Reading 4.1: Monitoring on AWS

When operating a website like the Employee Directory Application on AWS you may have questions like:

- How many people are visiting my site day to day?

- How can I track the number of visitors over time?

- How will I know if the website is having performance or availability issues?

- What happens if my Amazon Elastic Compute Cloud (EC2) instance runs out of capacity?

- Will I be alerted if my website goes down?

You need a way to collect and analyze data about the operational health and usage of your resources. The act of collecting, analyzing, and using data to make decisions or answer questions about your IT resources and systems is called monitoring.

Monitoring enables you to have a near real-time pulse on your system and answer the questions listed above. You can use the data you collect to watch for operational issues caused by events like over-utilization of resources, application flaws, resource misconfiguration, or security-related events.

Think of the data collected through monitoring as outputs of the system, or metrics.

Use Metrics to Solve Problems

The resources that host your solutions on AWS all create various forms of data that you might be interested in collecting. You can think of each individual data point that is created by a resource as a metric. Metrics that are collected and analyzed over time become statistics, like the example of average CPU utilization over time below, showing a spike at 1:30.

Consider this: One way to evaluate the health of an Amazon EC2 instance is through CPU utilization. Generally speaking, if an EC2 instance has a high CPU utilization, it can mean a flood of requests. Or it can reflect a process that has encountered an error and is consuming too much of the CPU. When analyzing CPU utilization, take a process that exceeds a specific threshold for an unusual length of time. Use that abnormal event as a cue to either manually or automatically resolve the issue through actions like scaling the instance.

This is one example of a metric. Other examples of metrics EC2 instances have are network utilization, disk performance, memory utilization, and the logs created by the applications running on top of EC2.

Know the Different Types of Metrics

Different resources in AWS create different types of metrics. An Amazon Simple Storage Service (S3) bucket would not have CPU utilization like an EC2 instance does. Instead, S3 creates metrics related to the objects stored in a bucket like the overall size, or the number of objects in a bucket. S3 also has metrics related to the requests made to the bucket such as reading or writing objects.

Amazon Relational Database Service (RDS) creates metrics such as database connections, CPU utilization of an instance, or disk space consumption. This is not a complete list for any of the services mentioned, but you can

see how different resources create different metrics.

You could be interested in a wide variety of metrics depending on the types of resources you are using, the goals you have, or the types of questions you want answered.

Understand the Benefits of Monitoring

Monitoring gives you visibility into your resources, but the question now is, "Why is that important?" The following are some of the benefits of monitoring.

Respond to operational issues proactively before your end users are aware of them. It's a bad practice to wait for end users to let you know your application is experiencing an outage. Through monitoring, you can keep tabs on metrics like error response rate or request latency, over time, that help signal that an outage is going to occur. This enables you to automatically or manually perform actions to prevent the outage from happening— fixing the problem before your end users are aware of it.

Improve the performance and reliability of your resources. Monitoring the different resources that comprise your application provides you with a full picture of how your solution behaves as a system. Monitoring, if done well, can illuminate bottlenecks and inefficient architectures. This enables you to drive performance and reliability improvement processes.

Recognize security threats and events. When you monitor resources, events, and systems over time, you create what is called a baseline. A baseline defines what activity is normal. Using a baseline, you can spot anomalies like unusual traffic spikes or unusual IP addresses accessing your resources. When an anomaly occurs, an alert can be sent out or an action can be taken to investigate the event.

Make data-driven decisions for your business. Monitoring is not only to keep an eye on IT operational health. It also helps drive business decisions. For example, let's say you launched a new feature for your cat photo app, and want to know whether it's being used. You can collect application-level metrics and view the number of users who use the new feature. With your findings, you decide whether to invest more time into improving the new feature.

Create more cost-effective solutions. Through monitoring, you can view resources that are being underutilized and rightsize your resources to your usage. This helps you optimize cost and make sure you aren't spending more money than necessary.

Enable Visibility

AWS resources create data you can monitor through metrics, logs, network traffic, events, and more. This data is coming from components that are distributed in nature, which can lead to difficulty in collecting the data you need if you don't have a centralized place to review it all. AWS has already done that for you with a service called Amazon CloudWatch.

Amazon CloudWatch is a monitoring and observability service that collects data like those mentioned in this module. CloudWatch provides actionable insights into your applications, and enables you to respond to system-wide performance changes, optimize resource utilization, and get a unified view of operational health. This unified view is important.

You can use CloudWatch to:

- Detect anomalous behavior in your environments.

- Set alarms to alert you when something's not right.

- Visualize logs and metrics with the AWS Management Console.

- Take automated actions like scaling.

- Troubleshoot issues.

- Discover insights to keep your applications healthy.

Resources

- *External Site:* AWS: Amazon CloudWatch