

AWS ELASTIC LOAD BALANCER (ELB) AND AUTO SCALING

Elastic Load Balancer (ELB)

Elastic Load Balancer



Elastic Load
Balancing

- **Distributes** traffic across multiple EC2 instances, in multiple Availability Zones
- Supports **health checks** to detect unhealthy Amazon EC2 instances
- Supports the **routing and load balancing** of HTTP, HTTPS, SSL, and TCP traffic to Amazon EC2 instances

Elastic Load Balancer Types

Classic Load Balancer

PREVIOUS GENERATION
for HTTP, HTTPS, and TCP

Choose a Classic Load Balancer when you have an existing application running in the EC2-Classical network.

Application Load Balancer

HTTP
HTTPS

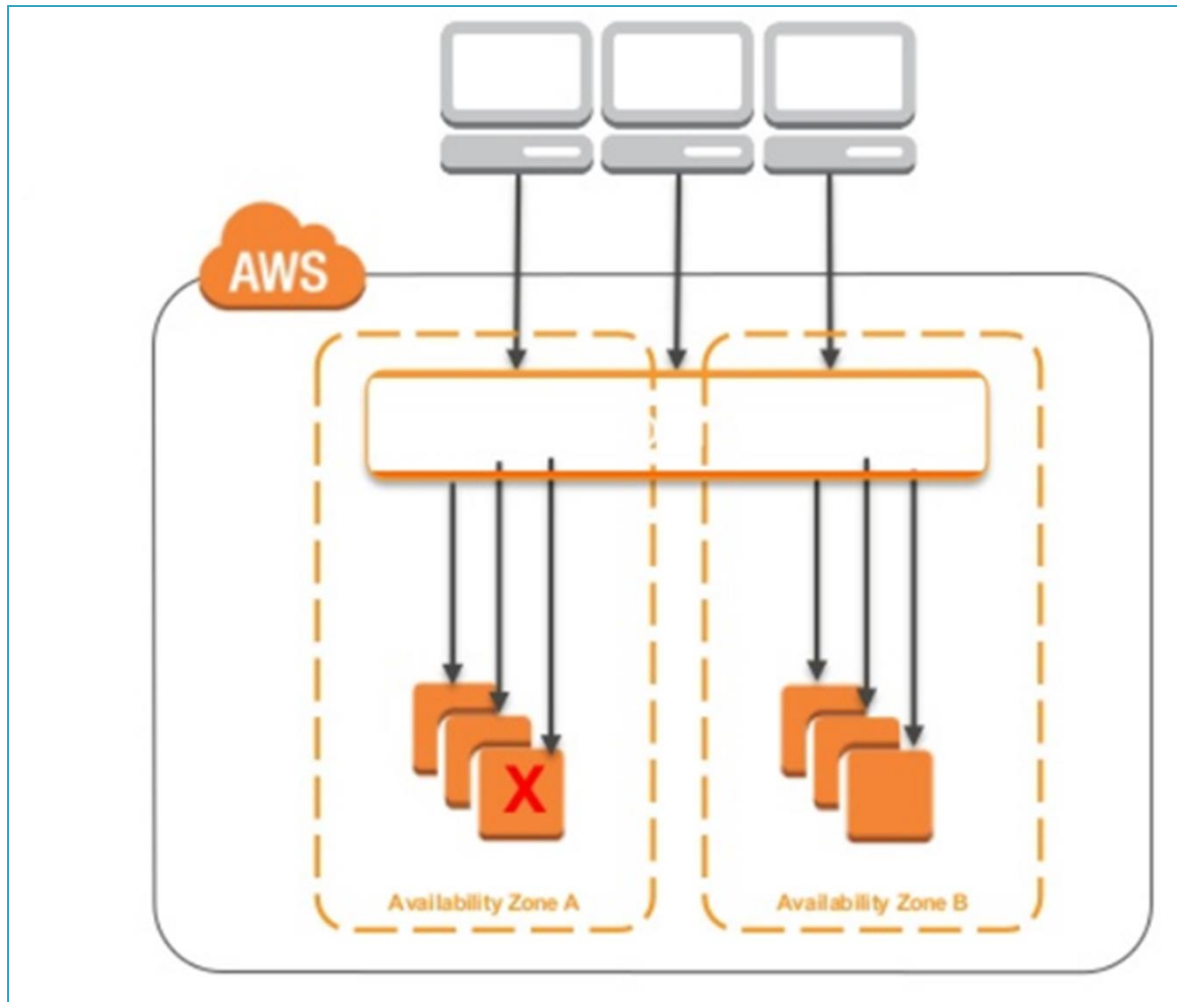
Choose an Application Load Balancer when you need a flexible feature set for your web applications with HTTP and HTTPS traffic. Operating at the request level, Application Load Balancers provide advanced routing and visibility features targeted at application architectures, including microservices and containers.

Network Load Balancer

TCP
TLS

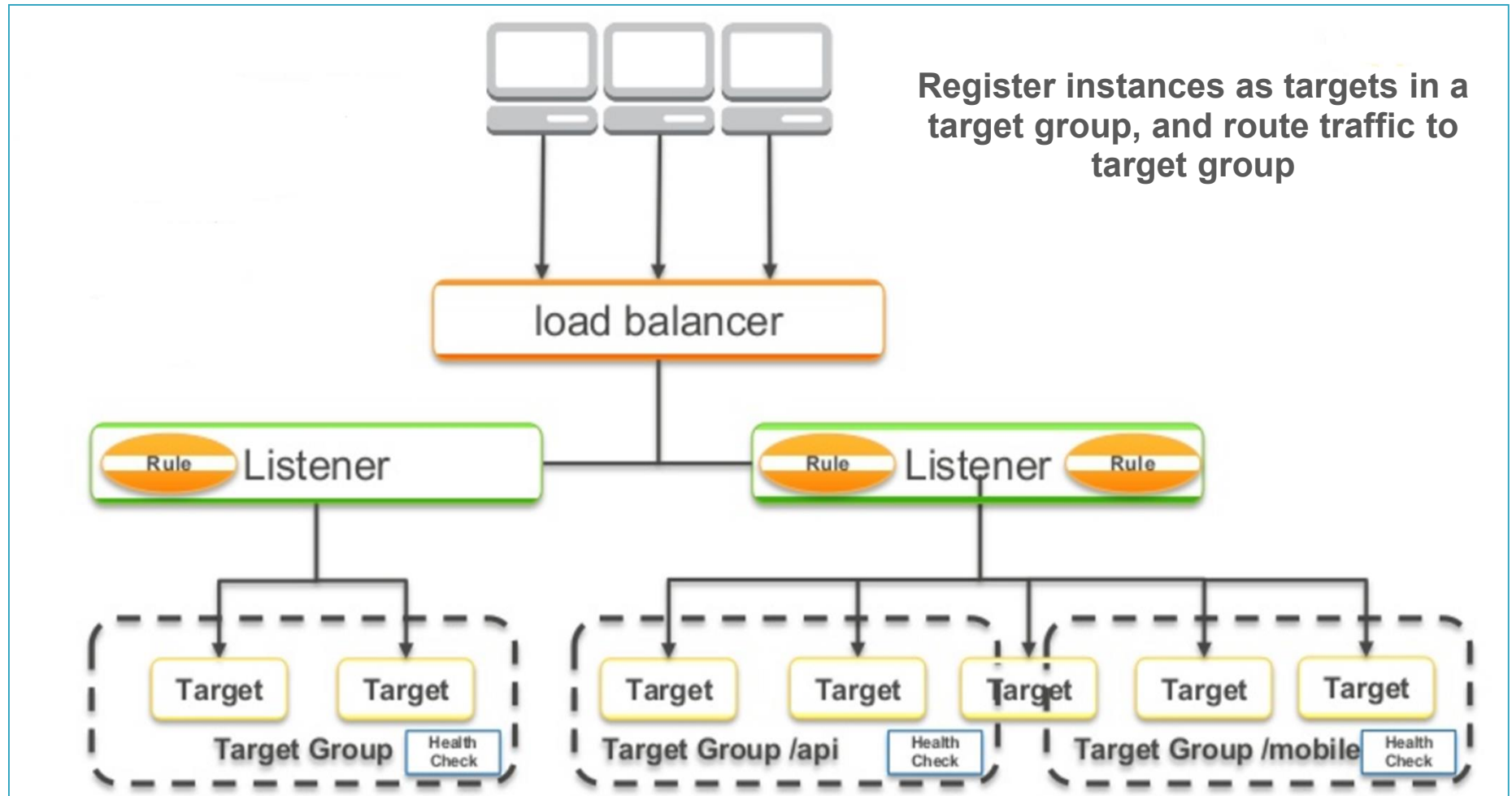
Choose a Network Load Balancer when you need ultra-high performance, the ability to terminate TLS connections at scale, centralize certificate deployment, and static IP addresses for your application. Operating at the connection level, Network Load Balancers are capable of handling millions of requests per second securely while maintaining ultra-low latencies.

Classic Load Balancer



Register instance with your load balancer

Application Load Balancer



Classic vs Application Load Balancer

Classic Load Balancer

benefits include support for:

- EC2-Classic.
- VPC.
- TCP and SSL listeners.
- Sticky sessions.
- OSI Layer 4
(network protocol level)

ALB benefits include support for:

- Path-based routing.
- Routing requests to multiple services on a single EC2 instance.
- Containerized applications.
- Monitoring the health of each service independently.
- OSI Layer 7
(application level)

Auto Scaling

Auto Scaling



Auto
Scaling

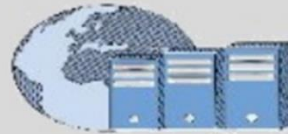
- **Scale** your Amazon EC2 capacity **automatically**
- Well-suited for applications that experience **variability in usage**
- Available at no additional charge

Auto Scaling Benefits

**Better Fault
Tolerance**



**Better
Availability**

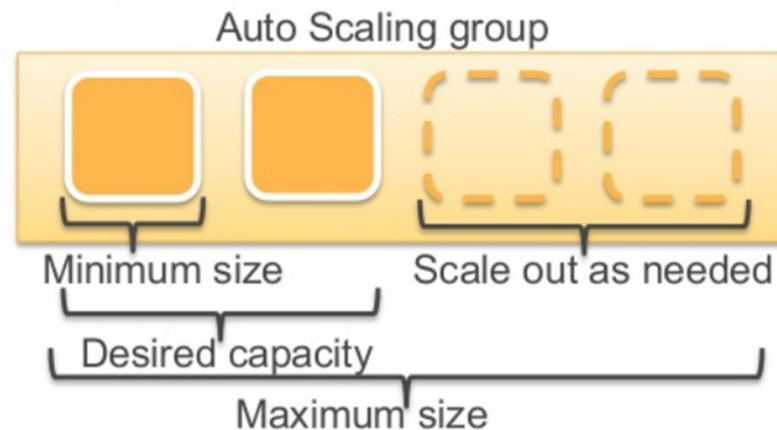


**Better Cost
Management**



Auto Scaling Groups

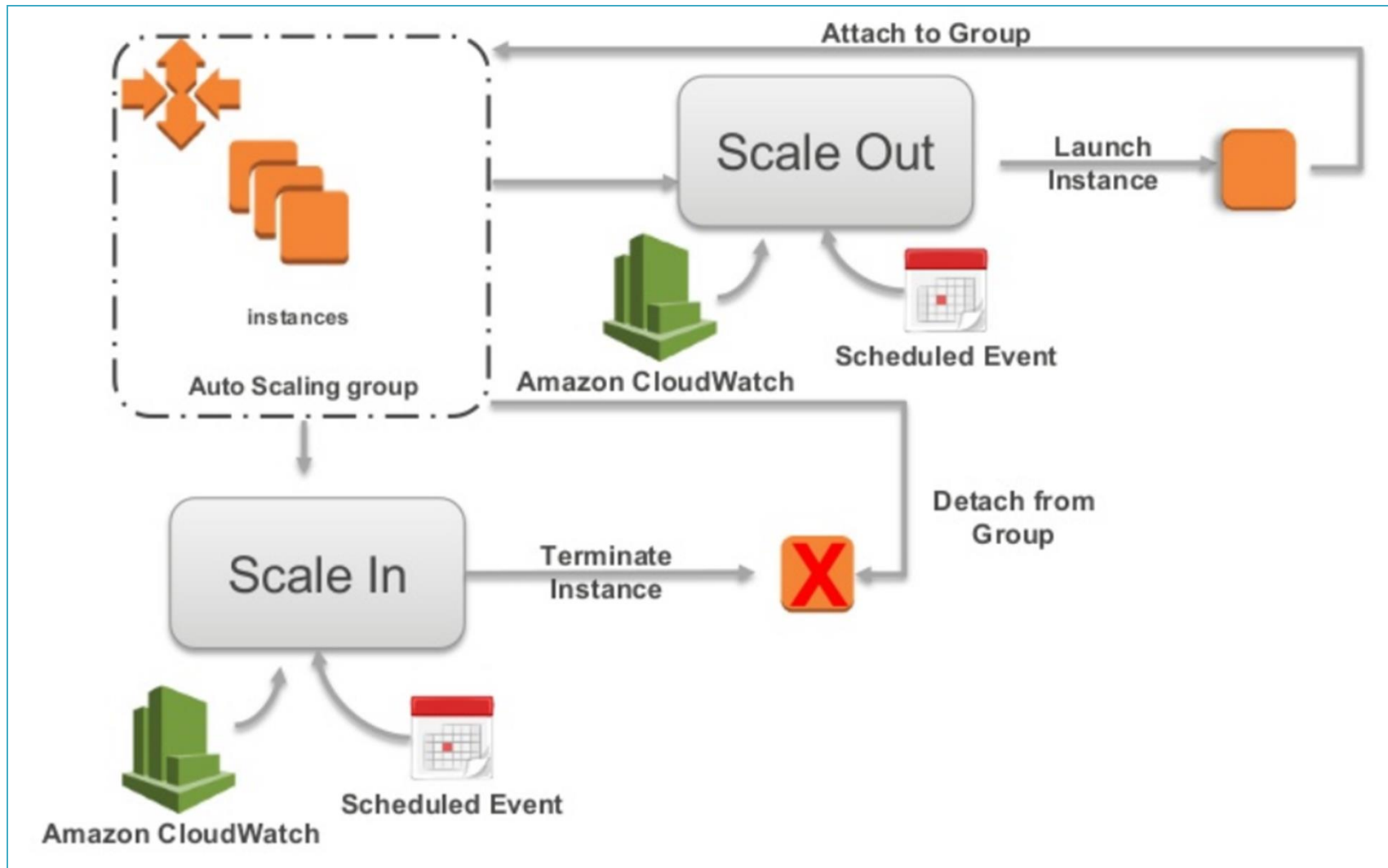
- Contain a collection of EC2 instances that share similar characteristics.
- Instances in an Auto Scaling group are treated as a **logical grouping** for the purpose of instance scaling and management.



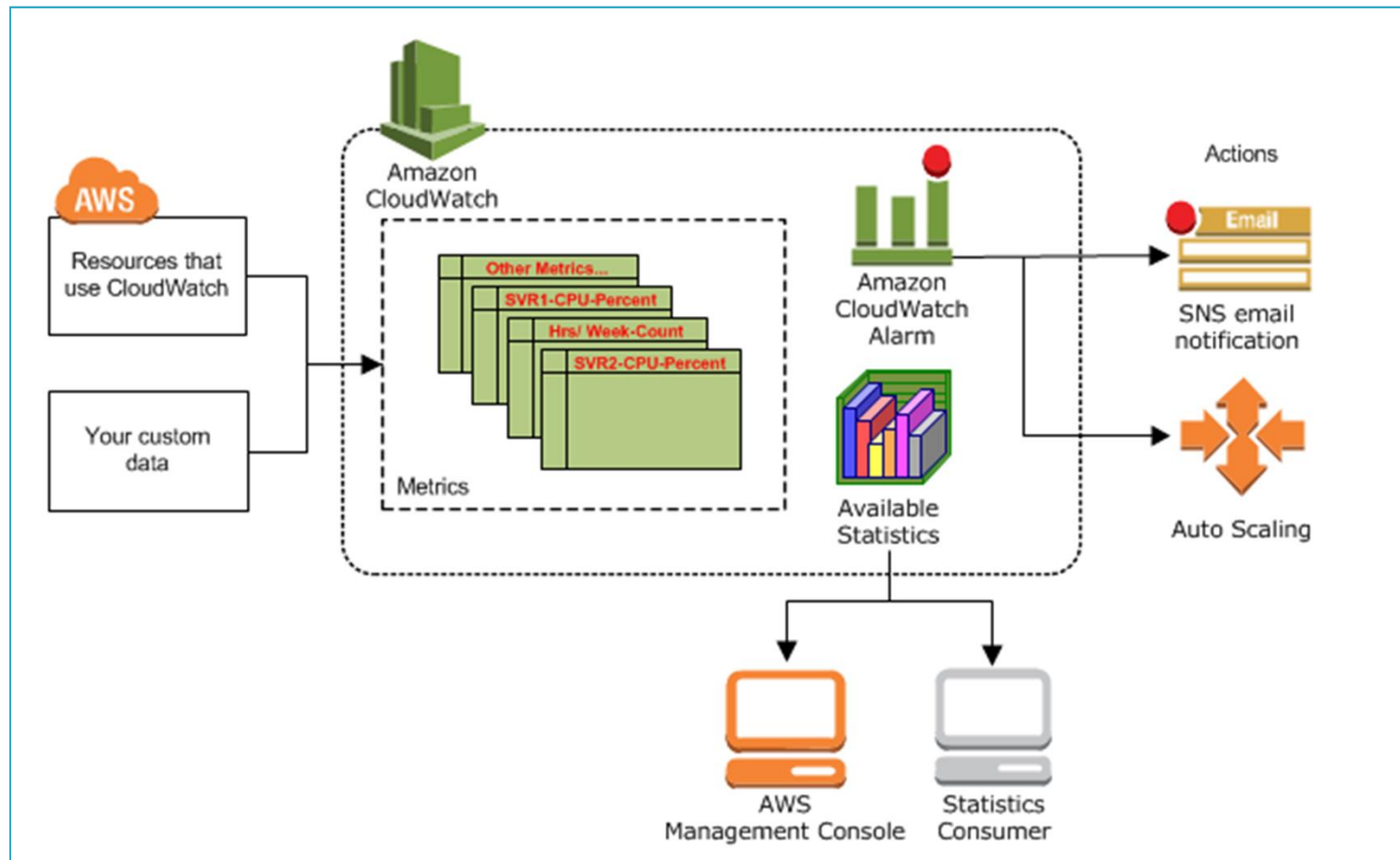
Dynamic Scaling

- You can create a scaling policy that uses **CloudWatch alarms** to determine:
 - When your Auto Scaling group should **scale out**.
 - When your Auto Scaling group should **scale in**.
- You can use alarms to monitor:
 - Any of the metrics that AWS services send to Amazon CloudWatch.
 - Your own **custom metrics**.

Auto Scaling Lifecycle



Cloud Watch



Thank You!