

Advanced AWS Workshop



Instructor: Govind Kumar

Date: 19-May-2025 | Time: 9:00 PM - 11:30 PM IST



Load Balancers

Learn about different types of AWS load balancers and their use cases.



Auto Scaling

Master AWS Auto Scaling to automatically adjust capacity based on demand.



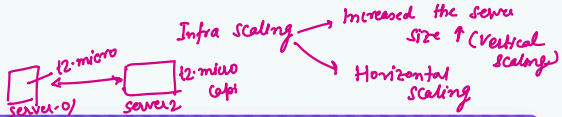
High Availability





Design resilient architectures using load balancing and auto scaling.

Load Balancers in AWS

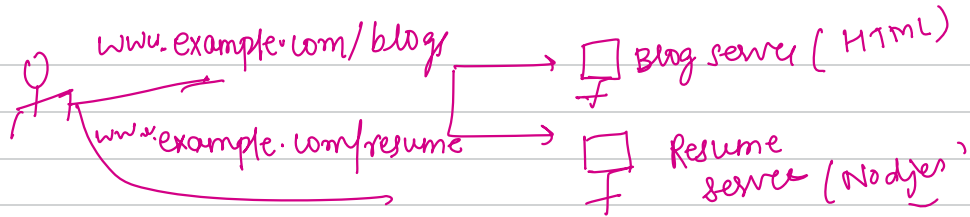
Comparison of AWS Load Balancer Types

AWS offers several types of load balancers to distribute traffic:

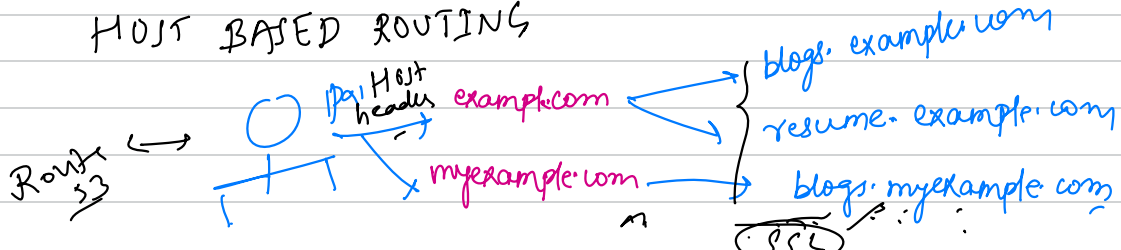


Load Balancer	Network Layer	Key Features
 Application Load Balancer	Layer 7 (HTTP/HTTPS)	<div>✓ Path-based Routing</div> <div>Host-based Routing</div> <div>WebSockets</div> <div>HTTP/2</div> <div>WAF Integration</div>
 Network Load Balancer	Layer 4 (TCP/UDP/TLS)	<div>Ultra-low Latency</div> <div>Millions of RPS</div> <div>Source IP Preservation</div> <div>Static IP</div> <div>Elastic IP</div>
 Gateway Load Balancer	Layer 3 (IP protocol)	<div>Virtual Appliances</div> <div>Transparent Inspection</div> <div>GENEVE Protocol</div> <div>Security Appliances</div> <div>Firewalls & IDS/IPS</div>
 Classic Load Balancer	Layer 4 & 7 (Legacy)	<div>EC2-Classic</div> <div>Legacy Support</div> <div>Basic Features</div> <div>Not Recommended</div>

PATH BASED ROUTING







HOST BASED ROUTING



Load Balancers in AWS

Comparison of AWS Load Balancer Types

AWS offers several types of load balancers to distribute traffic:

Load Balancer	Network Layer	Key Features
 Application Load Balancer	Layer 7 (HTTP/HTTPS)	<div>Path-based Routing</div> <div>Host-based Routing</div> <div>WebSockets</div> <div>HTTP/2</div> <div>WAF Integration</div>
 Network Load Balancer	Layer 4 (TCP/UDP/TLS)	<div>Ultra-low Latency</div> <div>Millions of RPS</div> <div>Source IP Preservation</div> <div>Static IP</div> <div>Elastic IP</div>
 Gateway Load Balancer	Layer 3 (IP protocol)	<div>Virtual Appliances</div> <div>Transparent Inspection</div> <div>GENEVE Protocol</div> <div>Security Appliances</div> <div>Firewalls & IDS/IPS</div>
 Classic Load Balancer	Layer 4 & 7 (Legacy)	<div>EC2-Classic</div> <div>Legacy Support</div> <div>Basic Features</div> <div>Not Recommended</div>

When to Use Each Load Balancer Type:



Use Application Load Balancer When:

- ✓ You need HTTP/HTTPS load balancing
- ✓ You're running containerized applications
- ✓ You need path-based routing
- ✓ You're building microservices architectures

Example: Web applications, API services, containerized microservices



Use Network Load Balancer When:

- ✓ You need extremely high performance and low latency
- ✓ You need static IP addresses for your load balancer
- ✓ You need to preserve client IP addresses
- ✓ You're handling non-HTTP protocols

Example: TCP/UDP applications, gaming servers, IoT applications, financial trading



Use Gateway Load Balancer When:

- ✓ You need to deploy, scale, and manage virtual appliances
- ✓ You need transparent network traffic inspection
- ✓ You need to implement security controls across VPCs

Example: Firewall deployments, intrusion detection systems, deep packet inspection

Choosing the right load balancer depends on your application architecture, traffic patterns, and performance requirements. For modern web applications, ALB is often the best choice. For high-performance TCP/UDP applications, use NLB. For security appliances, GWLB is the appropriate choice.



Load Balancer Selection Guide

Use Cases

Key Features

Decision Tree



Application Load Balancer

- ✓ Web applications
- ✓ Microservices
- ✓ Container-based apps
- ✓ API services



Network Load Balancer

- ✓ TCP/UDP applications
- ✓ Gaming servers
- ✓ IoT applications
- ✓ Financial trading



Gateway Load Balancer

- ✓ Firewall deployments
- ✓ Intrusion detection
- ✓ Deep packet inspection
- ✓ Security appliances



Load Balancer Selection Guide

Use Cases

Key Features

Decision Tree

Feature	ALB	NLB	GWLB
Protocol	HTTP/HTTPS	TCP/UDP/TLS	IP (Layer 3)
Static IP	×	✓	✓
Path Routing	✓	×	×
Performance	Good	Ultra-high	High
WAF Integration	✓	×	×



Load Balancer Selection Guide

Use Cases

Key Features

Decision Tree

HTTP/HTTPS traffic?

Yes



Use ALB

No

Need security inspection?

Yes

 Use
GWLB

No

 Use NLB

Core Components and Functionality

AWS Auto Scaling provides a comprehensive set of components and features that work together to create a dynamic, responsive infrastructure. Explore the key elements below to understand how Auto Scaling adapts to your application's needs.

Key Components of Auto Scaling

Explore the essential building blocks that make Auto Scaling work

Launch Templates

Auto Scaling Groups

Scaling Policies

Advanced Features

Launch Templates

Launch Templates define the EC2 instances that Auto Scaling will launch, including all the configuration details needed to create an instance.



Versioning

Create and manage multiple versions of launch configurations



Parameter Subsets

Specify a subset of parameters when launching an instance



Flexibility

Supports multiple instance types and purchase options

```
aws ec2 create-launch-template \  
  --launch-template-name "web-template" \  
  --version-description "Initial version" \  
  --launch-template-data '{  
    "ImageId": "ami-0abcdef1234567890",  
    "InstanceType": "t2.micro",  
    "SecurityGroupIds": ["sg-0123456789abcdef0"]  
  }'
```

Real-World Implementation Examples



E-Commerce



Media Streaming



SaaS Application

Media Streaming Service

High Performance



Viewers



Network Load Balancer



Streaming Servers ASG

AZ1

AZ2



Scheduled Scaling



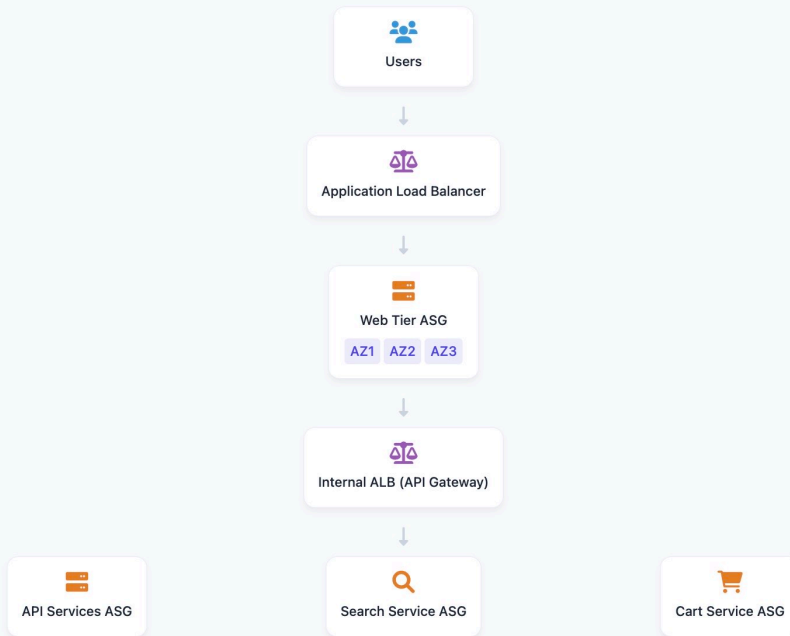
Predictive Scaling



Low Latency

E-Commerce Platform Architecture

High Availability



 WAF Protection

 Path-Based Routing

 Target Tracking Scaling



E-Commerce



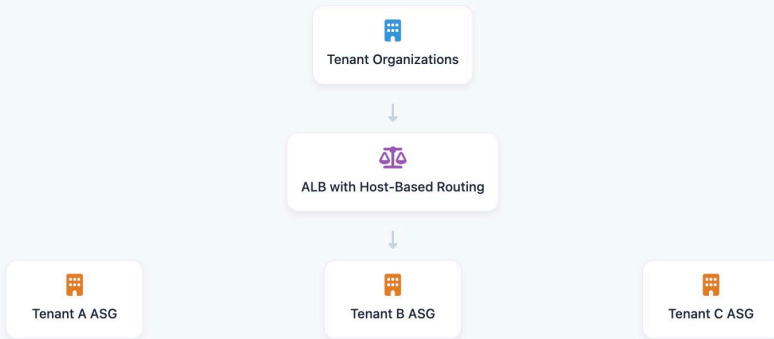
Media Streaming



SaaS Application

SaaS Multi-Tenant Architecture

Cost Optimized



\$ Mixed Instance Types

🕒 Response Time Scaling

🔒 Tenant Isolation

Best Practices and Design Patterns



Load Balancer



Use Multiple AZs

Deploy your load balancer and targets across multiple Availability Zones for high availability.

High Availability

Fault Tolerance



Configure Health Checks

Use application-specific health checks that verify actual functionality, not just TCP connections.

Reliability

Accuracy



Enable Connection Draining

Allow in-flight requests to complete when instances are deregistered from the load balancer.

User Experience

Request Completion



Use Security Groups

Restrict traffic to your load balancer and from your load balancer to instances.

Security

Access Control



Enable Access Logs

Capture detailed information about requests for troubleshooting and analysis.

Troubleshooting

Auditing



Auto Scaling



Architecture Patterns