

The Beginner's Guide to Machine Data in the Agentic Era



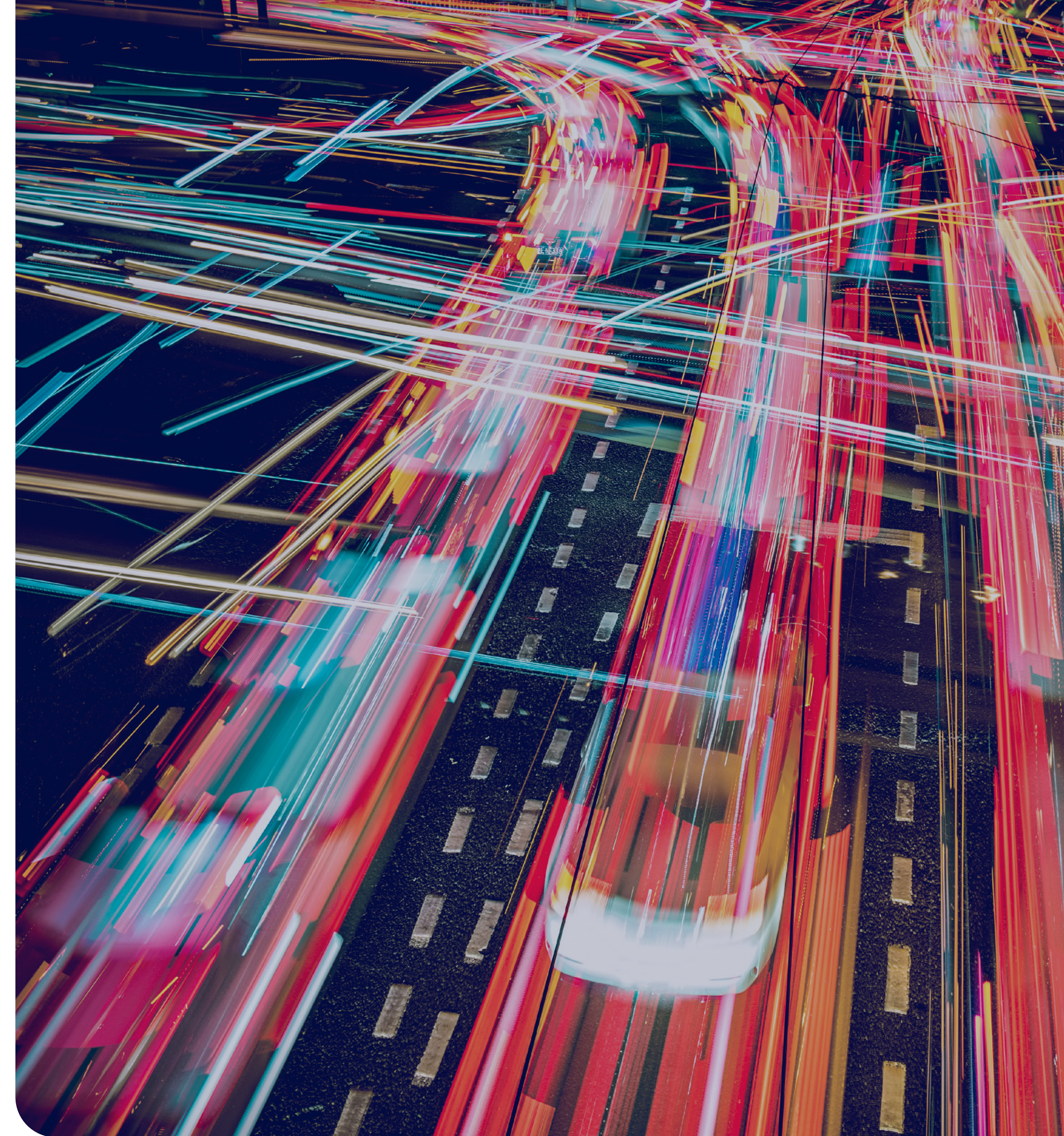
Modern digital systems are complex, distributed, and constantly in motion — and they generate an overwhelming amount of data even at the best of times. Every request and interaction produces a signal, capturing how technology behaves across operations and a sprawling enterprise environment.

This very data (known as machine data) is uniquely challenging because of “the 3 Vs” — that is, volume, velocity, and variety. It’s generated at massive scale, streams continuously, and comes in countless formats across logs, metrics, traces, and events. Unlike traditional enterprise data, which is structured and purpose-built for reporting, machine data is raw, unwieldy, high-frequency, and operational, created to reflect what’s happening *right now*, no matter where it resides across the barrels of the enterprise.

Of course, that’s where the difficulty begins. In most environments, machine data is fragmented across a patchwork of tools and systems, making it hard to capture and use effectively. At the same time, the cost of collecting and storing this data continues to rise as architectures become more distributed and data volumes grow exponentially.

The rise of artificial intelligence has only raised the stakes. AI-powered assistants can accelerate detection and automate investigation and response, yet these capabilities are only as strong as the data behind them. Without reliable, contextual machine data, teams are left operating without a clear picture: automation breaks down, insights become misleading, and decisions lose precision.

Organizations that establish a data strategy paired with a modern architecture can transform machine data into a strategic asset: powering digital resilience and accelerating innovation. And who wouldn’t want that?



What machine data really is (and why it matters)

Machine data refers to the raw, unfiltered output of digital systems. It captures what's happening in real time, straight from the source, without structure or interpretation.

This is what makes machine data both incredibly valuable *and* inherently difficult to work with. It is high in volume, generated at massive scale; high in velocity, streaming continuously; and high in variety, spanning countless formats and types. In other words, it's wild by nature: untamed, constantly changing, and not designed to be easily understood without the right tools and approach.

While machine data appears in many forms, it typically falls into four primary categories:

Logs

Logs record discrete events that occur within systems or applications. These entries provide detailed descriptions of system activity, including errors, configuration changes, authentication attempts, and application transactions. Logs are often the most detailed source of operational insight.

Metrics

Metrics are numerical measurements collected over time. They provide high-level visibility into performance indicators such as CPU usage, request latency, throughput, or error rates. Metrics are particularly useful for tracking trends and identifying anomalies.

Traces

Traces capture the path of a request as it travels through distributed services and microservices. They help teams understand how different components interact and where delays or failures occur within complex application architectures.

Events

Events represent state changes or alerts triggered by systems. These signals often indicate that something significant has occurred, such as a threshold being exceeded or a configuration change taking place.

Collectively, these machine data types form a comprehensive view of how the system is behaving at any given moment. What makes machine data uniquely valuable is the type of insight it provides.

Because of these characteristics, machine data powers highly critical workflows across the organization. Security operations teams rely on it to detect threats and investigate incidents, while IT operations teams use it to troubleshoot outages and maintain system performance. Finally, engineering teams depend on it to improve application reliability and user experience.

The challenge, however, is that machine data volumes can quickly become overwhelming. Without the right architecture, organizations struggle to separate meaningful signals from noise, correlate events across domains, and extract actionable insights from the data they collect.

Machine data is:

- **Time-stamped and precise:** It captures exactly when activity occurs and how systems behave at a given moment, forming a detailed digital footprint across systems.
- **Cross-domain:** It spans applications, infrastructure, networks, security tools, cloud environments, and edge systems.
- **Evidence-based:** It provides causal insight into what happened, where it occurred, and what changed.

The agentic era changed the equation

Modern digital systems aren't just growing, but *evolving*. What began as static, rules-based automation has progressed to generative AI, and now to something far more powerful: agentic systems.

In this new era, AI not only assists, but also acts on behalf of the agent. Agentic AI can do a myriad of things, from analyzing signals, to making decisions, and even executing workflows autonomously across environments. Rather than wait for instructions or strict directives, agentic AI interprets what's happening, and subsequently responds in real time — raising both the opportunity *and* the stakes.

On one hand, agentic AI promises faster detection, smarter remediation, and reduced manual effort. But on the other, it introduces a whole new level of dependency on data — that is, machine data. For agents to operate effectively, they need a constant stream of accurate, high-quality, and contextual information about what's happening across systems. Unsurprisingly, that's where the challenge compounds.

As mentioned before, modern environments already generate machine data at extreme volume, velocity, and variety. Now, organizations must not only collect and analyze that data, they must also prepare it to train, inform, and guide autonomous systems. This means more data to manage, higher expectations for data quality, and greater consequences when something goes wrong.

In fragmented environments, where data is siloed or incomplete, agentic systems lack the context needed to make reliable decisions. The result is a familiar problem: none other than *garbage in, garbage out*. Poor data leads to flawed reasoning, unreliable automation, and outcomes that can scale mistakes just as quickly as the data proliferates.

The rise of autonomous action also introduces new requirements around governance and trust. Organizations must ensure that the data feeding these systems is secure and compliant; and that the actions agents take can be understood and qualified.

When incidents occur in this context, the impact is immediate and seemingly far-flung. Service disruptions affect customers in real time, security breaches expose sensitive data, and compliance failures carry significant risk. With agentic systems in the loop, both the speed of response (and the potential blast radius) increases.

This is why modern operations require more than additional tools or dashboards. They require shared operational context: that is, a unified, real-time understanding of system behavior that connects applications, infrastructure, networks, and devices.

Only with this foundation can organizations ensure that both humans and AI agents are operating from the same, complete picture, allowing for better decisions, and outcomes they can trust.

Today's digital environments must support:

- Always-on services with strict uptime expectations
- Rapid application releases and continuous deployment
- Increasing cybersecurity threats and regulatory requirements
- Seamless digital experiences for customers and employees

What good looks like: an AI-ready machine data foundation

To fully activate machine data, organizations need a foundation designed for both human and AI-driven operations. An AI-ready operational data architecture should support several key capabilities:

Unified visibility across domains

Ideally, operational insight should not be restricted to individual tools or teams. Organizations need the ability to analyze signals across applications, infrastructure, networks, and security environments within a shared context. When teams can connect insights across domains, they can investigate issues more quickly and understand the full scope of system behavior.

Data in place when possible

Traditional architectures typically require all data to be centralized before it can be analyzed. In large environments, this approach can be expensive and operationally complex. Modern architectures support federated analysis, allowing organizations to query and analyze data where it resides. This approach reduces unnecessary data movement while still enabling unified insight.

Governance and trust

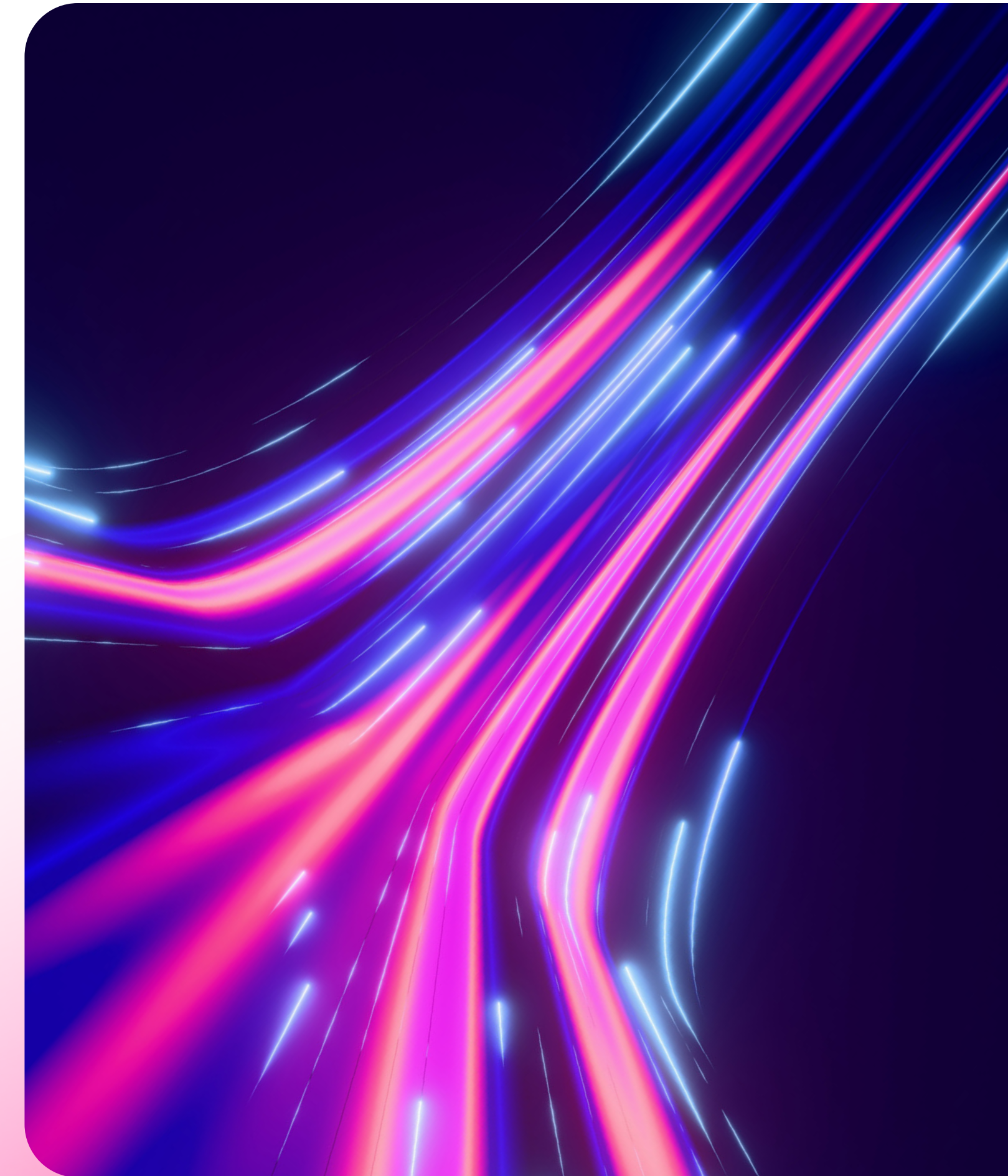
As operational data is the unequivocal foundation for AI systems and compliance workflows, governance becomes critical. Organizations need visibility into where data originates, how it is transformed, and who can access it. Capabilities such as data cataloging, lineage tracking, role-based access controls, and audit trails help ensure that operational insights remain trustworthy and compliant.

Context for AI

AI systems require more than raw data. They need context that explains relationships between systems, dependencies between services, and historical patterns in system behavior. Providing this context allows AI models to generate more accurate insights and recommendations.

Open and extensible architecture

Operational ecosystems evolve constantly. New tools, platforms, and services are introduced regularly. An effective data strategy must remain open and extensible, integrating broadly across technologies while avoiding vendor lock-in, allowing organizations to build a contextualized, AI-ready data foundation that supports unified intelligence across operational domains.



From insight to action

With the right data platform in place and the right steps, organizations can move beyond reactive monitoring toward a more strategic operational model. Historically, teams have really only focused on responding to incidents *after* they occur — largely thanks to the unwieldy amount of data they have to manage day-in and day-out. Add alert fatigue to the mix, and it becomes increasingly difficult to see what actually matters as and when an event occurs. And so, the cycle continues: an alert fires, someone digs through logs and metrics, and the fix is handled manually.

While this model can maintain system availability, it often results in slow response times and fragmented investigations. An AI-ready operational architecture enables a shift toward predictive operations. Instead of waiting for incidents to escalate, teams can identify anomalies earlier, correlate signals across domains more quickly, and resolve issues with greater confidence.

AI plays an important role in helping with this transition. AI-powered assistants can help analysts interpret complex telemetry, summarize investigation results, and recommend remediation steps. Autonomous agents can automate routine operational tasks such as incident triage or infrastructure adjustments.

However, responsible AI adoption requires transparency and governance. Organizations must ensure that automated decisions remain explainable and auditable. Human oversight should remain part of the operational loop, particularly for critical decisions.

Operational benefits may include reduced data duplication, faster mean time to resolution (MTTR), and improved readiness for compliance reporting and audits.



Where to get started

Building an AI-ready machine data foundation doesn't require a complete transformation overnight. In fact, the best place to start is by assessing what you already have in place.

Organizations can focus on a few critical questions that determine whether their data foundation is ready to support modern, AI-driven operations:

Do you have unified visibility?

Operational insight shouldn't be confined to individual tools or teams. To understand what's truly happening across your environment, you need the ability to analyze signals across applications, infrastructure, networks, and security (within a shared context, of course).

Without this unified view, investigations are slower, blind spots inevitably persist, and teams are forced to piece together fragmented data manually.

Can you search and analyze data where it resides?

Traditional approaches often require centralizing all data before it can be analyzed — a process that is regrettably costly, slow, and difficult to scale.

Modern architectures take a different approach, enabling federated search so teams can query and analyze data in place. This reduces unnecessary data movement while still providing the visibility needed to investigate and respond effectively.

Is your architecture open and extensible?

Operational environments are constantly evolving. New tools and services are introduced all the time, and your data strategy needs to keep up.

An open, extensible architecture ensures you can integrate across technologies, avoid vendor lock-in, and adapt as your environment changes. It also creates the flexibility needed to support both human workflows and AI-driven operations over time.



Benefits of having **AI-ready** machine data

When machine data is properly captured and prepared, it becomes far more than a byproduct of system activity but a strategic advantage. Organizations that invest in AI-ready machine data unlock faster insights, and better outcomes across the business. Including (but certainly not limited to):

Simplify operations and costs

AI-ready machine data brings clarity. By unifying data across fragmented systems, teams gain a single, consistent view of their environment, which eliminates the need to swivel between tools and manually piece together context.

This approach improves operational efficiency, and reduces the time required to detect and resolve issues. It also helps organizations optimize how data is collected, stored, and used, minimizing unnecessary data sprawl and controlling costs associated with data management.

Accelerate innovation

When teams have access to high-quality, contextual machine data, they can move faster and more confidently. Developers can release new features more quickly, knowing that they have the visibility to monitor performance and catch issues before it's too late.

AI systems and agents can also operate more effectively, using reliable data to generate insights, automate workflows, and support decision-making in real time. This creates a foundation for continuous improvement, where organizations can iterate and innovate without sacrificing stability.

Strengthen security posture

Security depends on visibility, and AI-ready machine data provides the depth and context needed to detect and respond to threats. By correlating signals across systems, organizations can identify anomalies, uncover hidden patterns, and investigate incidents more efficiently.

With AI in the loop, security teams can move from reactive to proactive: using machine data to anticipate risks, automate response, and reduce dwell time. A strong data foundation supports governance and compliance, ensuring that security efforts are both effective and accountable.

Resilience and innovation now share **the same foundation**

In modern digital environments, operational resilience and technological innovation are no longer separate priorities. Both depend on the same fundamental capability: trusted operational data that can be activated across the organization. By leveraging an AI-ready operational data architecture like Cisco Data Fabric powered by Splunk, organizations can transform machine data into a strategic asset.

With unified operational intelligence, teams can move beyond reactive monitoring toward predictive operations that improve reliability, security, and efficiency.

Organizations that invest in this foundation today will be better positioned to navigate the growing complexity of digital systems — and to thusly unlock the full potential of AI-driven operations.

Lead the data revolution

Stay two steps ahead. Give your teams the total visibility they need to act with confidence. Together with the Splunk Platform — the industry-leading data engine you need to unify security and observability — the Cisco Data Fabric architecture gives you the scale and flexibility to supercharge data management at AI speed.

Ready to future-proof your business and lead in the agentic AI era? [Explore the Splunk Platform](#).



Splunk, Splunk>, Data-to-Everything, and Turn Data Into Doing are trademarks or registered trademarks of Splunk Inc. in the United States and other countries. All other brand names, product names, or trademarks belong to their respective owners. © 2026 Splunk LLC. All rights reserved.

