Publication date:
1 March 2022
Author:
Tony Baer

Profiting from Log Analytics

An Omdia white paper for Amazon Web Services

In partnership with:

Brought to you by Informa Tech

Contents

Summary

Catalyst

Log analytics are becoming table stakes for enterprises whose business models require them to be always on. With volumes growing exponentially faster than those of conventional business data, machine data provides an opportunity to gain real-time insight into the operations of always-on enterprises. Common use cases have been as varied as managing IT infrastructure, overseeing application performance, and protecting against cyberthreats. But for most organizations, incorporating real-time log analytics across the enterprise is unfamiliar territory since most implementations have originated with specific applications or at the line-of-business level.

Omdia view

There are well-established point solutions for functions such as IT operations management, application performance management, and security information and event management (SIEM), but the single-purpose and proprietary nature of these solutions limits them to be highly siloed and costly to scale as data volumes grow. OpenSearch, an open source distributed search and analytics engine, is designed for diverse but often interrelated use cases for running always-on enterprises. It also has easy-to-use data ingestion and visualization tools with enterprise scalability. To optimize costs and time, customers can also use a fully managed service to deliver on the cloud's promise of IT simplification, especially when it comes to scaling. Amazon OpenSearch Service provides a managed service that extends on the open source platform to make the service scalable, easy to use, and economical.

Key messages

- The explosion of machine data has thrust logs into the spotlight. Generated by websites, applications, mobile devices, servers, sensors, and other Internet of Things (IoT) devices, logs can tell the story of your business. They can tell you what happens when a customer visits your website, your application is underperforming, or cyberattacks are occurring.
- OpenSearch is well-suited for machine data-based use cases (e.g., real-time application monitoring, observability, SIEM, clickstream analysis, and centralized logging) because of its flexibility, ease of use, and scalability.
- The cloud is a natural environment for deploying OpenSearch and related components, both because of the ability to tap scalable infrastructure and because of data gravity; much of the data already lives there.
- To gain maximum advantage from open source OpenSearch in the cloud, managed cloud services for open source OpenSearch promise the best of both worlds: the scale and flexibility of open source technology paired with the simplification of deployment and operation that managed services deliver.

The use cases for log analytics with OpenSearch

Real-time monitoring

ITOps and DevOps have become popular starting points for the use of log analytics. Having your logs in a central location that features rich search and the ability to locate errors across hundreds or thousands of machines and services with a simple interface are key to understanding the behavior of IT systems and applications. When sluggish performance drives customers to abandon their online shopping carts, developers can monitor logs and correlate events to help them nip problems in the bud. OpenSearch, a new open source project, was designed as a highly scalable system for providing fast access to large volumes of data to support these use cases. Using analytics with OpenSearch, developers can understand how their applications are performing. If crashes happen, OpenSearch can help them understand why they occurred. With the ability to readily search and perform analytics when outliers occur, developers reduce guesswork. That results in a virtuous cycle of improved productivity and customer experiences.

Observability

Having originated with digital online services such as Twitter and Etsy, *observability* emerged with the need to improve practices for system monitoring and management with internet-scale, distributed systems. The observability challenge has grown even more complex with the move to cloud native environments where applications are being refactored into microservices. Observability brings together traditionally siloed practices of collecting logs, which include the following:

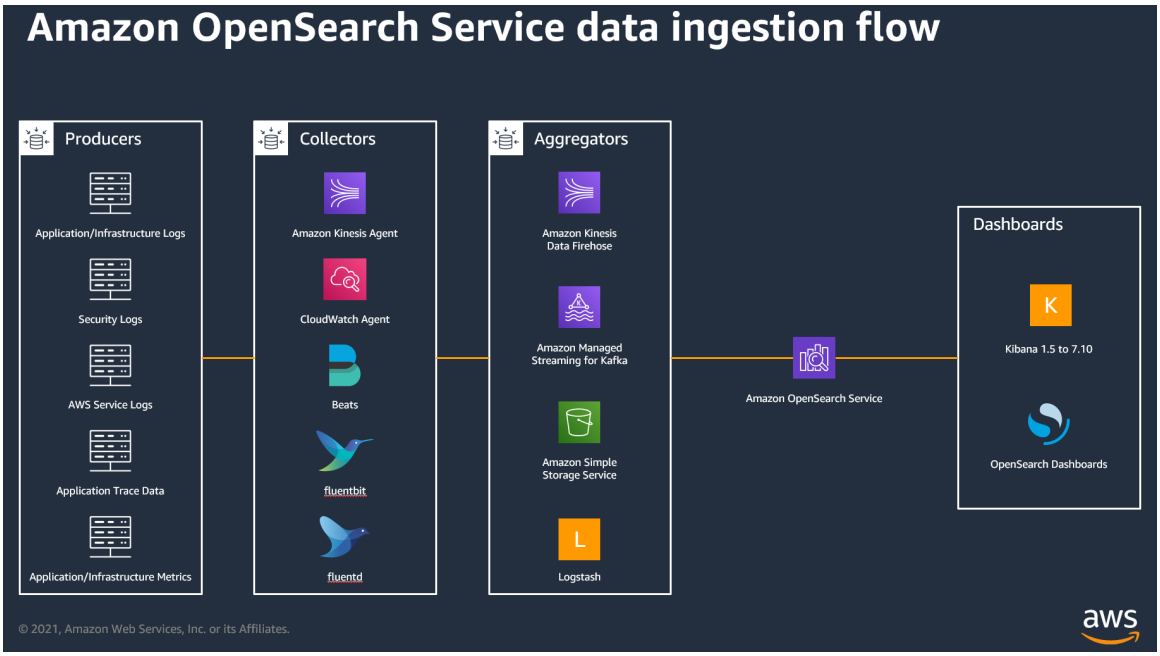
- Immutable records of system events
- Metrics, which take measurements of key attributes such as resource consumption
- Traces, which show the end-to-end journey of a process or application that consumes microservices

Observability is an emerging practice that has been embraced by DevOps practitioners to get a unified picture of what is happening in a complex environment. OpenSearch is well-suited to playing a key role in the analysis of log data that can be used for painting part of the picture.

SIEM and security analytics

SIEM systematically collects, identifies, categorizes, and analyzes incidents or events that affect cybersecurity. OpenSearch is a useful analytics tool for collecting and analyzing those events. It can work with tools or services that monitor IT infrastructure to collect log files. For instance, in the AWS cloud, Amazon OpenSearch Service can act as the frontend to Amazon GuardDuty, a threat detection service. In this scenario, Amazon GuardDuty is enabled in the AWS customer's account to monitor CloudTrail logs, virtual private cloud (VPC) flow logs, and domain name system (DNS) query logs, and it uses rules to generate findings about security-related incidents in the customer's AWS service. It delivers the event through an Amazon Kinesis Data Firehose stream to Amazon OpenSearch Service (which could run in a VPC), where the streaming data is indexed and subsequently visualized using OpenSearch.

Figure 1: Using Amazon OpenSearch Service for log analytics



Clickstream analytics

Running a website used by thousands of people results in the generation of huge torrents of logs. These logs provide the ground truth on the experiences that customers encounter while visiting a website. They show their navigation patterns and which patterns led up to customers placing an order or abandoning shopping carts. Clickstreams can also be used for monitoring the performance of a promotional campaign and can be correlated to when pages crash, showing HTTP 404 errors. The vast number of clickstream events would strain the capabilities of SQL data warehouses. By

contrast, leveraging the efficiency of the Lucene open source indexing engine (which powers OpenSearch), along with the RESTful API that interfaces to configuration, indexing, and querying, makes OpenSearch a natural solution for analyzing the clickstreams that gauge the effectiveness of a website.

Centralized logging

With its support of data collection from a wide variety of sources, including servers, virtual machines, and containers, the open source alternative to the Elasticsearch/Logstash/Kibana (ELK) stack is very popular. An open source solution is a natural collection point for storing logs, indexing them, and visualizing them in OpenSearch Dashboards in near real time. It works with a wide variety of data ingestion engines. The approach to centralized logging based on OpenSearch for collecting, analyzing, and displaying logs can span multiple AWS accounts and/or regions. The solution, which also uses OpenSearch Dashboards, works with other AWS managed services to deliver a customizable, multi-account environment to begin logging and analyzing the AWS environment and applications.

For managed services in the AWS cloud, Amazon OpenSearch Service works closely with AWS Centralized Logging, both for analyzing application-, infrastructure-, and security-related events across AWS services and for the operation of Amazon OpenSearch Service. In addition, data from Amazon CloudWatch can be streamed to Amazon OpenSearch Service to close the loop on real-time analysis.

Why OpenSearch?

High performance search engine

OpenSearch provides an open source alternative for delivering a high performance general-purpose search engine designed for performance and scale. It can deliver, within milliseconds, searches that can be performed as soon as documents are indexed. It has become highly popular as a general-purpose open source search engine because of its ability to index, search, and aggregate documents. Since it is based on search technology, OpenSearch is extremely flexible; it indexes every field, regardless of data type. Queries can then be aggregated to generate trend analysis and facilitate data exploration.

OpenSearch is extremely scalable. Developed as a community-driven, open source fork of Elasticsearch and Kibana, OpenSearch is designed for harnessing the power of scale-out clusters, and it can analyze billions of records in seconds. Its high concurrency allows OpenSearch-based solutions to handle large volumes of simultaneous queries. And, unlike proprietary solutions that charge by the volume of data and related features such as indexing, OpenSearch's open source model is extremely economical. However, customers that manually deploy and operate OpenSearch often want help scaling beyond using the tooling available from the free open source technology alone.

Part of a broader stack

OpenSearch is typically deployed with a core, distributed, JSON-based search and analytics engine that

- Supports a wide variety of searches.
- Encompasses structure, variably structured, geospatial, and metrics data.
- Provides an open source visualization engine that can display open data in a variety of forms such as histograms, line graphs, pie charts, and others.
- Leverages Logstash, a real-time data processing pipeline that ingests and transforms data streaming from multiple sources such as logs, metrics, web applications, and event hubs.

The result is a highly versatile monitoring and forensic platform that can be utilized for a variety of use cases, including but not limited to application and website search, enterprise search, log analytics, application performance monitoring, and security analytics.

Key capabilities for Amazon OpenSearch

OpenSearch expands the capabilities available under an open source Apache license by providing the following:

- **Advanced security:** Node-to-node encryption, five types of authentication, role-based access controls at multiple levels, Kibana multitenancy, and audit logging.
- **Alerting:** Log data monitoring and threshold notification.
- **Performance analysis:** Query numerous performance metrics for their cluster, aggregations, cross-cluster search, cross-domain OpenSearch searches, aggregations, and visualizations.
- **Machine learning:** Machine learning (ML) for real-time anomaly detection with the Random Cut Forest (RCF) algorithm, including notifications.

The solution: Managed services

Why managed services?

OpenSearch's scalability makes it a natural for cloud deployment where customers can take advantage of the elasticity of the cloud. However, using the core open source technology alone, scaling OpenSearch can be challenging. With *managed* cloud services, the technology provider takes a prescriptive approach in delivering packaged cloud services, simplifying operation for customers by automating installation and configuration and offering high availability and security.

How Amazon OpenSearch Service delivers time-to-value

Fully managed by AWS

Amazon OpenSearch Service provides the managed offering that allows the customer to focus strictly on the solution instead of on the tasks associated with deployment and operation that, on their own, do not add value.

As a fully managed service, Amazon OpenSearch Service simplifies the deployment, management, and scaling of OpenSearch clusters. It provides a secure environment that ensures customers are running on current production-ready releases and patches. The service provides a complete OpenSearch environment that supports production-ready open source tools and APIs. Amazon OpenSearch Service lets customers pay only for what they use. Compared to legacy tools, where you pay for items such as ingestion or indexing, Amazon OpenSearch Service is an implementation where you pay only for compute and storage.

A wide selection of supported versions

In addition to the new OpenSearch project, Amazon OpenSearch Service supports 19 versions of the original open source Elasticsearch, from 1.5 through the last 7.10 version, as well as multiple versions of OpenSearch. AWS' support includes full access to bug and security patches. OpenSearch customers *do not* have to upgrade to the latest version to get such fixes. Additionally, Amazon OpenSearch Service delivers visualization capabilities powered by OpenSearch Dashboards and Kibana (1.5 to 7.10 versions) and is fully compatible with the open source Logstash and FluentD.

Highly secure

Amazon OpenSearch Service is highly secure because security is built into the service. AWS gives customers the option to encrypt data in motion and at rest, isolate through a VPC, and enforce fine-grained access control. It has numerous certifications, including being HIPAA eligible and compliant with PCI, DSS, SOC, ISO, and FedRAMP standards. Access to management APIs for creating and scaling domains is controlled by AWS Identity and Access Management (IAM) policies.

Flexible storage options

Amazon OpenSearch Service enables customers to optimize cost and performance thanks to flexible storage options that are unique. With the UltraWarm for Amazon OpenSearch Service, Amazon offers a new option for reducing the cost of storage. It allows data to be stored and interactively analyzed using OpenSearch and OpenSearch Dashboards while reducing the cost per gigabyte by up to 90% over existing Amazon OpenSearch Service hot storage options. While hot storage is still used for indexing and providing the fastest access to data, UltraWarm complements it with less expensive, durable storage for older data that is accessed less frequently. With UltraWarm, the same interactive analysis experience is maintained.

Integrates with AWS portfolio

Amazon OpenSearch Service also provides integration with other AWS offerings. Examples include the following:

- **Streaming:** Amazon Kinesis Data Firehose or Amazon CloudWatch can be used for the native ingest of logs to Amazon OpenSearch Service. Additionally, Amazon CloudWatch can monitor a domain's performance and set alerts regarding the need for scaling.

- **Database migration:** Customers can also take advantage of the Amazon Database Migration Service (Amazon DMS) to migrate database tables to Amazon OpenSearch Service.

- **Streaming data from other AWS sources:** AWS Lambda can be used to develop programs in a serverless environment for streaming data from Amazon S3 storage, Amazon DynamoDB, or Amazon Kinesis Data Streams to Amazon OpenSearch Service for analysis. AWS recently added the option for Amazon Kinesis Data Firehose to deliver data more securely to Amazon OpenSearch Service instances that run in an Amazon VPC.

Customer examples

Observability at Pinterest

The observability team at Pinterest relies on Amazon OpenSearch Service to monitor and issue alerts for new software deployments on the main Pinterest site. In the span of a year, Amazon OpenSearch Service helped Pinterest scale its data ingestion capabilities from 500GB of data per day to 1.7TB per day. During that time, by adopting UltraWarm for Amazon OpenSearch Service, Pinterest's observability team was able to reduce the costs of its deployment, monitoring, and alerting operation by as much as 30% (with further cost reductions likely in the near future) while reducing the burden of low value work on engineers and increasing team productivity. As a result, Pinterest

can quickly and efficiently monitor and issue alerts for more than 20 new software deployments every day, helping the company deliver quality features to hundreds of millions of daily users.

SIEM at Sophos

Sophos, a worldwide leader in next-generation cybersecurity, protects its customers from today's most advanced cyberthreats. Sophos developed a large-scale security monitoring and alerting system using Amazon OpenSearch Service. Collecting machine data from a variety of sources, the system enriches it with identifiers, ingests it into Amazon OpenSearch Service to enable searching and trend analysis for anomalous events, and generates dashboards using OpenSearch Dashboards to visualize those trends. Sophos continues to use Amazon OpenSearch Service; it concludes the service is "dead simple" to deploy, requires "minimal ongoing maintenance," and is easy to upgrade.

Clickstream analytics at the *Financial Times*

The *Financial Times* uses clickstream analytics to see in real time which articles are the most popular. It is using the insights to drive editorial and daily planning decisions to better align with the interests of its readership.

Real-time monitoring and centralized logging at Autodesk

Autodesk, a leading provider of 3D design and engineering software, required a solution to ensure its customers had the best experience with its online software as a service (SaaS) solution. Autodesk adopted Amazon OpenSearch Service because it lowers the overall cost of employing logging solutions, provides better performance, and is more scalable. The solution uses Amazon OpenSearch Service for search and log analytics in conjunction with Amazon Kinesis Data Firehose for transporting log data, Amazon Kinesis Data Analytics for real-time monitoring of streaming log data, and Amazon CloudWatch for aggregated cloud-traffic metrics. It has given Autodesk vastly improved visibility, translating to better service levels for customers.

Conclusions

The explosion of machine data has made log analytics a powerful tool for assessing the health of enterprise operations in real time. Search has emerged as a powerful, unified alternative to fit-for-purpose tools for real-time tasks such as managing IT infrastructure, preventing cyberthreats, managing application performance, and conducting clickstream analysis. OpenSearch is well-suited for log analytics thanks to its efficient approach to data ingest and flexible indexing, its scalability, and its ability to deliver real-time analytics. For the fastest time-to-value, managed services are the most effective path for enterprises seeking to take advantage of OpenSearch's power, scale, and versatility. Manually setting up and operating OpenSearch clusters is time-consuming and complex. Amazon OpenSearch Service provides a managed alternative that addresses the need for enterprises to gain rapid time-to-value with a service that is fully based on open source code.

Appendix

Methodology

This report was compiled from the author's studies of log analytics platforms, open source licensing issues, and use cases reported by classic Elasticsearch customers.

Author

Tony Baer
Principal, dbInsight LLC
tony@dbinsight.io

Get in touch

www.omdia.com
askananalyst@omdia.com

Omdia consulting

Omdia is a market-leading data, research, and consulting business focused on helping digital service providers, technology companies, and enterprise decision-makers thrive in the connected digital economy. Through our global base of analysts, we offer expert analysis and strategic insight across the IT, telecoms, and media industries.

We create business advantage for our customers by providing actionable insight to support business planning, product development, and go-to-market initiatives.

Our unique combination of authoritative data, market analysis, and vertical industry expertise is designed to empower decision-making, helping our clients profit from new technologies and capitalize on evolving business models.

Omdia is part of Informa Tech, a B2B information services business serving the technology, media, and telecoms sector. The Informa group is listed on the London Stock Exchange.

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Omdia's consulting team may be able to help your company identify future trends and opportunities.

About dbInsight LLC

dbInsight provides an independent view on the database and analytics technology ecosystem. With 25+ years industry background, we have delivered insights to a market that is now being transformed by data, cloud, and AI.

Copyright notice and disclaimer

The Omdia research, data and information referenced herein (the “Omdia Materials”) are the copyrighted property of Informa Tech and its subsidiaries or affiliates (together “Informa Tech”) or its third party data providers and represent data, research, opinions, or viewpoints published by Informa Tech, and are not representations of fact.

The Omdia Materials reflect information and opinions from the original publication date and not from the date of this document. The information and opinions expressed in the Omdia Materials are subject to change without notice and Informa Tech does not have any duty or responsibility to update the Omdia Materials or this publication as a result.

Omdia Materials are delivered on an “as-is” and “as-available” basis. No representation or warranty, express or implied, is made as to the fairness, accuracy, completeness, or correctness of the information, opinions, and conclusions contained in Omdia Materials.

To the maximum extent permitted by law, Informa Tech and its affiliates, officers, directors, employees, agents, and third party data providers disclaim any liability (including, without limitation, any liability arising from fault or negligence) as to the accuracy or completeness or use of the Omdia Materials. Informa Tech will not, under any circumstance whatsoever, be liable for any trading, investment, commercial, or other decisions based on or made in reliance of the Omdia Materials.