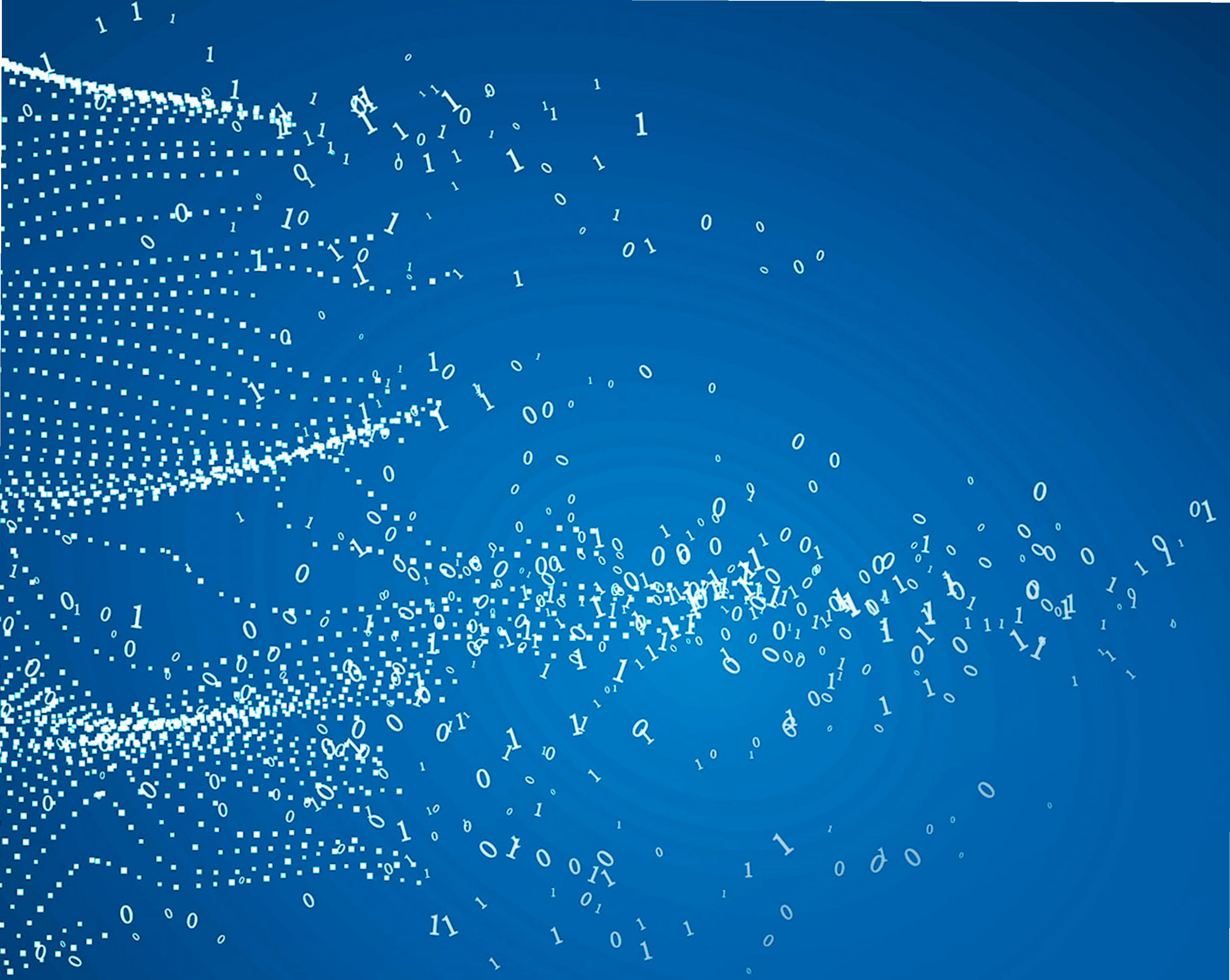


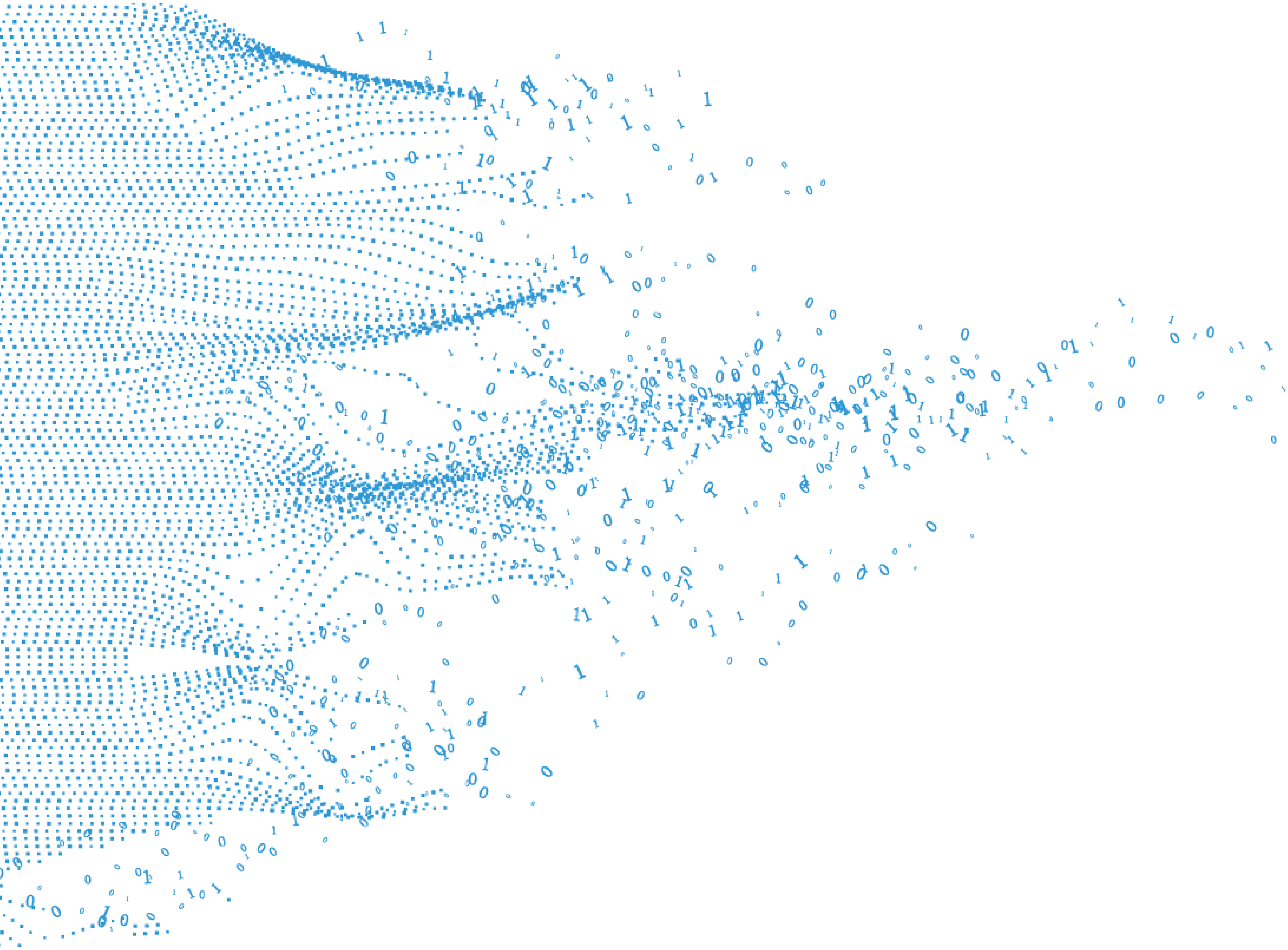
AIOps: A Guide to Operational Readiness

It's All About the Data

ScienceLogic

AIOps: A Guide to Operational Readiness, It's All About the Data

CONTENTS



Introduction – Making AIOps Work for You

The demands on IT organizations to deliver robust digital customer experiences – while taming a complex and dynamic IT infrastructure – have never been greater. Adoption of new technologies, from microservices to containers along with multi-cloud deployments, have created a highly ephemeral environment of IT workloads traversing a complex array of application and infrastructure components. All of these changes have rendered legacy tools obsolete, leaving visibility gaps and stalling automation initiatives. In response, companies are turning to Artificial Intelligence for IT Operations (AIOps), in the hopes of accelerating automation with a new generation of tools that combine big data with machine learning and artificial intelligence to move at machine speed.

The case for AIOps is clear; however, the concern among enterprise IT executives is less about whether AIOps will help them succeed, but more about how they can make it work for them today. One of the major challenges IT operations (IT Ops) teams face is data quality. Despite collecting more data now than ever before, according to Digital Enterprise Journal over 70 percent of the operational data available to IT Ops teams is not actionable.

Two aspects of current methods of operational data handling leave gaps that prevent data from being actionable.

> First is the timeliness of data.
> Given the transient nature of today's IT landscape, anything less than real-time data will result in inaccurate decision-making, which in turn will result in automation failures.
>
> Second, the data collected often lacks the required context to bring meaning to your data. As we described in our companion paper, ***Unlock the Power of AIOps: From Vision to Reality***, context ensures that relationships between IT elements and services are captured and tracked in real-time to account for adjacencies and relationships in decision-making.

Contextual relationship data is critical to a successful AIOps strategy. It can take multiple forms and notifies IT Ops of error conditions, such as:

- **Which production services are impacted and how important are they to my business?**
- **Did these events happen at the same time?**
- **Is there a dependency? How many other items are related?**
- **What change preceded this event?**

So, the central challenge of AIOps is: How to collect, organize and contextualize data – in real-time, such that it becomes actionable.

A Guide to Operational Readiness for AIOps

To make AIOps work, there are several steps that IT Ops teams can take to significantly improve the quality of data, which lead to a successful implementation of AIOps in the enterprise.

The following is an operational guide with five key steps designed to ensure that IT Ops teams can maximize their AIOps investment by using operational data that is accurate, timely, clean and contextualized.

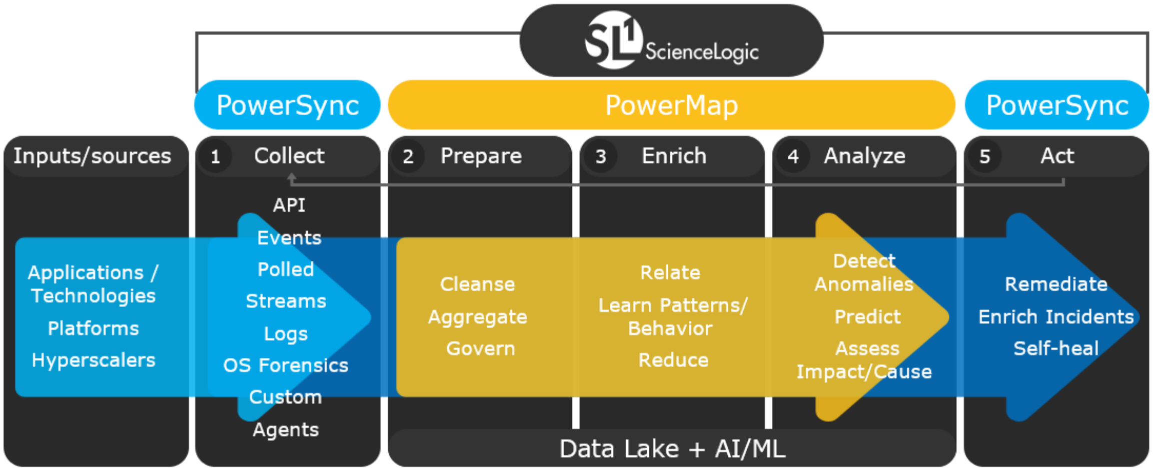
These are, in sequence:

- **Data Collection**
- **Data Preparation**
- **Data Enrichment**
- **Data Analysis**
- **Data-Driven Action**

We will describe each and relate it to the role they play within an AIOps framework.

ScienceLogic's 5 Steps to AIOps

A Bi-Directional Architecture for Learning and Notifications



Data Collection

Data collection includes the initial and continuing discovery of data from various sources, including agents, devices, applications, and services. IT Ops teams typically adopted simple network managing protocol (SNMP) polling, which is a relatively slow and cumbersome approach to discovery and collection. The limitations of SNMP include security aspects and lack of ubiquity, so it has always been necessary to add other data collection protocols like Secure Socket Shell (SSH), Windows Management Interface (WMI), PowerShell, and many others.

But for today's more cloud-centric management systems that manage virtual machines, cloud components or containers - beyond just hardware devices - a more programmatic approach is now possible with the widespread use of application program interfaces (APIs). Unlike traditional network or server element managers, APIs provide extensive data sharing and enable automation via a programmatic interface. Beyond APIs, we expect to see widespread use of publish and subscribe (Pub/Sub) interfaces to share data much faster, for a real-time, on-demand approach that streams operational data to the manager.

Once initial discovery is complete, device or element classification begins, enabling the establishment of appropriate monitoring policies. Classification ensures that the ongoing collection and monitoring process matches the type of asset (i.e., routers, storage devices, servers, or cloud elements are all monitored differently). While it may sound intuitive, this a slow, manual process with traditional monitoring systems. The data collected can be in multiple forms, including events, logs, time-series performance data, configuration records, application data and more.

While monitoring acquires data periodically, the discovery process must be continuous to allow for devices that are added, deleted or moved. This sort of domino effect also affects infrastructure relationships, so discovery must be ongoing, comprehensive and accurate. AIOps demands continuous, real-time knowledge of the current state and health of the IT environment to provide contextualized, actionable operational data.

Data Preparation

In this section, we'll discuss data deduplication and the importance of establishing a common data model. These two aspects of data cleaning and readiness contribute to preparing clean, actionable data.

AIOps cannot succeed if data is incomplete, imprecise, or out of alignment.

Data Deduplication

Duplication occurs when multiple sources and environments send alerts about the same instance. IT Ops can dramatically reduce duplication and cut through the alert noise by looking at the number of events matching a given criteria and how frequently they occur. Deduplication leads to a vast reduction in data processing, bandwidth, and storage of operational data. By eliminating duplicate data, IT Ops teams are not blinded by event noise during problem triage. Now retained data becomes actionable – a key criteria for AIOps success.

Establish a Common Data Model

A common data model enables different metrics, events and configurations to be connected, grouped and aggregated. Establishing a common data model is important when trying to correlate different behaviors across the IT estate given that servers, operating systems, storage devices, networks, cloud instances, and applications all behave differently and require their own sets of metrics - but consistent data models. Additionally, the common data model serves as the foundation for context or topology; understanding how everything is connected and related in a given IT ecosystem.

For AIOps to benefit the enterprise, IT Ops needs a single data lake with a comprehensive, normalized view across applications and infrastructure. Fragmented data silos make it extremely difficult or impossible to apply holistic artificial intelligence (AI) or machine learning (ML). For example, a hybrid IT deployment may feature web servers in a public cloud that are connected to a sensitive on-premise database that must remain in a fixed geographic location. Multiple disparate clouds and on-premise tools would make a consistent view of applications across such a hybrid environment impossible and also complicate the operational procedures required for effective problem-solving in a mixed environment.

Time synchronization is also a crucial requirement. Alignment of data in the correct time sequence, regardless of time zone or daylight-saving status, is critical for accurate analysis. Correctly sequenced data will show whether events are symptoms of other events, or not related.

Data Enrichment

The most critical element of data enrichment for AIOps is context. Context brings additional insight to raw data by adding metadata related to device or service metrics. This metadata may be simple asset records or more importantly, a set of relationships, connections or associations between that device or service and other related devices or services.

Context can take multiple forms and notifies IT Ops of error conditions, such as:

- **Which production services are impacted and how important are they to my business?**
- **Did these events happen at the same time?**
- **Is there a dependency? How many other items are related?**
- **What change preceded this event?**

Infrastructure Mapping

Infrastructure mapping refers to understanding the deployment, location, logical and physical dependencies of infrastructure elements. This understanding provides critical insight into the behavior of the infrastructure as well as the interactions between applications and their underlying infrastructure. For example, the attachment of storage units to virtual machines may be monitored by both a storage tool and a virtualization manager. The storage tool will be unaware of VM associations, and the virtualization manager will have no view of associated storage. A holistic approach offers a complete picture, as it provides the relationship between each component. When an application's performance becomes degraded due to a slow connection between a VM and its storage, neither point tool will show a problem.

Similarly, a lack of context can lead to incorrect prioritization of problems. By adding contexts such as associated dependencies and connections, we may determine whether a failed device is an essential component of a mission-critical business service, or a more routine level of follow-up will suffice.

Application Mapping

A full understanding of an application's topology and its relationship to the underlying infrastructure are key aspects of AIOps. Since AIOps addresses the holistic IT landscape – not just applications or infrastructure alone – application mapping is a crucial component of the data enrichment process. Mapping the topology of an application to its underlying infrastructure shows the physical and/or logical relationships between elements and allows IT Ops to pinpoint a particular element of the business service that is not functioning properly. It also provides visibility into the business-level impact of fault conditions. Application topology will come from many different sources. And since no single source will have maps for all applications, a complete picture may be fused from several disparate sources, including APM, NPM, Orchestration, etc.

Business Service Mapping

A mission-critical business service - such as a major e-commerce website - may include hundreds of smaller services (credit card validation, order placement, shopping cart, account profiles, etc.) and might comprise thousands of devices. In such cases, it is important to enable executive oversight with views of service health, availability and risk, rather than arcane views of a specific server or network device metrics that become meaningless and obfuscate the real picture of service performance. The ability to map out the components of mission-critical business services and track them as a group enables operations teams to focus on the right problems and prioritize their efforts on maintaining and supporting those services that most contribute to revenues and customer satisfaction. Where possible, a predictive view enables operations teams to forecast upcoming capacity bottlenecks or performance issues – ensuring the availability of critical business services is not uncompromised.

Data Analysis

The vast amounts of operational data collected by IT management systems place a significant burden on operations teams and incur significant analysis cost, in terms of both staffing and compute and storage resources. It is essential to reduce the amount of data required by eliminating spurious, non-actionable data and consolidating where possible. AIOps applies machine learning to solve problems rapidly. In cases where a machine learning engine is fed training data, faster results will apply if a training dataset is focused and contextualized, rather than fed into the ML engine in a raw state. Today's ML engines may falter or fail to achieve results when too much data is applied.

Dynamic Baselining

Dynamic baselining is the ability to determine what is considered normal and what is anomalous within the environment. Dynamic baselining is critical to businesses because the definition of operationally normal can shift due to external factors - such as a sales promotion - or internal factors, such as a malfunctioning device related to other infrastructure components. By eliminating alerts that would otherwise occur due to an outdated or inaccurate baseline, organizations can reduce operational noise and define "operationally normal" conditions.

Thresholding

When considering operational data, much emphasis is placed on alerts from thresholds or exceptions. Thresholds should only pertain to things that matter to the business. Too many thresholds or poorly set thresholds will create noise, regardless of how relaxed those thresholds are or whether a machine sets them via dynamic baselining.

For IT Operations teams, setting accurate thresholds can make the difference between being productive, or creating meaningless work due to 'alarm fatigue.' Incorrectly set thresholds can create extra work when the level of peak behavior fluctuates more than expected, potentially causing 'critical' threshold alerts that are not service affecting while missing genuine problems masked by the event noise.

Event Correlation

Event correlation refers to the ability to understand which events are related and how they are related. Event correlation can be used to isolate event noise from what is real and identify the root cause of a cascade of events to could occur. For example, if a web server CPU utilization is spiking, its memory utilization may also spike. These events may be a coincidence, or they may be related.

With event correlation, we can understand the time sequence of the events and use deductive reasoning or rules to see if the demand on the web server caused the memory spike.

Data-Driven Action

Once data is collected, organized and contextualized, decisions can be made with real insight, based on timely and accurate data. Automated actions can make changes, recommendations or notifications to ecosystem components or users. Automation can also streamline and speed workflows, eliminate manual process errors and free up skilled staff for more strategic projects.

Incident Automation and CMDB Enrichment

Automation can also be applied to enrich incidents with valuable context – where a fix may involve third-party triage, hardware replacement or a software patch. AIOps can run automated diagnostics to establish connectivity to a device, capture current device status, log entries leading up to the event, and attach detailed configuration information. By automatically including these in the upstream record of the incident, human operators have more information to support rapid triage and are freed to focus on non-automatable tasks.

With clean, contextualized data applied to incidents in an IT service management system, false positive incidents are significantly reduced, and you can automate the correct prioritization and routing of incidents to the right operations teams. These steps can save significant costs by eliminating false positives and reducing rework concerning incident reassignment.

Beyond incident automation, a CMDB can also be updated and enriched with real-time data, in context. Typically, a CMDB is populated only every few hours or daily – leaving too many gaps in data for today's highly ephemeral IT environment, especially where multiple clouds, containers and microservices are involved. By frequently updating a CMDB with contextualized data, the value of the CMDB is significantly enhanced and ITSM deployments can proceed much more smoothly.

Self-Healing

Perhaps the most prominent action that AIOps ecosystems can take relates to self-healing. Self-healing refers to the remedy of service-impacting issues automatically, based on an actionable insight or event triggering the automation. From stopping or restarting a VM to spinning up additional capacity, there are many potential use cases.

Automatic remediation - before the customer knows - is crucial, but requires the ability to act and react at machine speed. Automated remediation requires knowledge of business service impact – which is the insight into the consequences of an infrastructure fault. It involves looking at the entire IT landscape, understanding the interconnectedness of the devices, and establishing plans and procedures to mitigate failure or performance degradations.

Conclusion

These five steps (Data Collection, Data Preparation, Data Enrichment, Data Analysis, and Data-Driven Action) are the essential building blocks needed to prepare data for a successful AIOps implementation.

Continuous, real-time data collection, effective cleaning and deduplication, and normalization into a time-synchronized, common data model that's enriched with context all combine to make operational data actionable and serve as a reliable basis for insight-driven automation.

Raw data volumes can be reduced through intelligent baselining and topology inference, such that AIOps can create a base of self-healing IT applications and infrastructure, upon which digital transformation initiatives can succeed.

With clean, contextualized data driving an AIOps approach, IT organizations regain control over their destiny and deliver on the promise of AIOps. They can also achieve significant improvements to IT service delivery, with gains such as rapid fault isolation and reduced application outages, all while achieving dynamic business service visibility and vastly improved service agility. As a result, CIOs are empowered to deliver on strategic digital transformation initiatives while making substantial gains in business productivity, cost, and customer satisfaction. For the business, this leads to greater differentiation, business agility and a more innovative, customer-centric culture.

About ScienceLogic
ScienceLogic is a leader in IT Operations Management, providing modern IT operations with actionable insights to predict and resolve problems faster in a digital, ephemeral world. Its solution sees everything across cloud and distributed architectures, contextualizes data through relationship mapping, and acts on this insight through integration and automation. Trusted by thousands of organizations, ScienceLogic's technology was designed for the rigorous security requirements of United States Department of Defense, proven for scale by the world's largest service providers, and optimized for the needs of large enterprises.
https://ScienceLogic.com