risk assessment to drive innovation. By analyzing vast datasets, AI can uncover patterns and events humans can easily overlook (e.g., alert fatigue), offering a strategic advantage in decision-making. For instance, AI can predict the potential return on investment (ROI) of technology deployments, assess the cybersecurity risks associated with different IT configurations, or identify efficiency improvements in digital workflows. However, the growing vastness of data is evident as IT leaders seek to connect the dots between data sets in meaningful ways. Questions like the health of the interactions between individual IT components become paramount, as this may be an area of lack with the overproduction of cloud-native data sources like traces and logs (more on this later). This evidence-based approach helps IT leaders defend their decisions, showcasing the value of IT investments in performance and security. However, data quality is often a hindrance.

Data abstraction in the context of AIOps refers to the process of simplifying complex data from various sources, such as metrics, logs, and traces as they traverse through data pipelines, to extract meaningful insights. Unfortunately, this 'abstraction' includes hiding the underlying complexities of raw data across disparate types and presenting a potentially distorted representation for analysis.

Furthermore, IT and data leaders can emphasize how AI-driven decision-making enhances agility and innovation. AI models can rapidly process and analyze added information, allowing businesses to adapt quickly to changing market conditions, emerging threats, and opportunities. This agility is critical for maintaining competitive advantage and ensuring digital transformation (DX) initiatives deliver sustained value. By presenting these dynamic capabilities to the board, CIOs and CISOs (and staff) can illustrate how AI tools support strategic planning and contribute to building a resilient and forward-thinking digital infrastructure. However, full AI readiness will require more fact-finding and strategizing on the data front.

The benefits and rewards can be very innovative. But what risks and hurdles prevent your organization from being ready for AI? Data and time to value! Even the savviest, data-driven large enterprises need holistic and strategic frameworks to manage data fully in preparation for AI. This paper will simplify the landscape by focusing on data that fuels AI in IT Operations.

There are some fundamental issues to prioritize when reviewing data sources for AIOps. For example, IT organizations that utilize data curated and filtered closer to the source of production can realize benefits compared to more static IT observability data like metrics, logs, and traces. Data from the source is often more accurate and timelier, leading to faster, more informed decision-making (AI-model and human-based). Data with fewer levels of abstraction also tend to retain more context, which can improve the predictive capabilities of AI models, allowing for proactive rather than reactive management of IT systems. (Call out in visual design in the paper: data abstraction = AIOps data sources such as metrics, logs, and traces, which layer insights as they pass through data pipelines or OpenTelemetry feeds) Furthermore, data with more layers of abstraction characterized by most observability pipelines today can often distort more nuanced insights into security threats, preventing CISOs from developing robust defenses. Overall, higher quality data with fewer layers of abstraction leading to potential distortion of AI outputs can lead to increased operational efficiency, enhanced security postures, and a better understanding of system performance and user interactions in support of 'user experience.'

For this paper, traditional, cloud-native "observability pipelines" refer to those pipelines primarily focused on foundational monitoring (metrics, logs, and traces) data types. There has been an ongoing overproduction of data from these observability pipelines, leading to data ingest issues for AIOps.

In a landscape where data-driven organizations expect continued growth in their volume of data from multiple streams or pipelines, achieving trust and cost control in data collection is no small feat. It necessitates a flawless governance framework that addresses data collection, storage, processing, and dissemination amidst rising costs. This is another critical aspect of preparedness for AIOps. IT leaders must leverage the data they have today. Still, it may be necessary to include complementary data sources with fewer layers of abstraction and the inherent capacity to expose interactions across components and the health of components within systems. Source data is a powerful complement when combined with the power to filter and fine-tune data through curation. In so doing, AI models can 'cut through the noise' to find anomalies faster and more consistently.

Achieving trusted and reliable AI hinges on robust data and AI governance, which are bulwarks against various pitfalls. Are you ready to pivot into 'tomorrow' (diagram below) through AI? How prepared is your organization to prioritize and manage different data types? Cloud-native application developers helped push forward the urgency to implement more "observability" and, in so doing, prioritized logs, metrics, and traces. As trust in the outputs of AI and data quality represent potential roadblocks, should we now shift our attention to complementary curated data sources filtered directly from the source?

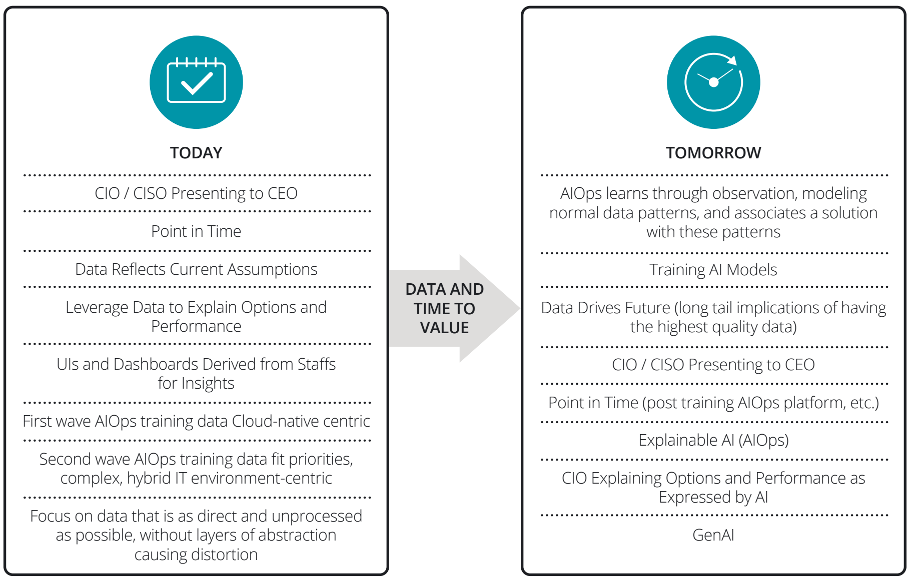


Figure 1: High Quality Data – The Rising Stakes!

As this paper further digs into AI for IT Operations (AIOps) specifically, the volume and complexity of data become even more pronounced and layered. Crystallizing the objective for AI in the automation and enhancement of IT Operations focuses our attention on the core of observability data. For this paper, the definitions of each observability data source will remain at a high level. Each data set adds to the next, filling in the picture (table below). This 'data stacking' to glean further insights is one aspect that defines these sources' 'additive' nature.

Data Sources for AIOps – Machine Data ShortList	Key Questions Addressed – Together Providing a Fuller Picture
Events	Did an issue occur?
Metrics	Do I have a problem based on that issue?
Traces	Where is the problem?
Logs	What might be causing the problem?

Next, let us unpack what has led us here in the IT industry. First, consider the 'data supply chain' in the typical enterprise, whereby observability improvements were the first order of business. Suppose one agrees that the general progression from monitoring to observability to AIOps is a transitional path most companies are pursuing. How do we break up the status quo, as it were, and ensure data quality is being addressed across observability pipelines as they feed data lakes (simple diagram below), which seed AI model training requirements? Could layers of observability pipelines and pipeline management tools for managing observability data contribute to the general 'data abstraction' and distortion that could stymie AI models post-ingestion? Can other complementary data sources help clarify and refine data sets like metrics, logs, and traces? Given that there are no clear frontrunner options on the consumption preferences of AIOps solutions, some companies define AIOps as adding more AI/ML capabilities into their observability pipelines. But is there a way for your business to reassess available data sets today and become AI-ready? A clear path forward is necessary – where AI models can process ingested contextual, operational insights more directly from data sources that expose the interactions and dependencies across components. To become fully AI-ready, your infrastructure platform providers (in support of AIOps) must also ingest data generated from the interactions within IT environments and between users and IT services. This data must be curated from the source to retain the critical interactions and dependencies across components. Data that is less 'additive' with fewer 'layers of abstraction' should be considered to improve observability pipelines' forensic capabilities.

The following image shows how 'observability pipelines,' which focus on metrics, logs, and traces as the pillars of data collection, have facilitated the overproduction of data. In turn, companies must deploy pipeline management tools to help streamline or control overproduction. In so doing, the contextual insights most meaningful to AI could be blurred or removed.

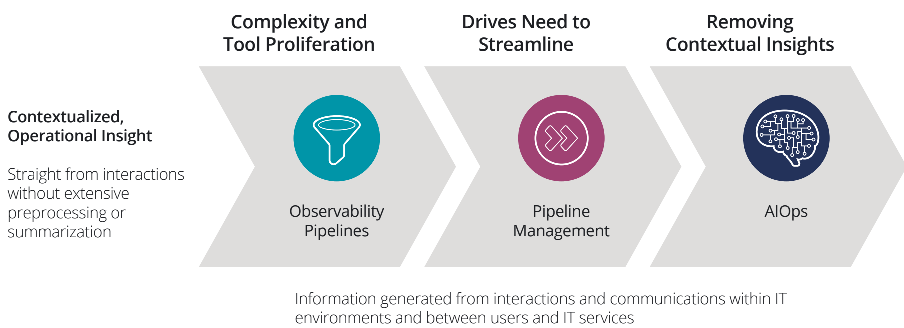


Figure 2: Time, Processing, Filtering and Layers of Abstraction Distorts Insights.

The need for artificial intelligence in infrastructure and operations is pressing. Tool proliferation, data volume growth, and duplicate data sets could 'clog up' data pipelines, obscure critical information to feed AI pattern recognition and compromise trust in AI model outputs.

Next, consider a complementary path forward. As can be seen below, IT, in general, is in a state of flux and directed change. User-focused objectives (rooted in improving the user experience of digital businesses—far right below) are top of mind and even highlighted in AIOps reference architectures from major IT analyst firms.

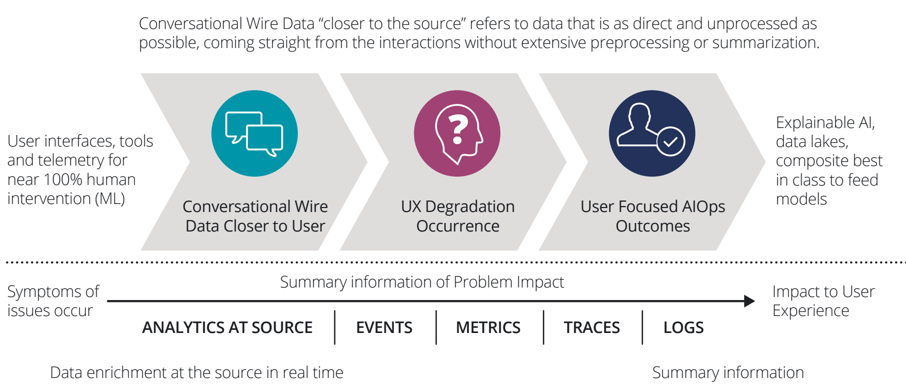


Figure 3: Worldview of Data Options to Feed AI – Partial Snapshot.

Implementing adaptive and curated data analytics at the source, where data originates, can significantly enhance the quality of data streaming into data storage solutions for AIOps (Artificial Intelligence for IT Operations). Being AI-ready includes data management methods holistically and prior planning to include data types with fewer layers of abstraction. Organizations can detect and correct issues faster by adapting and analyzing data at its point of origin, reducing noise, and ensuring that high-quality, filtered data is streamed to the AIOps platform. This approach minimizes latency and potential data degradation during transmission or transformation processes. Superior data quality at the source leads to more accurate and reliable AI-driven insights, as the AIOps system can more effectively correlate, analyze, and act upon cleaner and more consistent data streams. This immediate, source-level data refinement enhances the overall effectiveness of AIOps, enabling more precise monitoring, faster anomaly detection, and more proactive and predictive IT operations.

High-quality data curated directly from the source can also augment and improve the training of AI models. When trained with enriched source-level data, AI systems can better understand the operational landscape, leading to better predictive analytics and decision-making. This high-quality data complements other data sources integral to AIOps, such as events, metrics, traces, and logs, by providing a robust foundation for the AI models. Together, they create a comprehensive data ecosystem that supports nuanced AI analyses. For instance, high-quality source data could help AI models learn the normal operational baseline more accurately, improving the model's ability to detect deviations signaled by events and metrics. Traces and logs, enriched by detailed and accurate source data, allow for more precise fault isolation and root cause analysis. This constructive collaboration between high-quality source data and other data types enables AI models to operate more precisely, leading to more effective and timely responses to IT operational challenges.

So, what type of data source is most complementary in its roots and capabilities? Given that the data ecosystem is littered with necessary progressions of additive and abstracted data layers and pipelines, is there a less processed alternative that comes straight from interactions without extensive processing and summarization? Let's dig further.

Conversational Wire Data in the context of AIOps (Artificial Intelligence for IT Operations) refers to the information generated from interactions and communications within IT environments and between users and IT services. Conversational Wire Data offers rich, contextual insights into recurring issues, user experiences, and operational challenges, allowing AIOps models to fine-tune anomaly detection and predictive maintenance capabilities by understanding the nuances of real-world problems.

Conversational Wire Data

Collecting Conversational Wire Data closer to the source ensures that the information retains its original context and nuances, which is vital for training more effective AIOps models.

These models can then better understand user intentions, predict potential IT issues before they become critical, and automate responses or escalations as needed.

The goal is to enhance IT operations' efficiency, reduce downtime, and improve user experience by leveraging AI/ML to process and act on large volumes of Conversational Wire Data in real-time on a global basis for each customer environment.

Conversational Wire Data "closer to the source" refers to data that is as direct and unprocessed as possible, coming straight from the interactions without extensive preprocessing or summarization. AIOps and infrastructure platforms realize near-term advantages and there are ongoing game-changers for ingesting Conversational Wire Data patterns curated from the source. Perhaps most importantly, a complete data strategy for AI readiness requires flexibility around data collection volumes. Look for solutions that curate targeted data feeds by AIOps use cases or applications to regain control of AIOps data streaming costs and maintain higher quality data along the way. Highly curated data streaming can reduce data volumes significantly. It is conceivable to reduce tens of billions of records to only millions or much smaller ones. Targeted data feeds can represent a powerful way of fueling the most impactful AIOps use cases with the most appropriate data mappings and improve the value realized from your infrastructure platform and seed optimized AI models. Seek data providers for AIOps that provide complete control of streamed data volumes mapped to domain-specific use cases.

- **Advantages – Control Data Collection Volumes**

  Many companies are now seeking a balance across their observability data collection practices (cost v benefit) to fuel AIOps. The control measures to manage cost efficiencies for data collection will drive a new paradigm. It may become necessary to revisit the consumption of duplicate and additive data sources like metrics, logs, and traces to prioritize curated data enriched and filtered specifically for complex hybrid IT environments (including the cloud). Conversational Wire Data can reach into hybrid environments, including the cloud, to glean additional insights for more cost-effectiveness. As you adjust your AIOps and observability data collection practices, you must curtail the consumption of data sources like metrics, logs, and traces and prioritize more cost-effective, higher-quality sources. "Dial in" the volumes most appropriate for your use cases.

- **Advantages – Streaming Conversational Wire Data 'Today'**

  By analyzing Conversational Wire Data, AIOps can identify patterns and correlations between reported issues and underlying system states, enabling more accurate root cause analyses and proactive remediation strategies.

- **Advantages – Streaming Conversational Wire Data 'Tomorrow'**

  Leverage Conversational Wire Data to continuously improve the accuracy and reliability of AIOps models through filtered data streaming feedback loops. This ensures that operational decisions are aligned with the evolving needs and experiences of end-users and enhances overall IT service delivery. Fine-tune the engine over time with the highest quality data alternative.

Find the Signal

Cloud-native application dev teams require traditional observability data from applications to effectively manage modern cloud environments' complex and dynamic nature. Cloud-native apps are typically distributed, consisting of microservices that operate in containers or serverless architectures, making them inherently more ephemeral than traditional monolithic applications. This distribution across multiple services and environments necessitates visibility into aspects of the application's performance. But from an AI standpoint, can you also detect behaviors and interactions crossing components in the cloud and those not in the cloud in full hybrid IT? As cloud-native applications dynamically scale, adapt, and evolve, developers and operations teams need immediate insights into their workings to meet performance expectations. One caution for becoming AI-ready is ensuring the data function is not driving the AIOps use cases. Data can support business use cases for AIOps but should not be the sole determining factor. To ensure adequate data coverage across users, IT services, resources, and malicious threats, augmenting existing tools with Conversational Wire Data is vital.

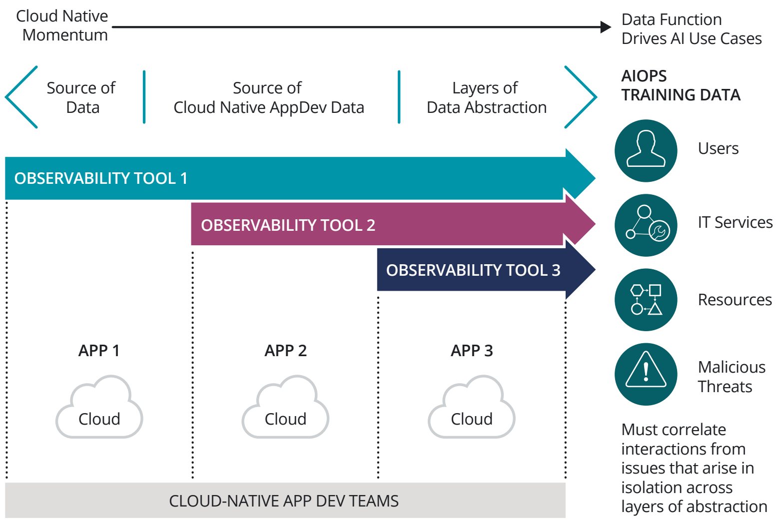


Figure 4: How We Have Gotten Here as an Industry.

One unintended consequence of extensive cloud-native observability is the proliferation of multiple observability data pipelines across different environments, each tailored to collect, process, and analyze combinations of specific data types like logs, metrics, and traces. While these pipelines are crucial for monitoring aspects of the cloud-native ecosystem, they can lead to fragmented views and siloed data streams, complicating the holistic understanding of hybrid environments. The proliferation of these pipelines across various platforms, tools, and teams can obscure insights and hinder the ability to correlate data across different components. Consequently, it becomes challenging to diagnose issues in hybrid IT environments and understand the interdependencies of services.

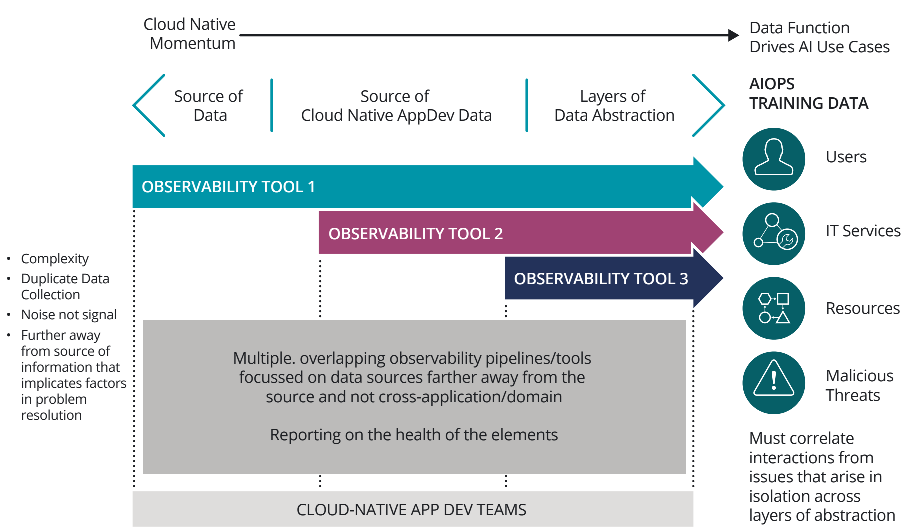


Figure 5: How We Have Gotten Here as an Industry.

Given the massive data complexity and overlapping and duplicate nature of multiple pipelines with too much focus on a narrow definition of telemetry, it is no wonder that AI models struggle to find clarity in the noise. This noise can lead to distrust of AI outputs and jeopardize the potential of AI in general as companies transition into Generative AI (GenAI) to maximize business results. Conversational Wire Data that captures the dependencies and the health of interactions across the board could assist in environments where organizations leverage a mix of on-premises data centers, private clouds, and public cloud services to create a flexible, scalable, and efficient IT infrastructure.

The need to see the health of interactions across components is critical for AI models and operations teams alike:

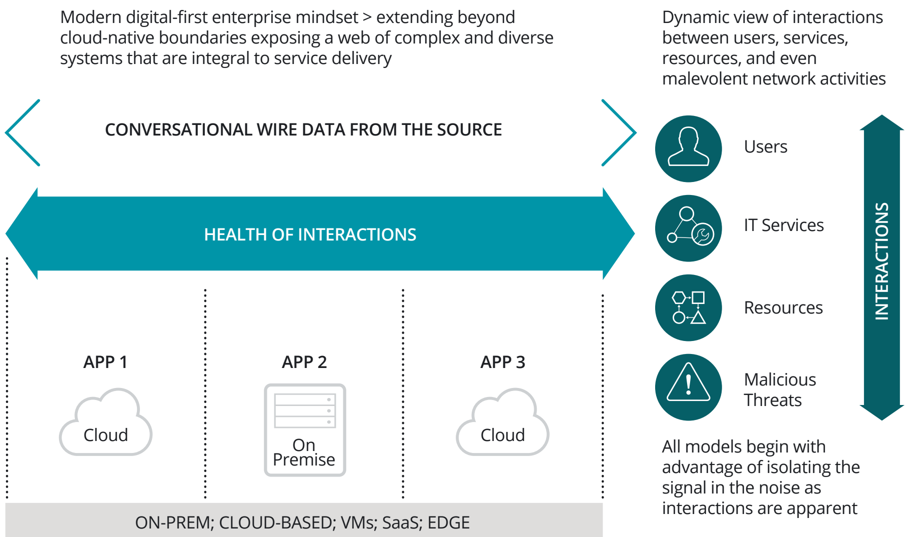


Figure 6: The Missing Piece.

Maintaining a dynamic view of interactions between users, services, resources, and even malevolent network activities is essential when managing overall data ecosystem strategies, especially when deploying AI. Such comprehensive visibility allows organizations to effectively govern their data ecosystems, which are complex and constantly evolving with user behaviors, service updates, resource allocation changes, and security threats. AI systems are adept at analyzing and learning from these interactions, adapting to new patterns, and optimizing operations. By maintaining a dynamic view of the health of components, organizations can ensure that the AI models are trained on up-to-date and relevant data, which is crucial for making accurate predictions and ultimately taking intelligent actions. For instance, dynamically tracking network activities in cybersecurity enables AI systems to differentiate between benign user behaviors and potential security threats, leading to more effective threat detection and response strategies.

Moreover, the necessity of this dynamic approach becomes particularly pronounced in an environment where real-time decision-making is critical. For example, in high-frequency trading platforms, the ability to dynamically analyze and act upon the ever-changing interactions between market conditions and trading activities can lead to significant financial gains or losses. In such scenarios, AI, fueled with Conversational Wire Data, can provide the power and speed to process and analyze vast amounts of data in real time, allowing immediate insights and actions. Additionally, in cloud computing, where resources are scaled and optimized on the fly, a dynamic view enables AI to make instantaneous resource allocation decisions to meet current demand without over-provisioning, thus ensuring cost-efficiency and performance optimization. In the most demanding hybrid IT environments, the dynamic observation of interactions across the data ecosystem is beneficial and critical for the strategic deployment and effective operation of AI-driven systems.

Many organizations can benefit from adding a complementary data source, like Conversational Wire Data, to help find the signal across traditional observability pipeline data sets to reflect the health of components. Another best practice for AI readiness is adopting an AI sensor and streamer AIOps architecture to complement existing infrastructure platforms and AIOps for fine-tuning ingestion.

- **Sensor approach defined:** consider scalable, purpose-built sensors that transform Conversational Wire Data from the source to generate high-quality, contextual data for analytics and curation.
- **Streamer approach defined:** consider a use case-centric data ingestion engine approach that performs adaptive feature extraction on granular Conversational Wire Data to deliver enhanced, curated data feeds to AIOps and data lakes for further AI/ML processing.

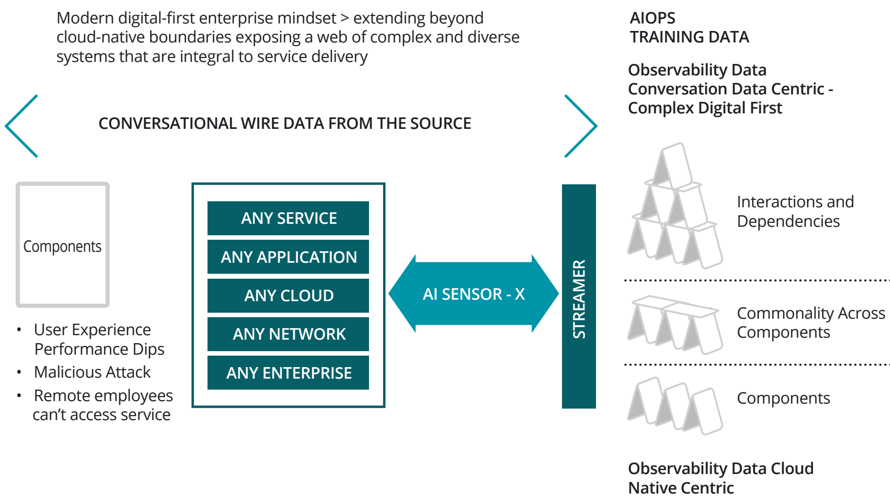


Figure 7: The Missing Piece.

Improve existing infrastructure platforms by aligning this AI architecture to case studies with the highest business value. Selecting the highest impact use cases to drive AIOps results is an obvious preferred practice. Still, data management practices for AIOps should find opportunities to leverage data sources that can articulate the commonalities, interactions, and dependencies across components. The additional control of highly enriched and use case-driven data streaming could allow your organization to minimize redundant observability pipelines across infrastructure platforms.

Collecting Conversational Wire Data transformed closer to the source ensures the information retains its original context and nuances, vital for training more effective AIOps models. These models can better understand user intentions, predict potential IT issues before they become critical, and automate responses or escalations as needed. Curated and enriched data or metadata provides a detailed, high-level overview of network traffic, for example, and system interactions. For infrastructure platform providers, enriched metadata offers insights into the operational state of the network, server, and application performance. This increased visibility helps identify patterns, trends, and anomalies. The goal is to enhance IT operations' efficiency, reduce downtime, and improve user satisfaction by leveraging AI to process and act on large volumes of Conversational Wire Data globally for each customer environment in real-time. The new data imperative driven by AI necessitates low-noise data, which exposes applications and services' connections, performance, and experiences across hybrid IT environments.

NETSCOUT

Corporate Headquarters
NETSCOUT Systems, Inc.
Westford, MA 01886-4105
Phone: +1 978-614-4000
www.netscout.com

Sales Information
Toll Free US: 800-309-4804
(International numbers below)

Product Support
Toll Free US: 888-357-7667
(International numbers below)

NETSCOUT offers sales, support, and services in over 32 countries. Global addresses, and international numbers are listed on the NETSCOUT website at: www.netscout.com/company/contact-us

NSWP_010_EN-2402 09/2024