Overview

A comprehensive log management and analysis strategy is mission critical, enabling organizations to understand the relationship between operational, security, and change management events and maintain a comprehensive understanding of their infrastructure. AWS customers have access to service-specific metrics and log files to gain insight into how each AWS service is operating, and many services capture additional data, such as API calls, configuration changes, and billing events. Log files from web servers, applications, and operating systems also provide valuable data, though in different formats, and in a random and distributed fashion. To effectively consolidate, manage, and analyze these different logs, many AWS customers choose to implement centralized logging solutions using either self-managed tools or AWS Partner Network (APN) offerings. These solutions provide a streamlined view of application, system, and AWS log information in the pursuit of operational excellence.

This document provides high-level best practices for log management as well as information and considerations for selecting a centralized logging solution using AWS services or third-party offerings. The following sections assume basic knowledge of Amazon Elastic Compute Cloud (Amazon EC2), Amazon Simple Storage Service (Amazon S3), Amazon CloudWatch (CloudWatch), Amazon Elasticsearch Service (Amazon ES), as well as a general understanding of application and system logging.

General Best Practices

When planning a centralized log management strategy, first identify business and compliance requirements, such as log monitoring and review processes, access control granularity, and log aggregation, alerting, reporting, and retention requirements. Look for a solution that is scalable and that can support new log types as you expand your use of AWS services and cloud technologies. Consider these additional best practices for implementing a log management solution:

- Define log retention requirements and lifecycle policies early on, and plan to move log files to cost-efficient storage locations as soon as practical.
- Incorporate tools and features that automate the enforcement of lifecycle policies. For example, Amazon Simple Storage Service (Amazon S3) is a cost-efficient log-storage location that includes built-in lifecycle capabilities, enabling customers to automatically retire logs to less expensive storage tiers (Amazon S3 Standard - Infrequent Access, Amazon Glacier) as necessary.
- Before implementing a custom-built solution, such as an open-source ELK stack running Elasticsearch, Logstash, and Kibana on local servers, consider the additional tasks, costs, and dependencies associated with managing and maintaining its components. Custom architectures might offer design flexibility, but managed services and tools can significantly reduce operational complexity.
- Automate the installation and configuration of log shipping agents to consistently capture system and application logs and support dynamic scaling of Amazon EC2 instances. Use Amazon EC2 user data scripts or configuration management software to perform these tasks. Alternatively, include the agent as part of the Amazon Machine Image (AMI)
- Organizations operating hybrid architectures should choose a solution that integrates with both on-premises and AWS workloads. Whether you choose to consolidate AWS and on-premises logs or to manage them separately, implement a log management solution that provides the visibility you require across all operating environments.

Application on AWS

The AWS Cloud provides flexible infrastructure and tools to support both sophisticated partner offerings and self-managed centralized-logging solutions. In general, the scale of the services to be monitored as well as an organization’s experience, budget, business requirements, and preferences for completeness and polish will determine which approach is most appropriate. The following sections describe some third-party products for log management as well as the key AWS services and open-source technologies to include in a self-managed centralized logging solution.
Third-Party Solutions

The AWS Partner Network offers a variety of comprehensive log-management solutions that can help make it easy for organizations of any size or stage of development to manage, analyze, retain, and archive logs. When selecting a third-party product, look for a solution that is easy to configure, includes a flexible method for data ingestion, and incorporates functionality for event searching, monitoring, alerting, and data visualization (e.g. real-time data dashboards). This approach may be appropriate for customers in the following situations:

- They have an existing partner tool in place for managing on-premises logs and want to extend their solution to incorporate logs from cloud resources. Organizations who already leverage popular partner technology can easily consolidate on-premises and cloud log data to visualize hybrid application performance in a single user interface.

- They have advanced alerting or reporting requirements but do not have the dedicated development and system administration resources to create or manage this capability.

- They have encryption, user management, security, or scalability requirements that Amazon ES and Kibana cannot presently meet. For example, some partner solutions are tailored to satisfy security requirements such as log-data encryption at rest or network isolation to a single VPC. These solutions can be the most efficient option for customers who manage protected personal or health information and who must comply with PCI and HIPAA standards.

Explore the AWS Marketplace for a comprehensive list of partner offerings, including popular options from Splunk, Sumo Logic, Datadog, Loggly, and Elastic.

AWS Services

Many customers choose to build their own centralized logging solution using AWS managed services. This can be a cost-effective and scalable way to help organizations meet their log-management needs. A serverless design can further reduce the overhead associated with managing individual solution components. This section introduces AWS services commonly used in log-management architectures.

Note that AWS offers a prescriptive centralized logging solution that customers can deploy in minutes using AWS CloudFormation. This automated solution uses native AWS services and open source tools to capture, consolidate, and visualize log data. For more information, see the Centralized Logging webpage on the AWS website.

Amazon Elasticsearch Service

Elasticsearch is a popular open-source search and analytics engine from Elastic that provides a quick time to value and is well supported by a vibrant open-source community. AWS offers Amazon Elasticsearch Service (Amazon ES) as a managed service that makes it easy to deploy and operate Elasticsearch in the AWS Cloud. Amazon ES manages the capacity, scaling, patching, and administration of Elasticsearch clusters while providing direct access to the Elasticsearch API. The service is integrated with CloudWatch Logs, so there are no additional requirements to write code for movement or transformation of log data.

Amazon ES provides integrated and managed access to Kibana, a data visualization plugin for Elasticsearch. Customers can create a variety of Kibana charts and dashboards for large volumes of data, and can load and use dashboards developed by the Kibana and AWS user communities. Note that Kibana does not provide native access control and must be secured with an additional mechanism, such as an Nginx web proxy (see the AWS Security Blog for detailed information). A third-party log management and visualization tool might be more appropriate for customers who cannot work within these limitations.

Amazon CloudWatch Logs

Amazon CloudWatch Logs enables customers to monitor, store, and access log files from Amazon EC2 instances, AWS CloudTrail, and other sources. Customers can retrieve log data from CloudWatch Logs using the Amazon CloudWatch console, the CloudWatch Logs commands in the AWS CLI, the CloudWatch Logs API, or the CloudWatch Logs SDK. The CloudWatch Logs agent can be easily installed and configured on Linux and Windows.
instances to send application and system log files to CloudWatch. It is best practice to use EC2 roles to grant the CloudWatch Logs agent the necessary permissions.

Amazon CloudWatch can also collect detailed system performance metrics from EC2 instances and provide those metrics to dashboards and API consumers such as Amazon Simple Notification Service and Auto Scaling triggers.

Customers can subscribe to real-time CloudWatch Logs event feeds which they can either process themselves with Amazon Kinesis and AWS Lambda, or deliver directly to Amazon ES using an AWS-provided Lambda function that connects CloudWatch Logs to Amazon ES (see Real-time Processing of Log Data with Subscriptions in the Amazon CloudWatch Logs User Guide).

Amazon Kinesis Firehose

Customers who have large amounts of log data to process can use Amazon Kinesis Firehose as a serverless log ingestion and delivery mechanism. Amazon Kinesis Firehose is a fully managed service that enables customers to deliver real-time streaming data to destinations such as Amazon ES, Amazon S3, and Amazon Redshift. Firehose is designed to handle large amounts of incoming data and can generate bulk indexing requests to an Amazon ES domain.

Unlike self-managed log processing components, such as a Logstash cluster, Firehose does not require any servers, applications, or resource management. Customers configure individual data producers to send log data to a Firehose delivery stream continuously, and Firehose manages the rest.

Amazon S3

Many organizations choose to export log data from CloudWatch Logs to Amazon S3. Amazon S3 offers customers a durable, highly scalable location to store log data and to consolidate log files for custom processing and analysis. Amazon S3 is the best choice for long-term retention and archiving of log data, especially for organizations with compliance programs that require log data to be auditable in its native format.

Once log data is in an Amazon S3 bucket, define lifecycle rules to automatically enforce retention policies and move these objects to other, cost-effective storage classes, such as Amazon S3 Standard - Infrequent Access (Standard - IA) or Amazon Glacier.

Resources

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Elasticsearch Service</td>
<td><a href="https://aws.amazon.com/elasticsearch-service/">https://aws.amazon.com/elasticsearch-service/</a></td>
</tr>
<tr>
<td>Amazon S3 Documentation</td>
<td><a href="https://aws.amazon.com/documentation/s3/">https://aws.amazon.com/documentation/s3/</a></td>
</tr>
<tr>
<td>Amazon Kinesis Firehose</td>
<td><a href="https://aws.amazon.com/kinesis/firehose/">https://aws.amazon.com/kinesis/firehose/</a></td>
</tr>
<tr>
<td>Kibana Website</td>
<td><a href="https://www.elastic.co/products/kibana">https://www.elastic.co/products/kibana</a></td>
</tr>
</tbody>
</table>