Application Data Analytics for Modern Applications

How Sumo Logic's Unified Machine Data Platform Changes the Management Game

An ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) White Paper
Prepared for Sumo Logic

July 2016

EMA™
IT & DATA MANAGEMENT RESEARCH, INDUSTRY ANALYSIS & CONSULTING

Application Data Analytics for Modern Applications: How Sumo Logic's Unified Machine Data Platform Changes the Management Game

Table of Contents

Application Data Analytics for Modern Applications: How Sumo Logic's Unified Machine Data Platform Changes the Management Game

Overview

> ***telemetry*** *– an automated communications process by which measurements and other data are collected at remote or inaccessible points and are subsequently transmitted to receiving equipment for monitoring*[1]

The word "telemetry" has traditionally been associated with industrial automation. Smart meters, aircraft control systems, and self-driving cars all generate real-time metrics that, when monitored, gathered, and stored remotely, provide the raw material for the generation of a wide range of very diverse information.

Telemetry systems for smart meters gather usage-related information applicable to billing, capacity planning, and communicating with customers. Those for aircraft control systems power auto-pilot, landing, and hydraulics systems and provide notifications when parts or systems are about to fail. Self-driving cars rely on real-time analysis of hundreds of thousands of metrics per minute, gathered from cameras and sensors. From this data, internal guidance algorithms generate split-second responses to changing traffic conditions.

Traditionally, however, there is a separate unique type of telemetry, often called "operational metrics," generated by IT environments and the applications that execute over them. Virtually every component underlying applications produces time-series metrics; some produce logs or similar unstructured/semi-structured data as well. The purpose of such data has traditionally focused on IT support. Application Support and Operations teams need visibility to the technology environments they manage to monitor performance and availability, and to guide them during the troubleshooting process.

From this perspective, traditional performance monitoring/management solutions are essentially data collection and analytics systems, optimized to analyze and report on application execution in context with the operational metrics supporting the application. Ideally capable of real-time processing of diverse data generated across the execution environment, these products are designed to "understand" the application ecosystem and, in doing so, automate notifications and responses when problems occur.

As IT organizations begin to deploy a new generation of modernized applications, however, many find that incumbent performance management platforms cannot meet 100% of their application monitoring requirements. Applications are increasingly being hosted on virtual versus physical infrastructure, either on premises or in the public cloud. Recently, even virtual machines are being called "legacy containers"; approximately 15% of companies are already using container-based technologies, such as Docker, to deliver production services.

As modern applications become increasingly componentized and more loosely connected, two things start to happen. First, the volume of data they generate starts to grow exponentially, primarily because each component generates its own telemetry. Next, the types of data generated from these systems—which consists primarily of structured and semi-structured machine data—may or may not be recognized or correctly processed by incumbent performance management solutions. Finally, since existing in-house toolsets were never designed to process massive amounts of data at scale, there may well be monitoring gaps in the toolsets that make it difficult to cost-effectively support component-based, containerized, and API-connected applications.

[1] https://en.wikipedia.org/wiki/Telemetry

While some performance management vendors are adding support for basic log analysis to existing solutions, another approach is gaining momentum. Sumo Logic, a Software as a Service (SaaS) log management provider, has introduced a SaaS-based Unified Logs and Metrics (ULM) platform. Representing a new approach to Application Data Analytics, the creators of Sumo Logic have applied many of the principles behind the very successful science of Business Intelligence (BI) to the field of IT operational support. In addition, this data and analytics platform supports analysis and reporting functions which can be of significant value to a wide variety of both IT and Line of Business (LOB) applications.

This Enterprise Management Associates (EMA) white paper profiles Sumo Logic's ULM platform.

Telemetry, Big Operational Data, and the Real-Time Business

As applications are increasingly deployed to cloud platforms, virtual environments, containers, and similar platforms that abstract execution from the underlying infrastructure, there are multiple problems with a metrics-only approach to managing them. Increasingly, modern applications more closely resemble self-driving cars than they do traditional monolithic applications in terms of the volume of data they generate. Not only do they generate operational big data, they also generate data in diverse formats, most generally a combination of metrics and unstructured or semi-structured data.

This shift can tax the capabilities of traditional monitoring solutions, most of which are designed to support metrics-based data in operational environments that change relatively infrequently. Few have inherent support for unstructured data in dynamically changing operational environments, or for correlating this type of data with structured metrics for application support purposes.

As increasing numbers of companies define themselves as being in the midst of Digital Transformation, this combination of factors has become an issue impacting both IT and LOB. As software becomes increasingly business critical, Continuous Delivery practices are pushing new features into production very quickly—in many cases hundreds of times per day. This introduces another challenge to traditional monitoring solutions since modern software systems are increasingly characterized by both their lifecycle (a time-based dimension) and their stack (a technology-focused dimension). And from the management perspective and given the speed of change, visibility to both dimensions is necessary to building a foundation capable of supporting troubleshooting and root-cause analysis.

At the same time, in today's competitive business environment, the ability of IT organizations to accelerate software delivery has become a competitive differentiator. And the application management discipline—the ability to monitor performance/availability and solve production problems—has become a critical element in this process.

Figure 1 shows the primary bottlenecks impeding Continuous Delivery of new software features. The fundamental challenges relate to the adverse impact of escalating rates of production change. Manual troubleshooting processes are reducing the amount of time Development and Operations teams can spend on new digital initiatives. For this reason, IT practitioners have identified automation of the application management process as a key factor in accelerating the software lifecycle.

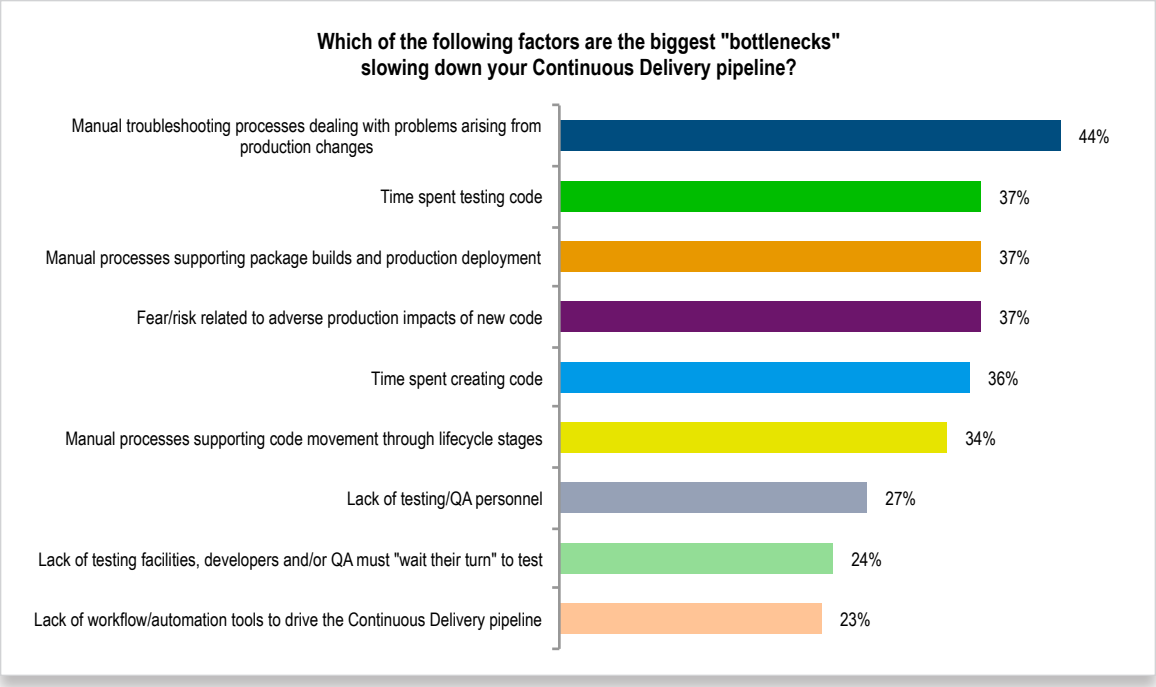

Figure 1. Manual troubleshooting is the top Continuous Delivery bottleneck.

Heterogeneity, scale, and integrations have all added complexity to application deployments; this complexity has dramatically increased the amount of data generated by instrumentation of application ecosystems. As Figure 2 shows, each new technology introduces a new type of machine data, and often a new set of tools. IT personnel are barraged by real-time data streams generated by toolsets supporting virtually every technology platform and silo, including time-based data such as that generated by lifecycle-focused and Release Management systems.

All of these factors combined underline the need for innovative, application-focused toolsets, particularly when those toolsets are purpose-built to support large volumes of diverse metrics in dynamically changing technology environments.

All of these factors combined underline the need for innovative, application-focused toolsets, particularly when those toolsets are purpose-built to support large volumes of diverse metrics in dynamically changing technology environments.

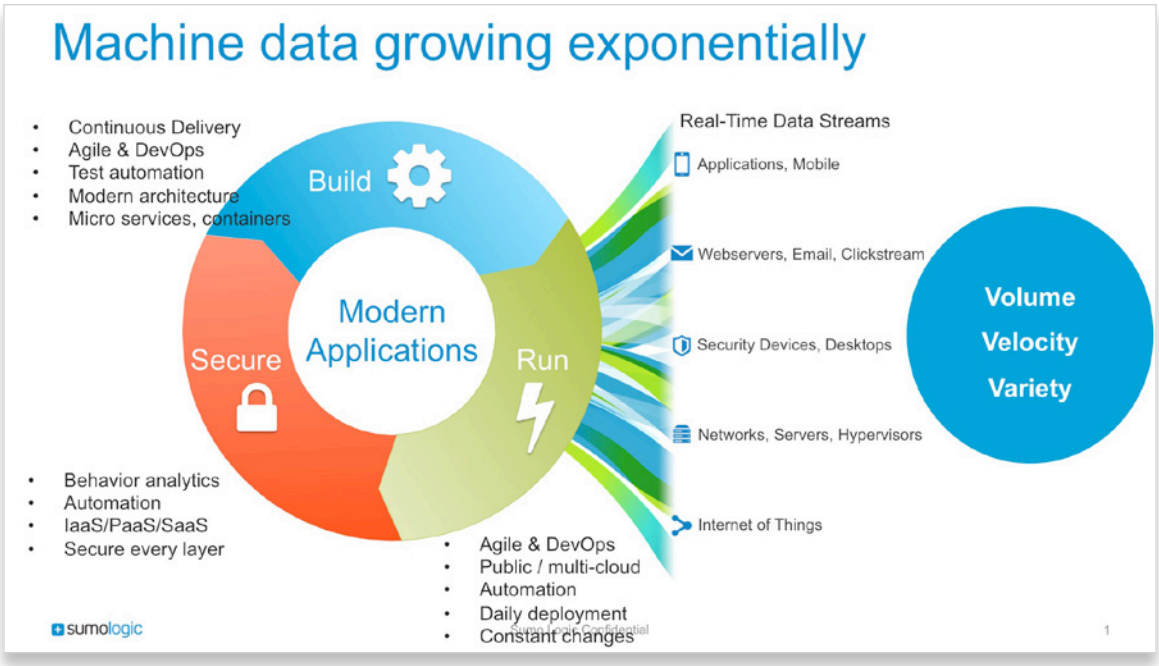

Figure 2. Volume of machine data grows with each new tool and technology.

Modern Application Environments Require a New Approach to Application Management

Application Management solutions have traditionally combined data collection from multiple technology silos with technologies such as synthetic transactions to trace and monitor end-to-end execution across the underlying ecosystem. There are several problems with this approach. By focusing primarily on structured data (such as time series–based metrics), these products have no inherent support for extracting or analyzing critical execution information from semi-structured and unstructured data. In modern component-based execution environments, this means that they lack visibility to the interplays between cause and effect that can be so critical to troubleshooting and root-cause analysis.

In short, companies that have successfully accelerated software delivery timeframes soon find they have won only half the battle. Many are finding that today's increasingly complex deployments are expensive to support and difficult to maintain at acceptable service levels. They are also struggling with blind spots in end-to-end execution, namely those elements of the application ecosystem that are either minimally instrumented or not instrumented at all.

Many companies are attempting to address this challenge by deploying tools supporting analysis of machine data, such as log files and messages, side by side with traditional performance management solutions. While this deployment model can be helpful in solving infrastructure-related problems, there is one enormous gap in this strategy. The two types of solutions remain separate and distinct, as the two types of data cannot be correlated and analyzed in context with one another. Another challenge is the fact that log-file analysis alone has no inherent support for end-to-end application management.

For a tooling environment to address the challenges presented by modern software-defined infrastructure and modern componentized applications, the following four common characteristics of complex applications need to be addressed:

- **Heterogeneity**
  Heterogeneity is difficult to support for a variety of reasons. Each platform generates structured metrics, and many also generate semi-structured and unstructured data in the form of log files and/or messages. Since traditional solutions see structured metrics only, there are monitoring gaps in the process. This means that IT teams often lack the accurate topology models that are essential to timely application support. The problem is exacerbated by hybrid applications, which can span on-premise and cloud-hosted platforms. These types of deployments create additional monitoring gaps, making root-cause analysis essentially an exercise in trial and error.
- **Scale**
  Scale is a growing problem impacting the delivery and management of business applications. Today's IT organizations, particularly those in enterprise-sized businesses and telecommunications carrier environments, are managing an unprecedented level of scale.

  In terms of visibility and control, dynamic scaling is a particularly difficult problem to solve. In virtual server and container-based systems, for example, scaling becomes a relatively simple matter of duplicating a container, a virtual machine, or a server cluster. From the management perspective, this means that, while the number of elements supporting the application can scale very rapidly to meet performance demands, the complexity underlying the application scales in equal proportion. This adds to the uncertainties relating to application topologies that increase the difficulties associated with application support.
- **Integration**
  Integration technologies have become the critical glue supporting end-to-end execution for many types of applications. Web Service, Enterprise Service Bus (ESB), and Application Programming Interface (API) connections are all widely used for purposes of application or data integration. Enterprise applications routinely exchange data with SaaS-based applications such as Salesforce and NetSuite. Mobile and containerized applications, in particular, have ushered in the era of the API, with API-based connectivity the primary interaction method for container-based microservices and often for interactions between mobile devices and the back-end systems supporting them.

  The quality of the end-user experience relies on the sum total of "all of the above"—in other words, on the performance and availability of the integration points themselves, the software supporting application execution, and the infrastructure supporting mobile, virtualized, and public-cloud services. This means that monitoring integrations has become a critical element of application support.
- **Speed (& Agility)**
  Execution environments are barraged with change from a wide variety of sources. Continuous Delivery practices have escalated rates of change on the software side. Elastic computing, cluster deployments, and container replication are all controlled by software, which means that the time delay traditionally associated with hardware deployments has now evaporated.

  At the same time, companies that are adept at technology agility can see significant impact on bottom-line revenue. Those that are most successful at accelerating delivery of software-based features and functions frequently report revenue growth significantly higher than their less agile competitors. However, while the impact to the business is almost always beneficial, the adverse impact to IT—and to production—can be substantial as well.

Sumo Logic: Cloud-Native Unified Logs and Metrics

The Sumo Logic platform (see Figure 3) has broad applicability as either a standalone Application Data Analytics (ADA) solution or as a complement to traditional Application Performance Management (APM) solutions. It was purpose-built to support real-time analysis of metrics, structured, and semi-structured data at scale and to analyze large volumes of metrics-based data in context with structured and semi-structured data. At the same time, the platform addresses a common problem of traditional management toolsets: It was engineered to support the high levels of granularity necessary for detailed, time-based analysis of ecosystem performance in context with code and infrastructure changes—both past and present.

Sumo Logic's latest platform release combines message-based and event-based "operational big data" with advanced IT- and business-facing analytics. The resulting combination is a distinctive machine data analytics platform supporting delivery of modern applications in complex, mission-critical execution environments. The solution is also distinctive in that it delivers Application Data Analytics via a SaaS model.

Sumo Logic's latest platform release combines message-based and event-based "operational big data" with advanced IT- and business-facing analytics. The resulting combination is a distinctive machine data analytics platform supporting delivery of modern applications in complex, mission-critical execution environments.

Sumo Logic utilizes two separate processing engines to parse and analyze the two types of data it collects: structured/semi-structured data and time-series data. The two types of data can then be analyzed, correlated, then normalized by time. A wide variety of reports can then display and sub-analyze the two types of data in context with one another and with time of occurrence. This provides IT support professionals with a valuable new data source enabling them to see causes and effects between infrastructure execution (metric data) and application execution (log data).

A host of features set it apart from competitors:

- **Single integrated platform supporting virtually all machine data types –** Sumo Logic has built its own analytics platform, which is designed to natively ingest and index diverse and dynamic data sets and analyze them in real time. Sumo Logic's machine learning approach applies equally to both log and telemetry data, so both can be analyzed in context with one another.
- **One integrated platform featuring a single pane of glass unifying logs, standard metrics, and custom metrics –** Sumo also provides correlation and visualization capabilities for custom metrics coded into applications.
- **Open and flexible –** Sumo natively delivers visibility to standard infrastructure metrics such as those delivered via APIs, WMI, etc.
- **Real-time data, not just sampled or summarized data –** Sumo customers can see data gathered via real-time streaming analytics two seconds after it hits Sumo's services.
- **Simple analysis and reporting –** One-click comparison between log patterns, easily understood reports, and single-click interactions with outlier predictive models are all features of the Sumo platform. It also time stamps all incoming data with arrival time for calculations of complex analytics concerned with accurate time comparisons.

- **Scalability, data retention, and support for enterprise-sized companies –** In addition to high-speed processing engines and support for large-scale data processing, Sumo has introduced features such as different retentions for different data sets and access control definitions for diverse data access across departments.
- **Lifecycle focus spanning development and production, supporting development/deployment and production scenarios –** As companies move from monolithic to distributed applications, it becomes increasingly important to integrate both log and metrics data sources during pre-deployment as well as during production. Sumo supports real-time machine learning capabilities for unified collaboration across the entire application lifecycle. Its open architecture allows API-level connections with JFrog, Chef, Docker, Puppet, Salt, Ansible, Jenkins, Azure, AWS, and others.
- **Cloud native –** The Sumo platform was initially engineered for public cloud hosting, and is still deployed and sold as SaaS. For customers, this means broad coverage with the convenience and cost-efficiency of a SaaS solution.
- **Extensive support for cloud technologies –** Sumo has functional integrations with a variety of SaaS vendors via direct cloud-to-cloud integration engines. Sumo is also deeply integrated with Akamai and AWS, including integrations with AWS Kinesis Streams and AWS CloudWatch, which gives users insight into AWS events and performance metrics.
- **Compliance/security –** Sumo Logic maintains the highest level of security certification to protect data, including CSA STAR, PCI DSS 3.1 Service Provider Level 1, ISO 27001, SOC 2, Type II Attestation, FIPS 140 Level 2, and HIPAA.
- **Intuitive, with minimal training required –** Combined data set provides value for multiple stakeholders including Business Intelligence teams, Line of Business, Technical Support, and Line of Business IT.
- **Price –** Customers pay based on the number of data points processed per minute versus by server.

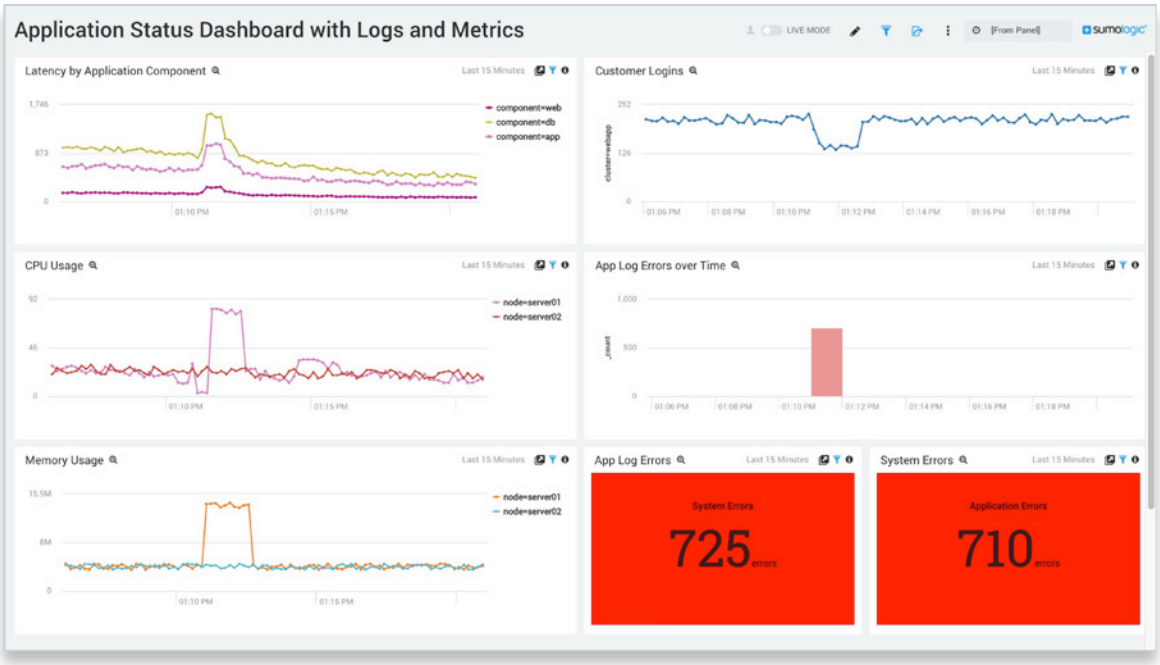

Figure 3. One comprehensive platform for all machine data—metrics and logs

EMA Perspective

As companies begin to deploy application components over a wide range of diverse platforms and technologies, the resulting polyglot of metrics, logs, and messaging creates a torrent of data. IT professionals report that they are drowning in data but still lack the information they require to streamline the process of application support.

Analytics are the key to solving this challenge. And particularly as application ecosystems become increasingly complex and dynamically changing, the value proposition of solutions capable of analyzing a wide range of data types and enormous volumes of machine data in real time will continue to accelerate. Indeed, EMA research is showing correlation/analytics tools as a number one "wish list" product for 2016, for exactly the reasons outlined in this paper.

Within today's rapidly changing business and IT landscapes, Sumo Logic continues to enhance its Application Data Analytics platform. With native support for analysis of combined quantitative (metrics-based) and action/messaging (semi-structured and unstructured-based) insights in context with one another, this combined approach positions Sumo Logic as distinctive in its class. Providing the continuous, real-time execution insights necessary for building, running, and securing modern applications, these capabilities extend the value proposition of existing APM installations while also providing robust management analytics as a standalone solution.

About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst firm that provides deep insight across the full spectrum of IT and data management technologies. EMA analysts leverage a unique combination of practical experience, insight into industry best practices, and in-depth knowledge of current and planned vendor solutions to help EMA's clients achieve their goals. Learn more about EMA research, analysis, and consulting services for enterprise line of business users, IT professionals and IT vendors at www.enterprisemanagement.com or blogs.enterprisemanagement.com. You can also follow EMA on Twitter, Facebook or LinkedIn.

Corporate Headquarters:
1995 North 57th Court, Suite 120
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com

3418.071816