Machine Data Analytics for AWS Services

AWS - Sumo Logic White Paper

Table of Contents

Abstract

This whitepaper helps IT administrators, DevOps managers, developers, Cloud/IT Operations, Security Professionals, and other IT professionals understand the benefits of using Sumo Logic, a multi-tenant, cloud-native machine data analytics platform, in their public, private, or hybrid cloud environment.

This paper provides an overview of:

- Machine Data Analytics
- Benefits of a Multi-Tenant, Cloud-Native Service in the Cloud
- Criteria for Assessing a Cloud Service
- How Sumo Logic Can Help

This paper concludes with example use cases that showcase how Sumo Logic can be benficial for your organization's workloads on Amazon Web Services (AWS).

Introduction

The AWS Cloud is no longer the future of information technology infrastructure, but rather a present day reality. As data growth continues to expand, organizations around the world are avoiding building, and in some cases, actively closing down, on-premises datacenters as paying for the total cost of ownership for such environments is becoming an unwieldy, or at the very least inefficient, use of capital. This trend can be observed with the increasingly rapid adoption of cloud services over recent years. According to the new Worldwide Semiannual Public Cloud Services Spending Guide from International Data Corporation (IDC), worldwide spending on public cloud services will grow at a 19.4% compound annual growth rate (CAGR)--almost six times the rate of overall IT spending growth–from nearly $70 billion in 2015 to more than $141 billion in 2019.

Some organizations worry about losing visibility into their workload when moving to the cloud. The reality is that when companies migrate to the AWS Cloud, they have the opportunity to leverage cloud-native services and tools that were designed specifically for the agility and scalability of the cloud, avoiding excessive cost, lengthy implementations, and the need to for additional internal IT resources to manage the platforms and the hardware. An example of this would be logging and monitoring services that were frequently considered too expensive or time consuming to utilize in an on-premises environment. Because the scalability of the AWS Cloud allows you to spin up new instances on-demand and leverage pay-as-you-go pricing, logging and monitoring has become not only more affordable, but more foundational than ever. Since logging and monitoring on AWS is less expensive and simpler to implement than on-premises, it is easier than ever to have complete coverage of your environment, meaning you don't need to miss out on any data.

Related to logging and monitoring, one area of opportunity is machine data analytics. Service that leverage AWS services:

- **Amazon Simple Storage Service (Amazon S3)** - A secure, durable, and highly-scalable cloud storage service
- **Elastic Load Balancing (ELB)** - An AWS service that automatically distributes incoming application traffic across multiple Amazon Elastic Cloud Compute (Amazon EC2) instances
- **Amazon CloudFront** - A global content delivery network (CDN) services that accelerates delivery of your websites, APIs, video content, and other web assets
- **AWS CloudTrail** - A web services that records AWS API calls for your account and delivers log files to you
- **Amazon Virtual Private Cloud (Amazon VPC) Flow Logs** - An AWS feature that enables you to capture information about the IP traffic to and from network interfaces in your VPC

These, and other AWS services, generate machine data in the form of log files and time-series metrics that can be analyzed in real time to improve visibility and mitigate security risk. Amazon CloudWatch (a monitoring service for AWS Cloud resources and the applications on them) aggregates these logs for high-level monitoring and alerting in AWS workloads. AWS Partner Network (APN) Advanced Technology Partner and AWS Security Competency Partner Sumo Logic applies advanced analytics and machine learning to logs and time-series metrics allowing organizations to gain real-time, full-stack visibility into cloud and hybrid environments. Sumo Logic does not require instrumentation and easily captures machine data from AWS. It pulls log files from from a variety of AWS services, including AWS CloudTrail and Amazon VPC Flow Logs, and centralized metrics from Amazon CloudWatch to provide continuous intelligence. This continuous intelligence can help companies accelerate the building, running, and securing of modern applications and enables them to achieve greater visibility intotheir workloads compared to an on-premises environment. Sumo Logic also supports cross-functional collaboration by correlating data from multiple data sources, showing data in the context of time-series metrics, thereby providing a common source of truth for monitoring and troubleshooting.

Machine Data Analytics

Machine data is data generated automatically by the activity of a computer, application, or device. This machine-generated data often come in the form of logs and can contain immensely valuable insights about the application/infrastructure and its health. The biggest problem with harnessing machine data is the sheer volume of data being generated. Raw machine data contains billions, if not trillions, of log and metric data points and is increasing in quantity at an exponential rate. The volume and velocity of this data growth can be difficult for single-tenant analytics solutions to handle. Additionally, machine data can come in a variety of formats and can be structured, unstructured, or semi-structured:

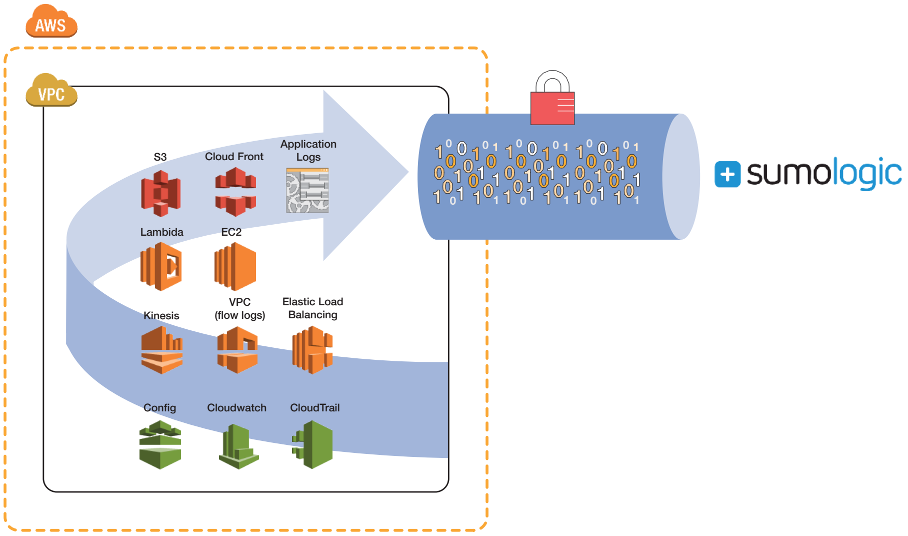
- **Structured data** refers to data that resides in a fixed field within a file, such as a field in arelational database or a time-series metric such as CPU utilization. Structured data can beeasily stored, retrieved and analyzed.
- **Unstructured data** refers to all those things that cannot be easily classified such as streamingdata, videos, images, blogs, and wikis.
- **Semi-structured data** is a cross between the two. It lacks the strict data model of structureddata but has tags or other markers that help you identify certain elements. Log files are a goodexample of semi-structured data.

With this in mind, it is important to use a data analytics platform optimized to handle all types of machine generated data, including custom metrics.

The volume, variety, and velocity of cloud generated data is constantly increasing and while old tools organizations are familiar with from legacy environments may not be able to scale to handle the data inputs in the AWS Cloud, cloud-native services and tools have been designed to do so, and can do so in real time.

As organizations transition application workloads to AWS, it is critical that they monitor the delivery, performance and security of those workloads. Organizations need continuous intelligence about their cloud-based workloads in the form of real-time, machine data analytics that generate operational, security and business insights and trends. This intelligence is critical for them to drive competitive advantage, add business value, innovate, and grow. More and more data is constantly being generated, and by moving to the AWS Cloud, organizations have the opportunity to re-think the instrumentation of their Big Data platforms.

By moving to AWS, organizations can automate much of their Big Data resources and gain better telemetry through services such as AWS CloudTrail, which logs user account access and activity, giving organizations visibility they couldn't get with an on-premises environment. Further, organizations gain access to a centralized flow of information through Amazon VPC Flow Logs and Amazon CloudWatch. Sumo Logic can ingest data directly from these sources and by applying advanced analytics and machine learning, allow customers to monitor and visualize their IP traffic, user behaviors, error codes, etc. and gain operational and security insights. Examples of insights could include real-time alerts on suspicious user activity, changes to your infrastructure, and security and network ACL changes from AWS CloudTrail feed. The Sumo Logic App can also integrate with Amazon CloudWatch at scale and provide operational visibility into network latency and failures, behavior trends and traffic patterns and provide anomaly alerts through ingestion of Amazon VPC Flow Logs.



In order to leverage machine data analytics, Sumo Logic uses machine learning to analyze data outputs from AWS sources to uncover patterns in log and time-series metrics data and surface anomalies and outliers, allowing organizations to get real-time feedback on potential business and security insights. With Sumo Logic, when an anomaly is detected in your AWS environment or edge devices, you can quickly drill down through data to find the source and cause of the anomaly in minutes. From there, you can quickly assess the situation and take actions as needed. For example, if Sumo Logic detects an abnormal spike in failed login attempts, you can quickly look to see if there is a specific account or accounts that are failing to log in, see where in the world the attempts are coming from, and issue a prompt for the account owner to change their login credentials if it appears their account might be compromised.

In addition, leveraging machine data can support and automate compliance initiatives such as PCI or HIPAA. Audit data, in the form of log files, can be used to track activities, identify suspicious usage, and query historic data to support audit requests.

A concern many organizations have when considering a move to the cloud is around maintaining full-stack visibility into a hybrid cloud environment. According to the RightScale Survey, “Cloud Computing Trends: 2016 State of the Cloud Survey”, 71% of respondents (which consisted of 1,060 IT professionals, with 42% representing enterprises with 1,000+ employees) have adopted a hybrid cloud environment. Sumo Logic addresses the question of maintaining full-stack visbility by analyzing all data streams with unified views from AWS, on-premises, and any other cloud environment an organization might have.

Full Stack Operational Visibility for AWS

Sumo Logic enables your organization to gain granular visibility across your entire stack, improving your ability to detect and remediate anomalies, plus be more flexible and agile. Additionally, as described in greater detail in Figure 1 below, the benefits of a multi-tenant, cloud-native service like Sumo Logic greatly outweighs those of single-tenant or on-premises based services. Benefits of using Sumo Logic on your AWS workloads include improved feature velocity, scalability, access to predictive analytics and the industry’s first real-time unified logs and metrics analytics platform, and easiness to get started. Additionally, all organizations receive the same level of security from Sumo Logic regardless of the organization’s size and usage level of Sumo Logic. Sumo Logic helps customers fulfill their responsibilities under the AWS Shared Responsibility Model by integrating closely with and aggregating data from several AWS sources to provide end-to-end visibility and actionable security related data. Additionally, Sumo Logic features AES 256-bit encryption and maintains some of the highest security certifications available, including:

- PCI DSS 3.1 Service Provider Level 1
- CSA Star certifications
- SOC 2, Type II attestation
- ISO 27001 certification
- E.U. - U.S. Privacy Shield
- Attestation of HIPAA compliance
- FIPS 140 compliance

Sumo Logic was born in the cloud and was purpose-built to support AWS workloads and to help your organization stay current and relevant. Additionally, Sumo Logic provides pre-built applications and integrations with your AWS environment so you can easily visualize and detect anomalies. Sumo Logic has already pre-built the following applications for and integrations with the following applications:

- **AWS CloudTrail** - Feeding CloudTrail data into Sumo Logic allows you to track and monitoryour AWS workloads for operational and security insights.
- **Amazon CloudFront** - Analyzing and correlating Amazon CloudFront data with the origin data and/or other data sets allows you to improve availability and end-user experiences whileenforcing rigorous security controls.
- **Elastic Load Balancing (ELB)** - Analyzing raw ELB data helps you determinelatency issues and optimize your system configuration based on information from across allAvailability Zones about source IPs and traffic to Amazon Elastic Compute Cloud (Amazon EC2) backend instances.
- **Amazon Simple Storage Service (Amazon S3)** - Examining critical elements of your Amazon S3 serviceincluding access logs and revealing information such as request type, resource name, and the time the request was processed.
- **Amazon VPC Flow Logs** - Ingesting your VPC Flow Logs directly into Sumo Logic continuousintelligence service allows you to monitor and visualize your IP traffic within your VPC for operational and security insights.
- **AWS Config** - The Sumo Logic app for AWS Config delivers real-time interactive visualizationsto track configuration changes made to critical resources in your AWS workloads.
- **Amazon Kinesis Connector** - The Sumo Logic Amazon Kinesis Connector enables real-timeAWS Kinesis data streams to be ingested by the Sumo Logic cloud-native platform.
- **AWS Lambda** - The Sumo Logic App for AWS Lambda enables monitoring and tracking ofkey operational indicators such as duration, memory utilization and number of requests acrossall your Lambda Functions
- **AWS Lambda Function for Sumo Logic** - Allows for high volume ingestion of data fromCloudwatch Logs and Amazon Kinesis

As noted above, Sumo Logic can be integrated with AWS Lambda, scaling to match the size of your code being run and providing the operational and security insights much like workloads running on Amazon EC2. Using Sumo Logic in this capacity can enable an organization to quickly test and debug code being tested in Lambda to ensure proper function and operation of that code. Additionally, with the Internet of Things (IoT) concept expanding at an exponential rate, drawing data generated through connected devices can provide even more valuable insights. Sumo Logic's integration with Lambda allows users to bring that Internet of Things generated data into their centralized data stream.

Unified Logs and Metrics

Sumo Logic is introducing the industry's first machine learning driven machine data analytics platform that can transform logs and metrics into real-time continuous intelligence for managing modern applications. This service leverages machine learning algorithms to unify log data and time-series metrics to uncover real-time insights into your application needs and new customer opportunities. Doing this helps provide immediate context to time series metrics to enable rapid root cause analysis and reduces the need to waste time switching between different tools to address the opporutnity, insight or security risk.

By being able to unify structured time-series data and semi-structured log data, Sumo Logic can provide your organization with real-time continuous intelligence. With Sumo Logic's machine data analytics platform, you gain access to a single, universal platform for DevOps, TechOps, InfoSec, and Line-of-Business (LoB) teams to access and control information relevant to functional needs, and you can improve cross-function collaboration through shared, real-time dashboards of KPIs and log metrics. You also gain top-to-bottom visibility across your application stack to contextually troubleshoot modern applications, plus the ability to quickly identify service level degradations on underlying infrastructure.

Sumo Logic's machine data analytics platform for unified logs and metrics can help you accelerate root cause analysis using machine data, plus utilize statistics and algorithms to analyze large volumes of machine data to transform it into actionable insights in real-time. It can also leverage the unified logs and time-series data to provide algorithmic support for machine learning at scale.

Criteria for Evaluating Machine Data Analytics Services

It is important to take a look at the variety of machine data analytics services on the market. Many were originally created for on-premises environments and have been "SaaS-ified." According to the IDC report, "Why Choose Multi-Tenant Cloud-Native Services for Machine Data Analytics," a common method of deploying these SaaS-ified services is through single-tenant packaged server and client software intended to be run from a company's datacenter. This single-tenant approach places the greatest burden on the customer because the customer is responsible for handling all aspects of implementation of the software. Everything from deploying, running, supporting, customizing, integrating and upgrading the software, to assembling the stack from bare metal up and ensuring security is the responsibility of the customer. Additionally, there is an up-front sunk cost and IT organizations must pay for the server and software regardless of whether it is used or not. Organizations should consider to selectively spending their scarce resources on building differentiated applications and services that makes them competitive and avoid managing low value / high risk operational aspects involving software stack management and related maintenance.

More recently, companies have been leveraging cloud services such as Amazon EC2 to take advantage of the high-performance and consistent services provided by AWS. These services are turnkey and ensure that the customers of Independent Software Vendors (ISVs) like Sumo Logic can focus on using the service and improving their workloads and applications without needing to worry about managing the software and server installation, updates, and so on. Figure 1 goes into additional detail about the differences between an on-premises, single-tenant, and multi-tenant service.

Pros and Cons of Various Software Deployment Options

	On-Premises	Single-Tenant Service	Multi-Tenant Cloud-Native Service
Pros	• Controld over the full stack	• Transfer operations responsibility to vendor • Dedicated compute resources	• Elasticity and socialbility without any constraint • Level playing field for all customers (same feature, same support quality, same service SLA) • Lower TCO • Up and running within minutes • Scalable secure solution
Cons	• Customer is responsible for • Troubeshooting • Scaling • Compliance and security • Upgrade • Scalability of management stack • Higher TCO • Lack of ease of use • Lack of scalability and elasticity without overprovisioning • Inability to drive continuous improvement and innovation	• Enterprise software unable to leverage new architecture (e.g., lack of security, compliance, multi-tenancy) • Different level of service, support, and security based on revenue contribution by the customer • Slow innovation and feature velocity • Lack of elasticity and scalability without constraints • Fewer features and lack of security for smaller customers	• Lack of control over the entire machine data analytics stack

Source: IDC, 2016

Viewing Figure 1, it is easy to see that multi-tenant, cloud-native services have a great number of benefits over single-tenant and on-premises based services with few of the cons. It isn't always immediately obvious whether a service is cloud-native or not, however. Since it is important to maximize your organization's workload efficiency by using services that can scale with it and are flexible enough to meet your needs, Figure 2 offers validating questions to ask when determining whether a service is right for your organization.

What Should I Ask?	Why Should I Ask?
Is it offered as a cloud-based service?	Determine that it is offered as a cloud-based service that you will not need to host the service yourself.
Is the service multi-tenant?	Confirm that the cloud solution is multi-tenant rather than single-tenant so you get all the benefits of elasticity, scale, feature velocity, and equal status regardless of customer spend.
How often am I allowed to go over the license limit?	If you are not able to go over the license limit, you may be unable to burst when you need it most or face delays for manual approvals and provisioning.
How often will I have access to new features? Will I need to wait for annual updates?	Access new capability just as soon as it's ready – or later: it's your schedule. Provide real-time feedback and get the features that matter faster.
How difficult is it to upgrade to a new release?	Releases should have the same features and 100% of integrations and customizations should be auto-upgraded, for consistent user experience.
Is on-boarding automated and configuration access consistent?	Automated on-boarding and consistent configuration access creates an improved, consistent user experience, simple access to resources, and fewer trouble tickets trouble tickets.
Which security certifications are available? Are those certifications and attestations applicable to all customers?	Validate that the cloud solution has security and compliance certifications in addition to those provided by the cloud provider. PaaS. For example, ensure that vendor service can demonstrate compliance with the U.S.-EU Safe Harbor framework, ISO 27001, SOC 2 Type II, HIPAA, and PCI DSS 3.0.

Source: IDC, 2016

Common Use Cases

Improve Visibility as You Moved Workloads to the AWS Cloud

Sumo Logic enhances customers' abilities to simplify and accelerate movement to the AWS Cloud. Because of the logging and monitoring capabilities of Sumo Logic, users can easily monitor new workloads for operational efficiency and compliance as they are moved from an on-premises environment to the AWS Cloud. Sumo Logic scales on-demand and streamlines everything from small to massive cloud deployments by providing real-time visibility into operational status, KPIs, usage metrics and compliance violations. The flexibility of Sumo Logic to provide visibility across public, hybrid, and on-premises environments plus its scalability allow customers to move to the cloud with great efficiency. Since Sumo Logic can continuously monitor workloads, you can ensure they are working and maintaining compliance as they move to the cloud without needing to pause migration to scale up monitoring services for each new workload.

Accelerate Root Cause Analysis

AWS services such as Elastic Load Balancing (ELB), Amazon EC2, AWS CloudTrail, and Amazon S3 can transfer machine data to Sumo Logic for root cause analysis. Through this integration, developers and operations teams can quickly detect anomalies and outliers and remediate them. Sumo Logic uses machine learning algorithms to identify patterns within millions of logs and surface anomalies using Sumo Logic LogReduce. DevOps teams can also use LogCompare to identify potential problems and resolve problems quickly by comparing two time periods and highlighting the areas of change that may have triggered the problem. For example, LogCompare can be used to test the impact of a new code push or cloud migration. Additionally, Outlier Detection allows for dynamic thresholds to be set on any number of KPIs/metrics to alleviate the operational burden and inaccuracy of static threshold based alerting.

Ultimately, by surfacing problems quickly, DevOps teams can accelerate release cycles and improve application performance.

Strengthen Security Posture with Advanced Security Analytics

As noted above, InfoSec teams can also leverage Sumo Logic's machine data analytics to strengthen their security postures through user behavior monitoring. Unlike traditional SIEMs, there is no need to write rules or alerts based on events. With Sumo Logic, security professionals can monitor user behavior and visualize or alert on anomalies to quickly drill down on an unusual spike in user activities. As an example, Sumo Logic could detect an abnormally high amount of login attempts and be used to determine characteristics of the login attempts. Characteristics could include the accounts attempting to log in, geographic locations of the login attempts, what is trying to be accessed, and why exactly an account can't access the data or service (wrong password, invalid access key, and so on), etc.

After determining what the suspicious activity looks like, you can view all the logs during the time frame of the anomaly, and use LogReduce to hide logs that aren't relevant to the anomaly, saving hours of sifting through irrelevant log data. With this, you can quickly drill down and determine what exactly the problem is, an example being a unique signature where a user makes an access key inactive. For more information on this use case, watch this demonstration of how Sumo Logic can be used to rapidly identify a user with compromised credentials.

Simplifying Compliance

Sumo Logic also helps users by simplifying compliance which is often one of the biggest barriers to cloud adoption. Administrators can monitor user access, platform configuration changes across all AWS and on-premises workloads, and generate audit trails to demonstrate compliance with internal security standards and industry regulations such as PCI and HIPAA. Pre-built apps and powerful machine learning algorithms automate cloud audits and quickly uncover compliance violations, outliers and anomalies in real-time. Many customers begin with a compliance initiative and then expand their use of data analytics to strengthen their security posture or improve monitoring and troubleshooting.

Conclusion

AWS provides flexible, scalable, and secure cloud services including a host of tools to make your migration to the cloud easy, quick and secure. Advanced Technology Partner Sumo Logic can assist you in making your migration efficient and help you gain operational, security and business insights across your AWS and hybrid environments. With AWS, you can get your cloud environment launched in minutes with a greatly reduced TCO and no upfront or sunken costs.

By using Sumo Logic on AWS, you can start taking advantage of real-time continuous intelligence from Sumo Logic's unified logs and metrics analytics platform virtually instantly. Additionally, Sumo Logic is a multi-tenant cloud native service that utilizes AES 256-bit encryption and allows you to gain full-stack visibility across cloud, hybrid, and on-premises environments, making it an ideal platform for your AWS workloads.

To see a demo of how Sumo Logic's machine data analytics for log and time-series metrics helps improve visibility and reduce the risks associated with moving to the cloud, click here.

About Sumo Logic:

Sumo Logic is a secure, cloud-native, data analytics service, delivering real-time, continuous intelligence across an organization's entire infrastructure and application stack. More than 700 customers around the globe experience real-time operational, business and customer insights using Sumo Logic for their DevOps, IT ops and security and compliance use cases. With Sumo Logic, customers gain a service model advantage to accelerate their shift to continuous innovation,increasing competitive advantage, business value and growth. Founded in 2010, Sumo Logic is a privately held company based in Redwood City, CA and is backed by Greylock Partners, DFJ, IVP, Sutter Hill Ventures, Accel Partners and Sequoia Capital. For more information, visit www.sumologic.com.

About AWS:

For 10 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud platform. AWS offers over 70 fully featured services for compute, storage, databases, analytics, mobile, Internet of Things (IoT) and enterprise applications from 35 Availability Zones (AZs) across 13 geographic regions in the U.S., Australia, Brazil, China, Germany, Ireland, Japan, Korea, Singapore, and India. AWS services are trusted by more than a million active customers around the world – including the fastest growing startups, largest enterprises, and leading government agencies – to power their infrastructure, make them more agile, and lower costs. To learn more about AWS, visit http://aws.amazon.com.