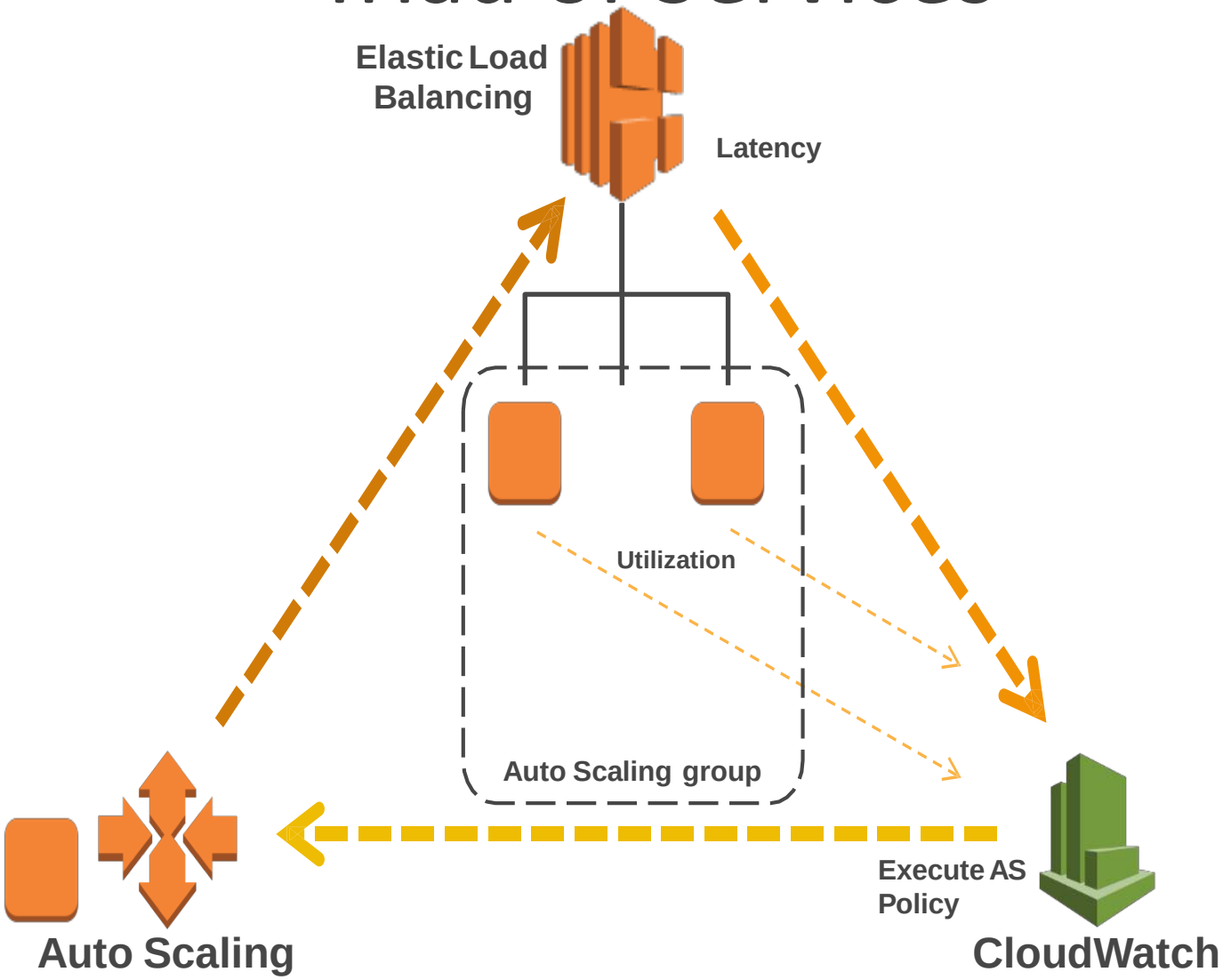
AWS Elasticity and Management

Triad of Services



AWS CloudWatch

- Basic monitoring (7 metrics, 5min)
- Detailed monitoring (10 alarms, 1 million API requests, 1min)
- Set alarms and alerts
- Notification via SES, SNS
- Custom Monitoring through API
- Integrate with Auto Scaling
- Mobile app for basic monitoring and management

AWS Autoscaling

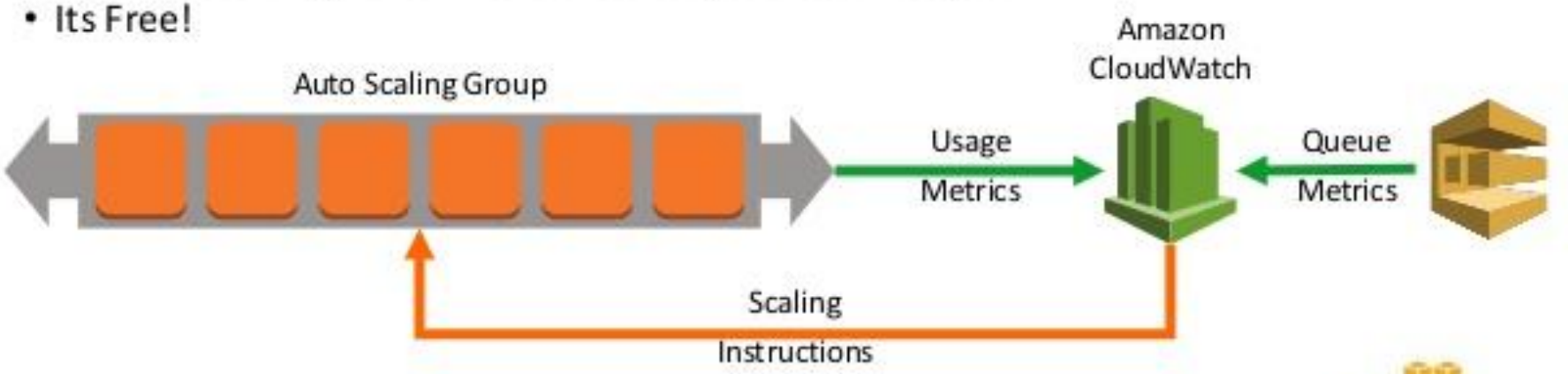
What is Auto Scaling

- Auto Scaling helps you maintain application availability and allows you to scale your Amazon EC2 capacity up or down automatically according to conditions you define. You can use Auto Scaling to help ensure that you are running your desired number of Amazon EC2 instances.

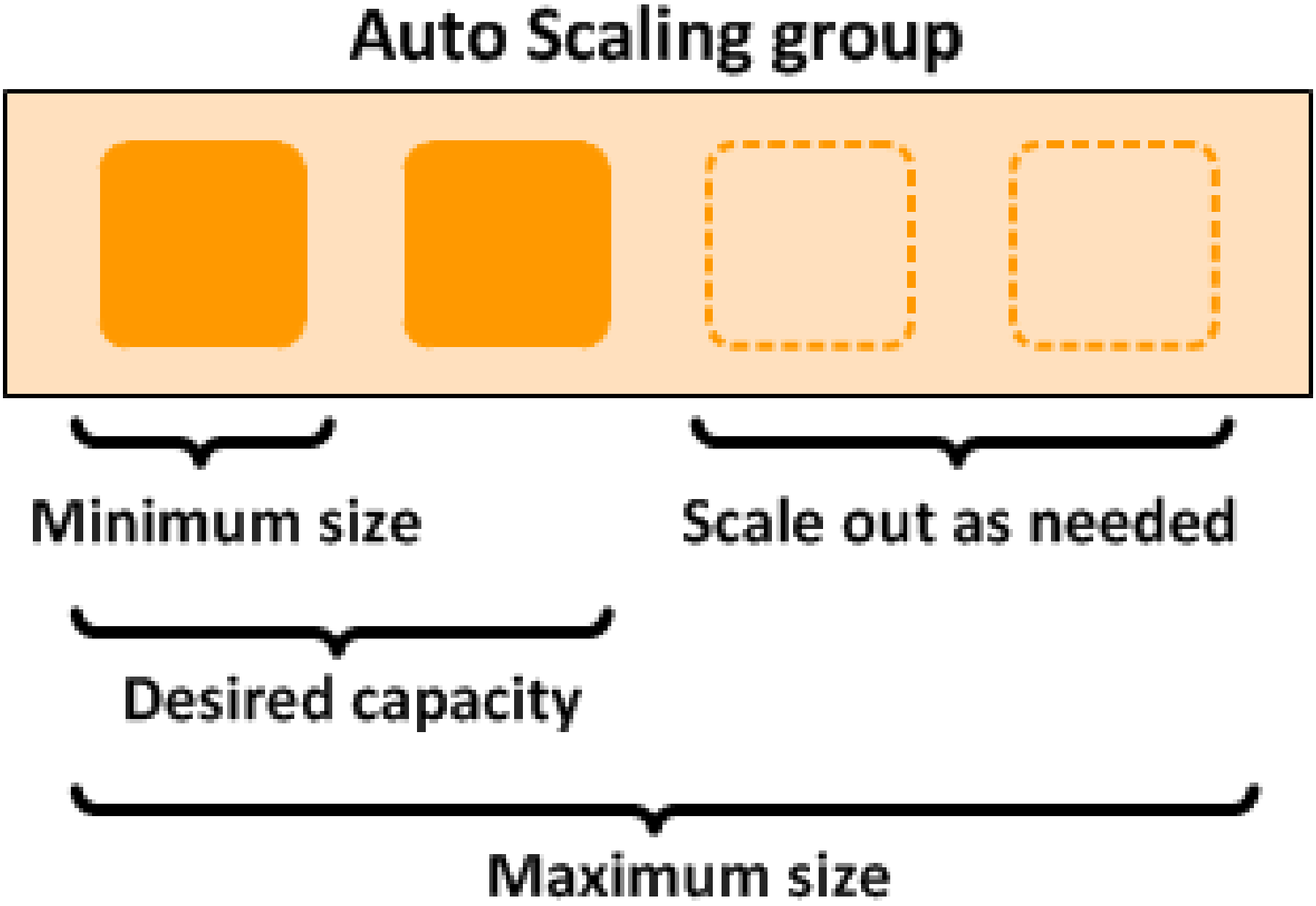
- Auto Scaling can also automatically increase the number of Amazon EC2 instances during demand spikes to maintain performance and decrease capacity during lulls to reduce costs. Auto Scaling is well suited both to applications that have stable demand patterns or that experience hourly, daily, or weekly variability in usage.

Auto Scaling

- Automatic resizing of compute clusters based on demand
- Define minimum and maximum number of instances
- Define when scaling out and in occurs
- Use metrics collected in Amazon CloudWatch to drive scaling
- Run Auto Scaling for On-Demand and Spot instance types
- Its Free!



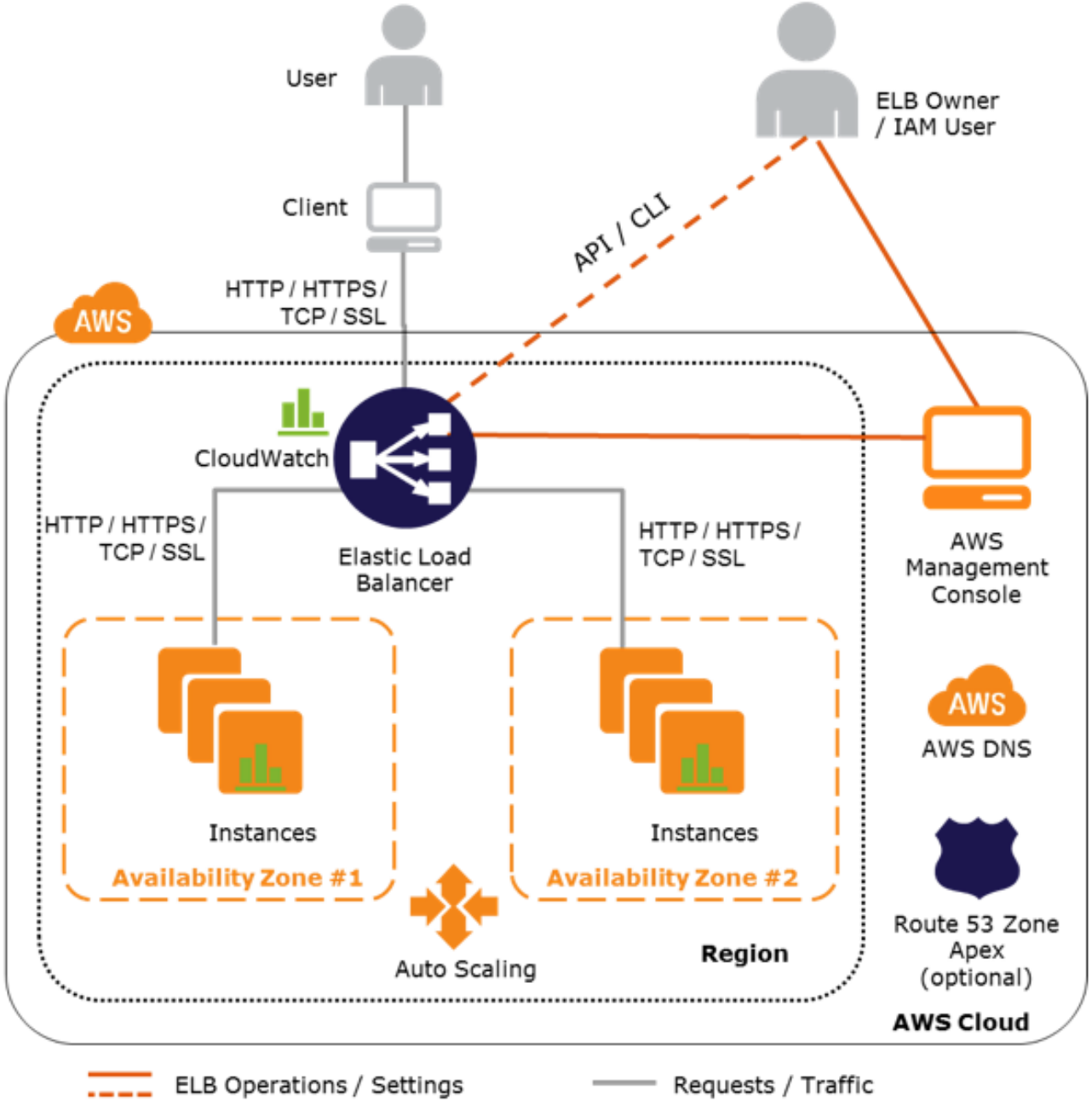
amazon web services



AWS Load Balancer

AWS Load Balancer

Elastic Load Balancing automatically distributes incoming application traffic across multiple Amazon EC2 instances. It enables you to achieve fault tolerance in your applications, seamlessly providing the required amount of load balancing capacity needed to route application traffic.



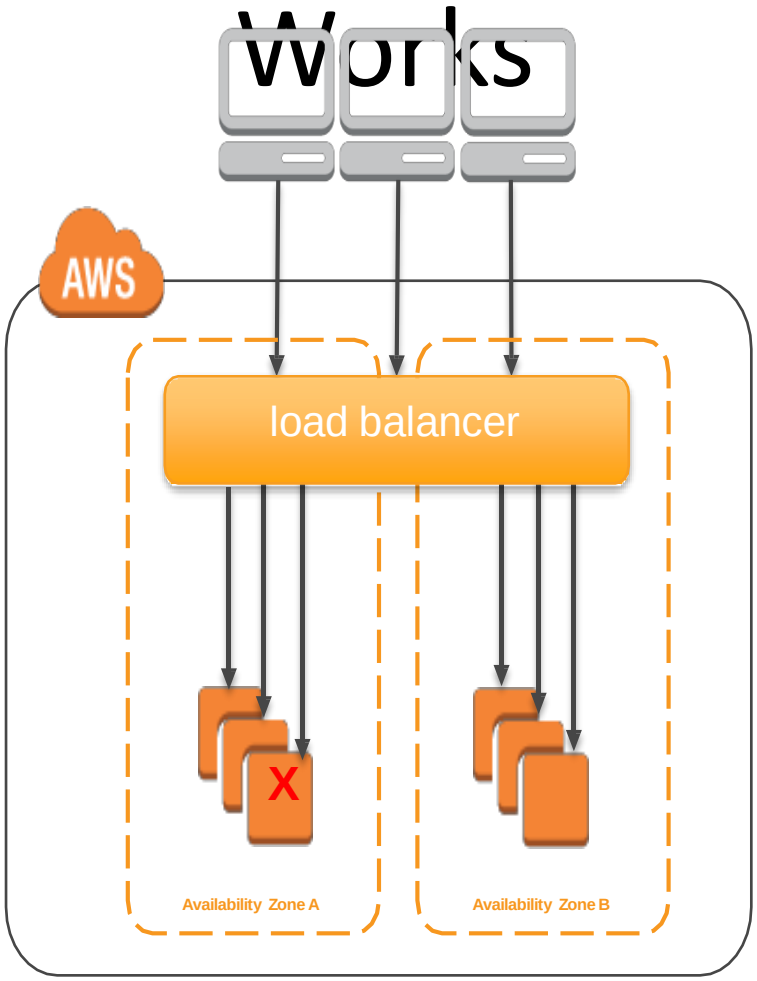
Elastic Load Balancing

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

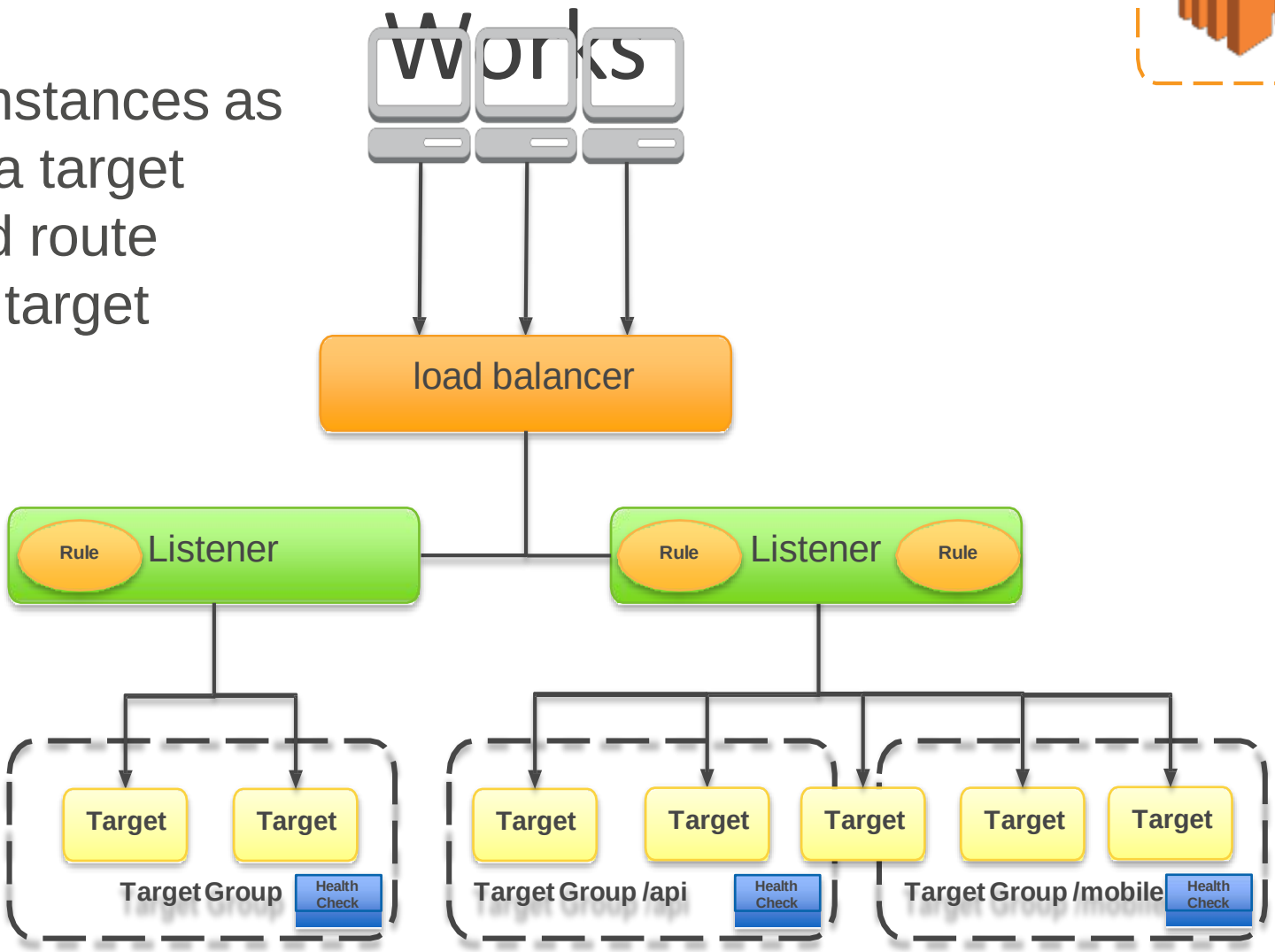
Classic Load Balancer - How It Works

Register instances with your load balancer.



Application Load Balancer – How Works

Register instances as targets in a target group, and route traffic to a target group.



Load Balancer Comparison

Classic Load Balancer benefits include support for:

- EC2-Classic.
- VPC.
- TCP and SSL listeners.
- Sticky sessions.

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.

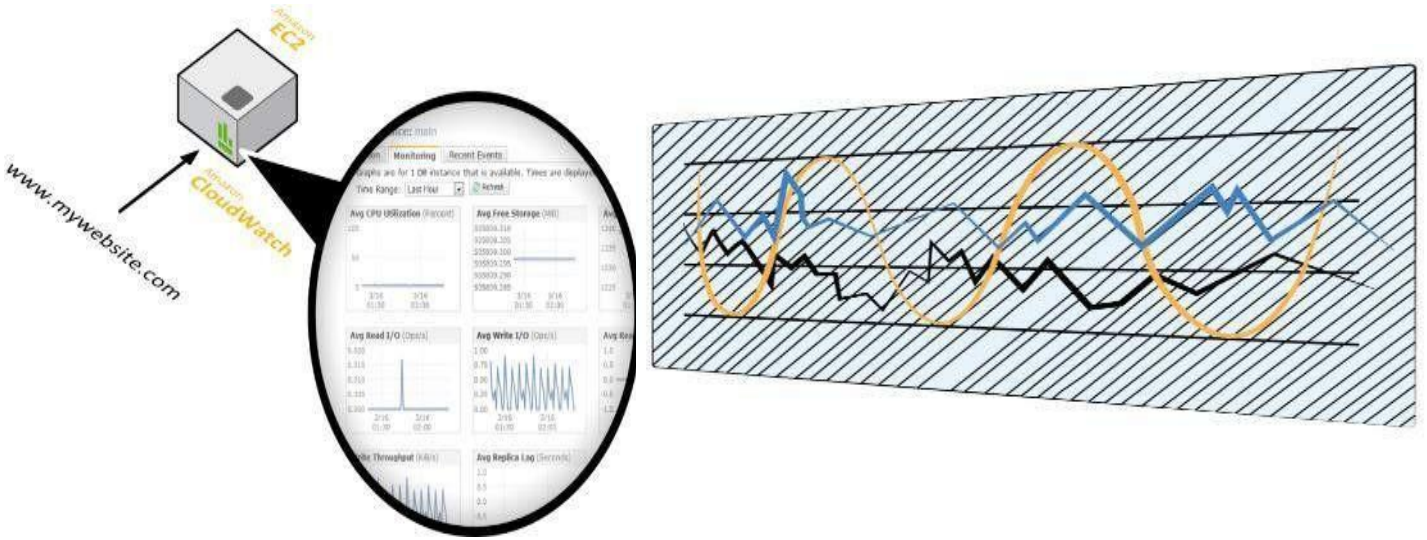
Amazon CloudWatch

Amazon CloudWatch

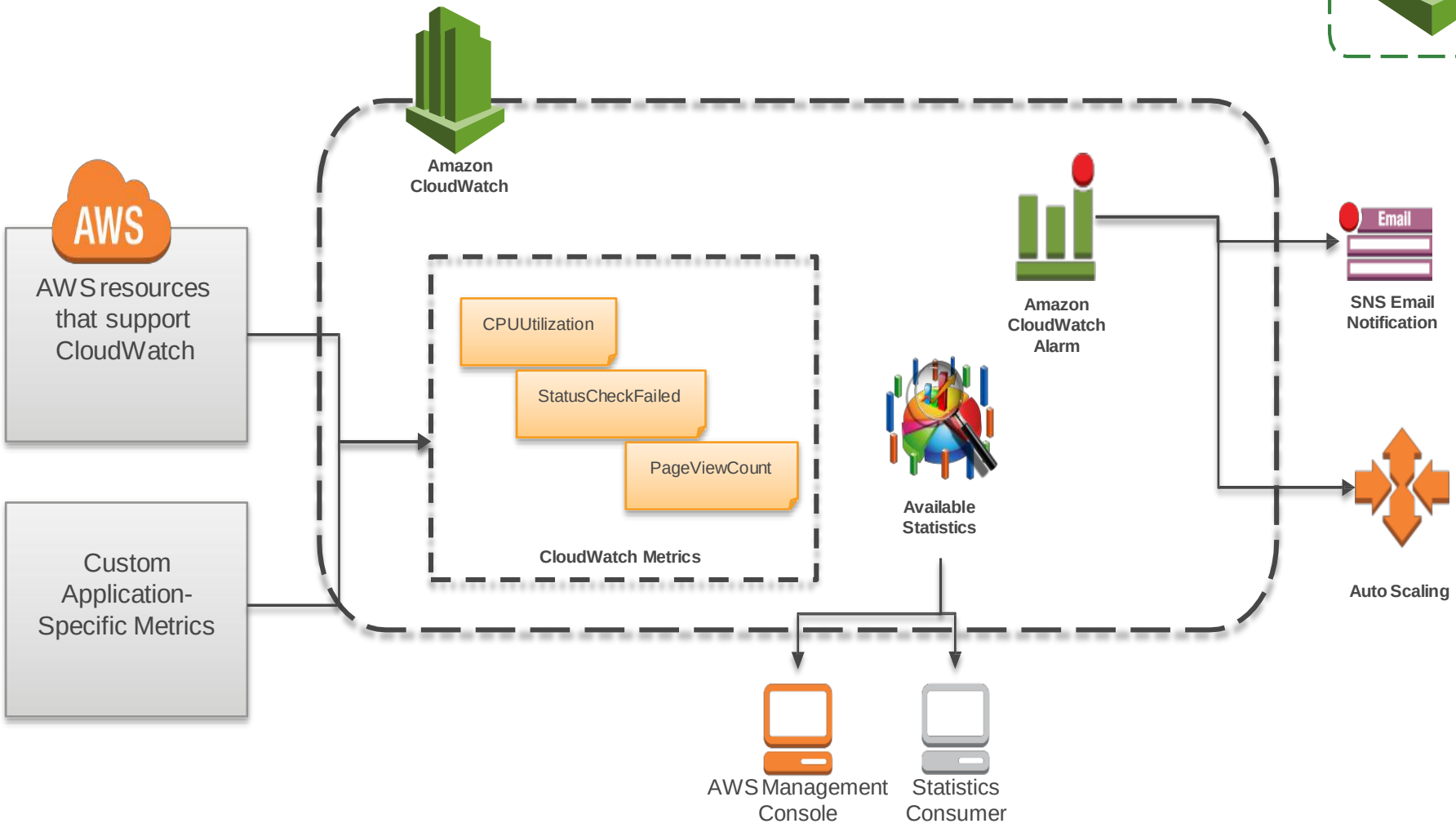
- A **monitoring service** for AWS cloud resources and the applications you run on AWS
- **Visibility into** resource utilization, operational performance, and overall demand patterns
- **Custom application-specific** metrics of your own
- **Accessible** via AWS Management Console, APIs, SDK, or CLI

Amazon CloudWatch Facts

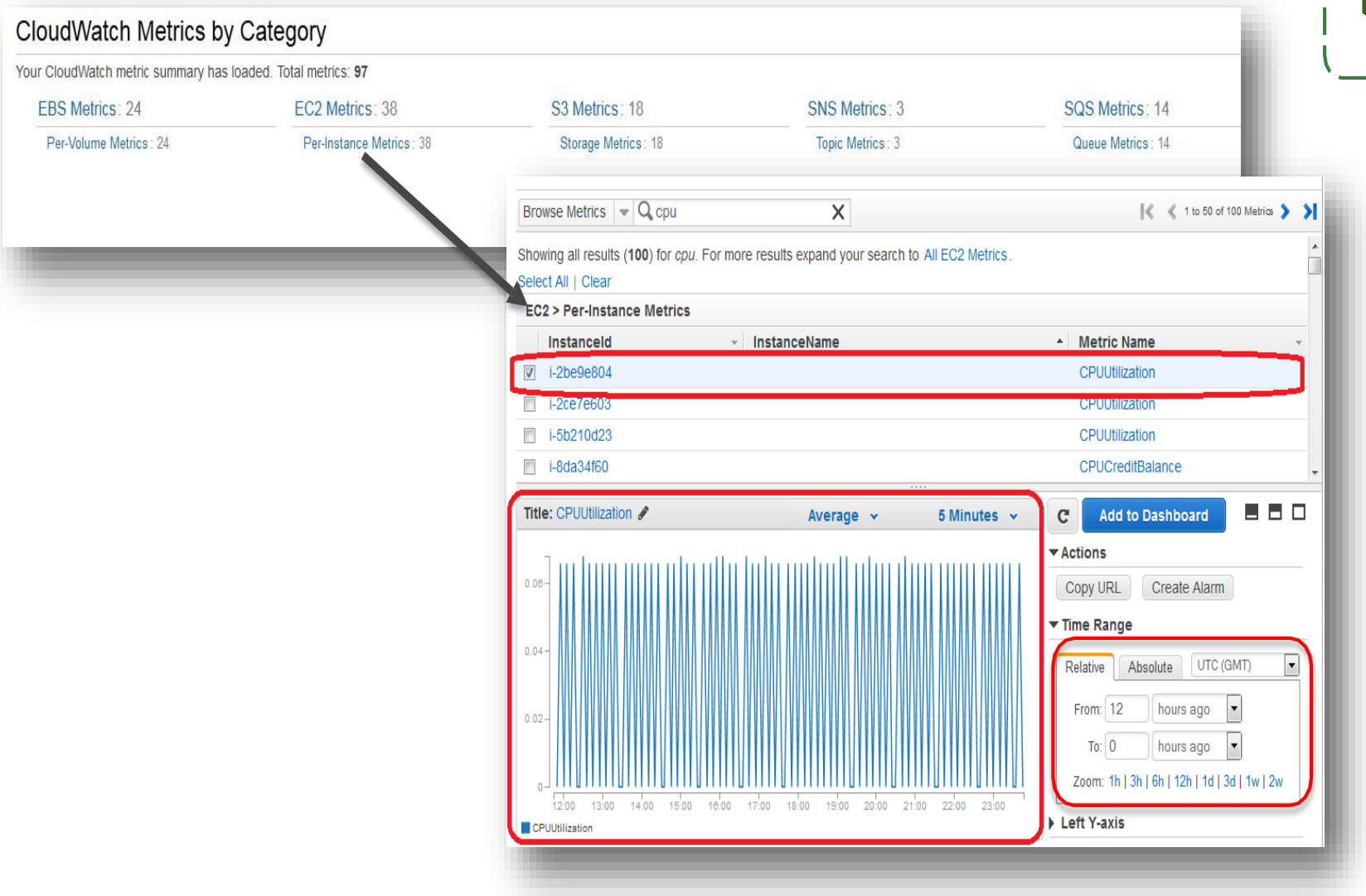
- Monitor other AWS resources
  - View graphics and statistics
- Set Alarms



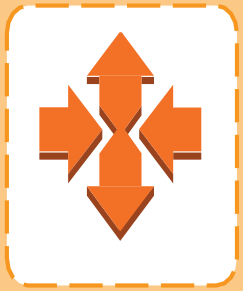
Amazon CloudWatch Architecture



CloudWatch Metrics Examples



Auto Scaling

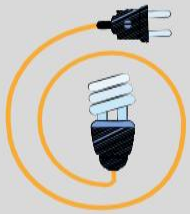


Auto Scaling

- **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

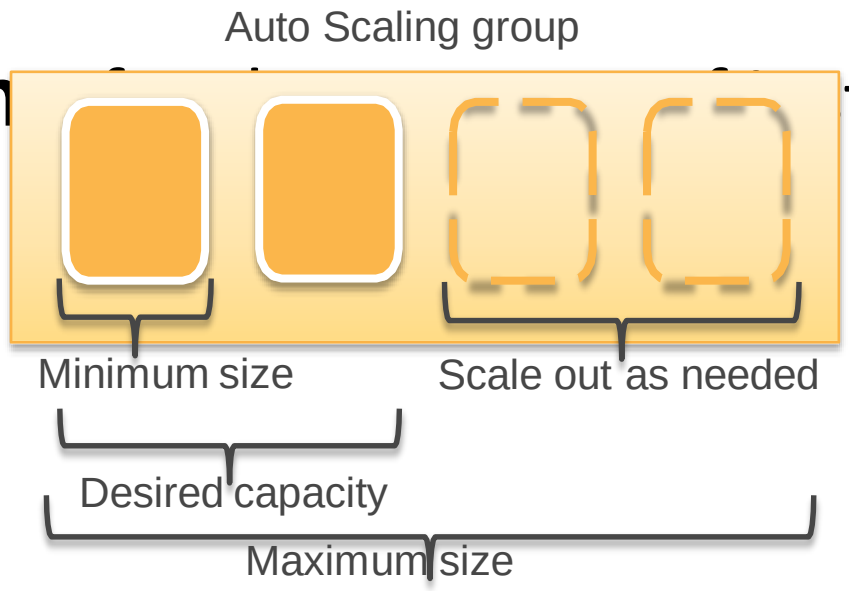
Better Cost Management

Launch Configurations

- A **launch configuration** is a template that an Auto Scaling group uses to launch EC2 instances.
- When you create a launch configuration, you can specify:
  - AMI ID
  - Instance type
  - Key pair
  - Security groups
  - Block device mapping
  - User data

Auto Scaling Groups

- Contain a collection of EC2 instances that share similar characteristics.
- Instances in an Auto Scaling group are treated as a
and management.
- **logical groupin** tance scaling



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 Basic Lifecycle