Monitoring and Optimization

Module 6

Units

6.1 Explore AWS CloudWatch (Monitoring)
6.2 AWS Load Balancer and Auto Scaling
6.3 Object Storage with Amazon S3

OVERVIEW : MODULE 06

Monitoring and Optimization

Learning Outcomes

- Gain an understanding of AWS CloudWatch as a monitoring service in the AWS ecosystem
- Explore CloudWatch features for monitoring resources and applications
- Understand the role of AWS Load Balancer in distributing incoming application traffic
- Explore Auto Scaling to automatically adjust the number of instances based on demand
- Understand how to leverage CloudWatch metrics for monitoring S3 buckets

6.1

Explore AWS CloudWatch (Monitoring)

Monitoring gives you insights into your applications that help you detect, investigate, and remediate problems faster.

Explore AWS CloudWatch (Monitoring)

Purpose of Monitoring

When we need to monitor ?

When operating a website like an employee directory application on AWS, you might have questions like the following:

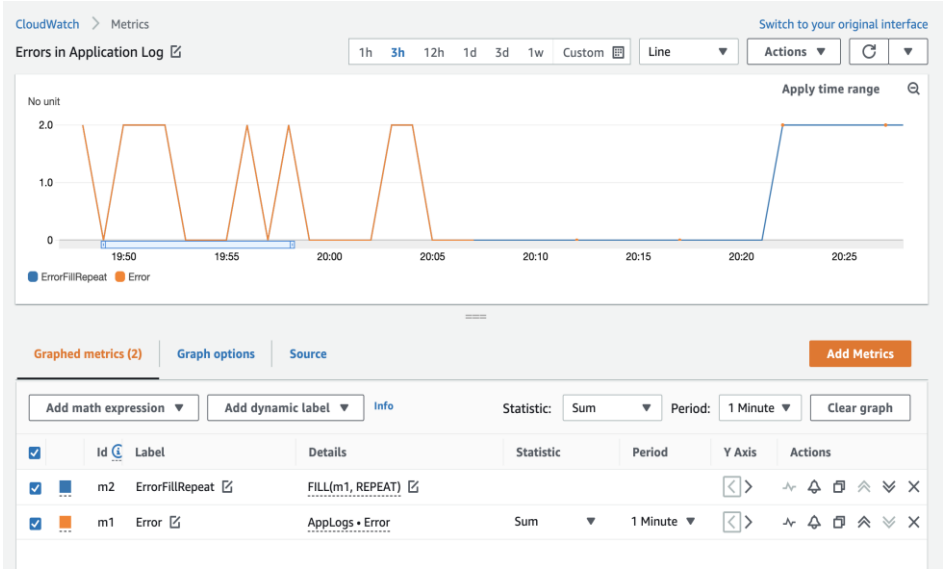
- How many people are visiting my site day to day?
- How can I track the number of visitors over time?
- How will I know if the website is having performance or availability issues?
- What happens if my Amazon Elastic Compute Cloud (Amazon EC2) instance runs out of capacity?

Explore AWS CloudWatch (Monitoring)

Use Metrics to Solve Problems

USE Metrices

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



Explore AWS CloudWatch (Monitoring)

Type of Metrices

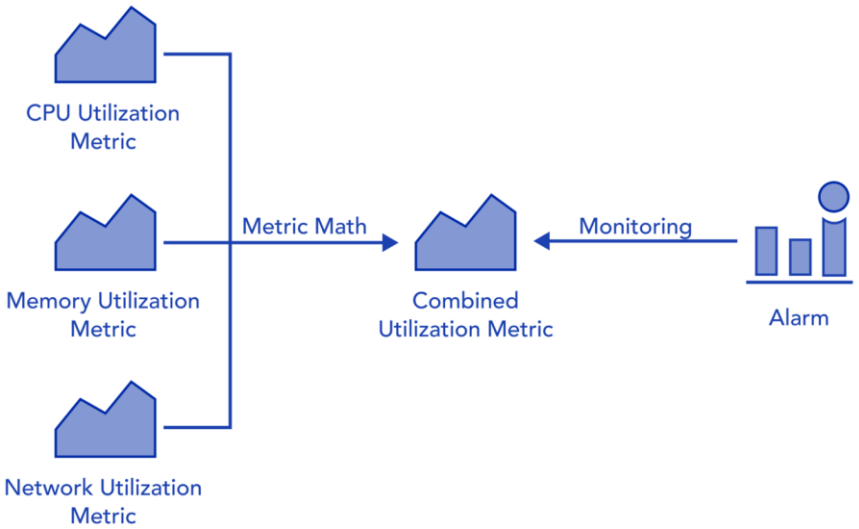
EC2 Metrices

Different resources in AWS create different types of metrics.

- CPU utilization
- Network utilization
- Disk performance
- Status checks

S3 Metrices

- Size of objects stored in a bucket
- Number of objects stored in a bucket
- Number of HTTP request made to a bucket



Explore AWS CloudWatch (Monitoring)

SUMMARY

- ✓ The importance of Monitoring
- ✓ How we can monitor our systems
- ✓ Services specific AWS CloudWatch metrices

Explore AWS CloudWatch (Monitoring)

Resources

- ✓ https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/monitoring_best_practices.html
- ✓ https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/monitoring_ec2.html
- ✓ https://aws.amazon.com/cloudwatch/
- ✓ https://docs.aws.amazon.com/AmazonCloudWatch/latest/monitoring/CloudWatch-Metric-Streams.html

Lesson Learning Outcomes

- ✓ Develop a foundational understanding of load balancing concepts and how AWS Load Balancers distribute incoming traffic across multiple instances to ensure high availability and fault tolerance
- ✓ Differentiate between the various types of AWS Load Balancers, including Application Load Balancer (ALB), Network Load Balancer (NLB), and Classic Load Balancer, and understand their use cases and features
- ✓ Learn how to configure and set up AWS Load Balancers, including defining listeners, configuring health checks, and managing routing policies to optimize traffic distribution

LESSON OVERVIEW

MODULE 6 MONITORING AND OPTIMIZATION

Lesson 6.2 AWS Load Balancer and Auto Scaling

- Traffic Routing with Elastic Load Balancing
- Application Load Balancer
- Network Load Balancer
- AWS Auto Scaling

6.2

AWS Load Balancer and Auto Scaling

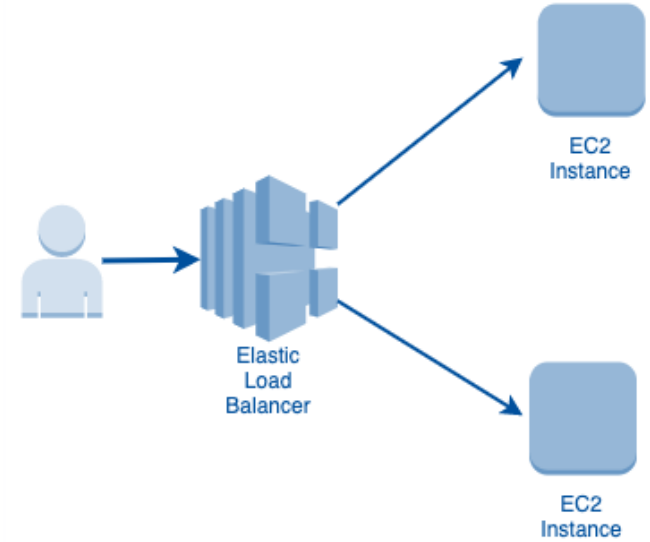
The Elastic Load Balancing (ELB) service can distribute incoming application traffic across EC2 instances, containers, IP addresses, and Lambda functions. faster

AWS Load Balancer and Auto Scaling

Traffic Routing with Elastic Load Balancing

ELB Features

The Elastic Load Balancing refers to the process of distributing tasks across a set of resources. The resources are EC2 instances that host the application, and the tasks are the requests being sent.



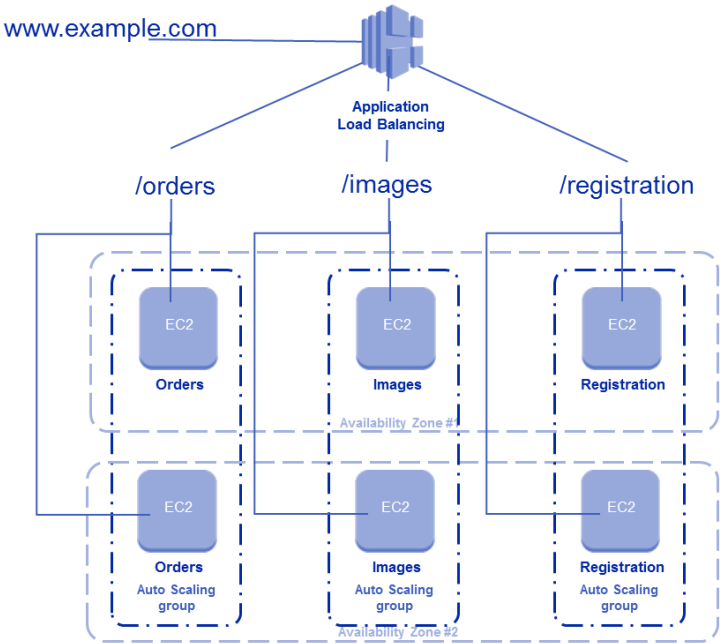
AWS Load Balancer and Auto Scaling

Application Load Balancer

Application Load Balancer serves as the single point of contact for clients. The load balancer distributes incoming application traffic across multiple targets, such as –

- EC2 instances
- In multiple Availability Zones

It functions at **Layer 7** of the **OSI model.**

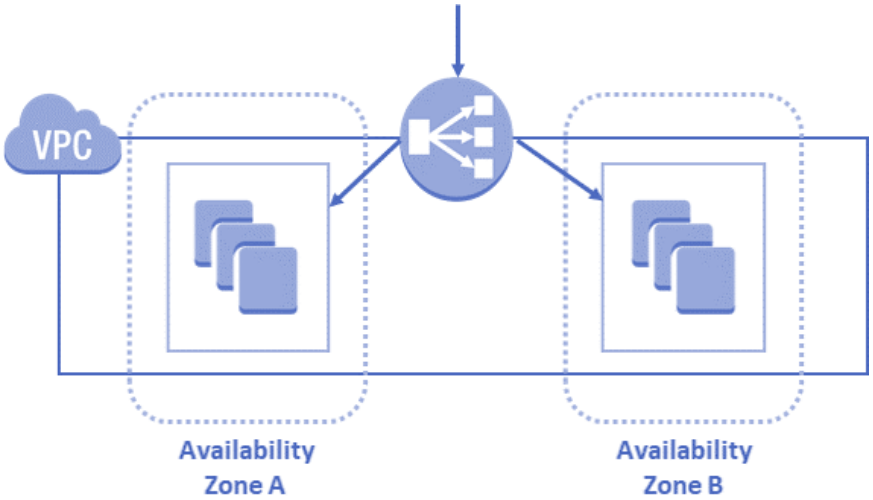


AWS Load Balancer and Auto Scaling

Network Load Balancer

A **Network Load Balancer** is ideal for load balancing TCP and UDP traffic. It functions at Layer 4 of the OSI model, routing connections from a target in the target group based on IP protocol data

- TCP and User Datagram Protocol (UDP) connection based
- Source IP preservation
- Low latency

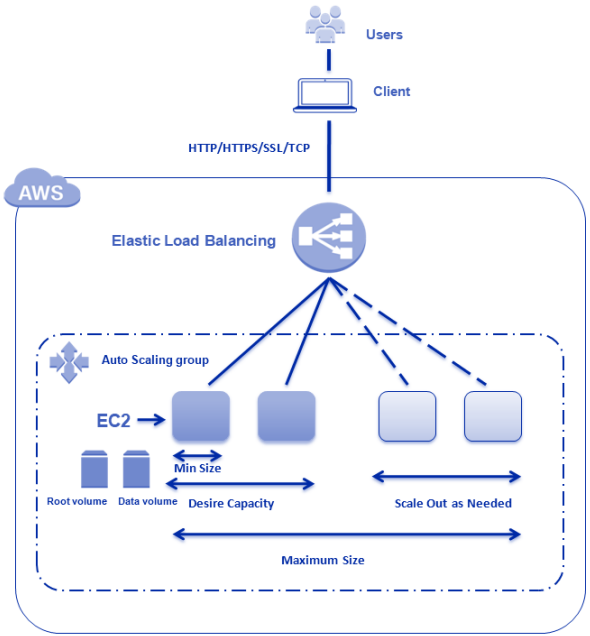


AWS Load Balancer and Auto Scaling

AWS Auto Scaling

Horizontal Scaling

Amazon EC2 Auto Scaling helps you maintain application availability. You can automatically add or remove EC2 instances using scaling policies that you define.



AWS Load Balancer and Auto Scaling

SUMMARY

- ✓ Understanding the AWS Elastic Load Belcher
- ✓ Discuss different type of Elastic Load Belcher
- ✓ How to handle unpredictable traffic

AWS Load Balancer and Auto Scaling

Resources

- ✓ https://aws.amazon.com/elasticloadbalancing/
- ✓ https://docs.aws.amazon.com/AmazonECS/latest/developerguide/load-balancer-types.html
- ✓ https://docs.aws.amazon.com/AmazonECS/latest/developerguide/service-load-balancing.html
- ✓ https://aws.amazon.com/what-is/load-balancing/
- ✓ https://aws.amazon.com/autoscaling/
- ✓ https://docs.aws.amazon.com/AmazonECS/latest/developerguide/service-auto-scaling.html

6.3

Hands-On Lab

Configure High Availability for Your Application