

CS 4037
Introduction to Cloud Computing
Lecture 30

Danyal Farhat
FAST School of Computing
NUCES Lahore

AWS Auto Scaling and Monitoring

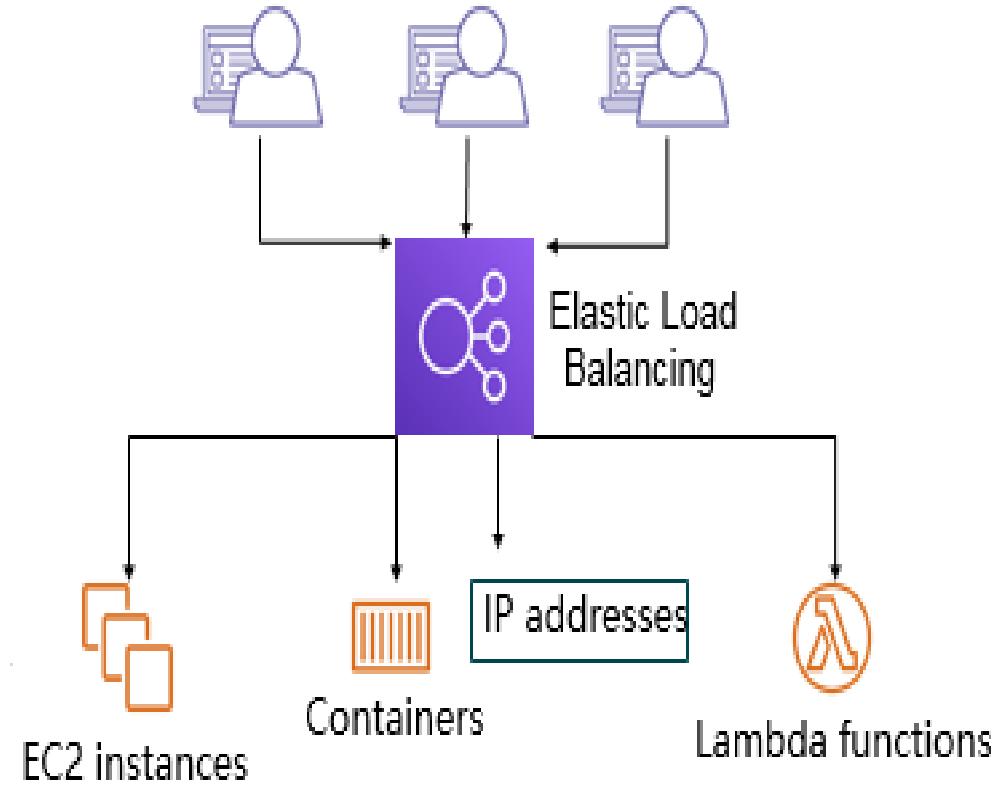
Lecture's Agenda

- **Elastic Load Balancing**
- Amazon CloudWatch
- Amazon EC2 Auto Scaling
- Certification Exams



Elastic Load Balancing

- Distributes incoming application or network traffic across multiple targets in a single Availability Zone or across multiple Availability Zones
- Scales load balancer as traffic to application changes over time



Types of Load Balancers

Application Load Balancer	Network Load Balancer	Classic Load Balancer (Previous Generation)
<ul style="list-style-type: none">• Load balancing of HTTP and HTTPS traffic• Routes traffic to targets based on content of request• Provides advanced request routing targeted at the delivery of modern application architectures, including microservices and containers• Operates at the application layer (OSI model layer 7)	<ul style="list-style-type: none">• Load balancing of TCP, UDP, and TLS traffic where extreme performance is required• Routes traffic to targets based on IP protocol data• Can handle millions of requests per second while maintaining ultra-low latencies• Is optimized to handle sudden and volatile traffic patterns• Operates at the transport layer (OSI model layer 4)	<ul style="list-style-type: none">• Load balancing of HTTP, HTTPS, TCP, and SSL traffic• Load balancing across multiple EC2 instances• Operates at both the application and transport layers.

Lecture's Agenda

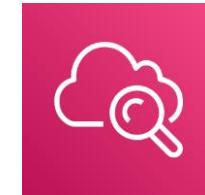
- Elastic Load Balancing
- **Amazon CloudWatch**
- Amazon EC2 Auto Scaling
- Certification Exams



Amazon CloudWatch

- **Monitors:**

- AWS resources
- Applications that run on AWS



- **Collects and tracks:**

- Standard metrics

- ✓ CPU Utilization (GHz and %)
- ✓ RAM Utilization (GBs and %)
- ✓ Storage Utilization (GBs and %)
- ✓ Network Inbound Throughput (IGW) (Mbps or Gbps)
- ✓ Network Outbound Throughput (IGW) (Mbps or Gbps)

- Custom metrics

- ✓ Network Inbound Throughput (NAT-GW) (Mbps or Gbps)
- ✓ Network Outbound Throughput (NAT-GW) (Mbps or Gbps)

Amazon
CloudWatch

Amazon CloudWatch (Cont.)

- **Alarms:**

- Send **notifications** to an Amazon SNS topic
- Perform Amazon EC2 **Auto Scaling** or Amazon EC2 actions



- **Events:**

- Define **rules to match changes** in AWS environment and route these events to one or more target functions or streams for processing

Amazon CloudWatch Alarms

- Create alarms based on:
 - Static threshold
 - Anomaly detection
 - Metric math expression
- Specify:
 - Namespace
 - Metric (CPU, RAM, Storage)
 - Statistic (Avg., Max, Min)
 - Period
 - Conditions (>, >=, <=, <)
 - Additional configuration
 - Actions

The screenshot shows the configuration interface for a new CloudWatch Metrics Alarm. The 'Statistic' is set to 'Average' over a '5 minutes' period. The 'Threshold type' is selected as 'Static', which uses a value as a threshold. The condition is defined as 'Whenever CPUUtilization is... Greater than threshold'. The threshold value is set to 100. There is also a link to 'Additional configuration'.

Statistic: Average

Period: 5 minutes

Conditions:

Threshold type:

- Static: Use a value as a threshold
- Anomaly detection: Use a band as a threshold

Whenever CPUUtilization is... Define the alarm condition:

- Greater: > threshold
- Greater/Equal: >= threshold
- Lower/Equal: <= threshold
- Lower: < threshold

than... Define the threshold value: 100

Must be a number

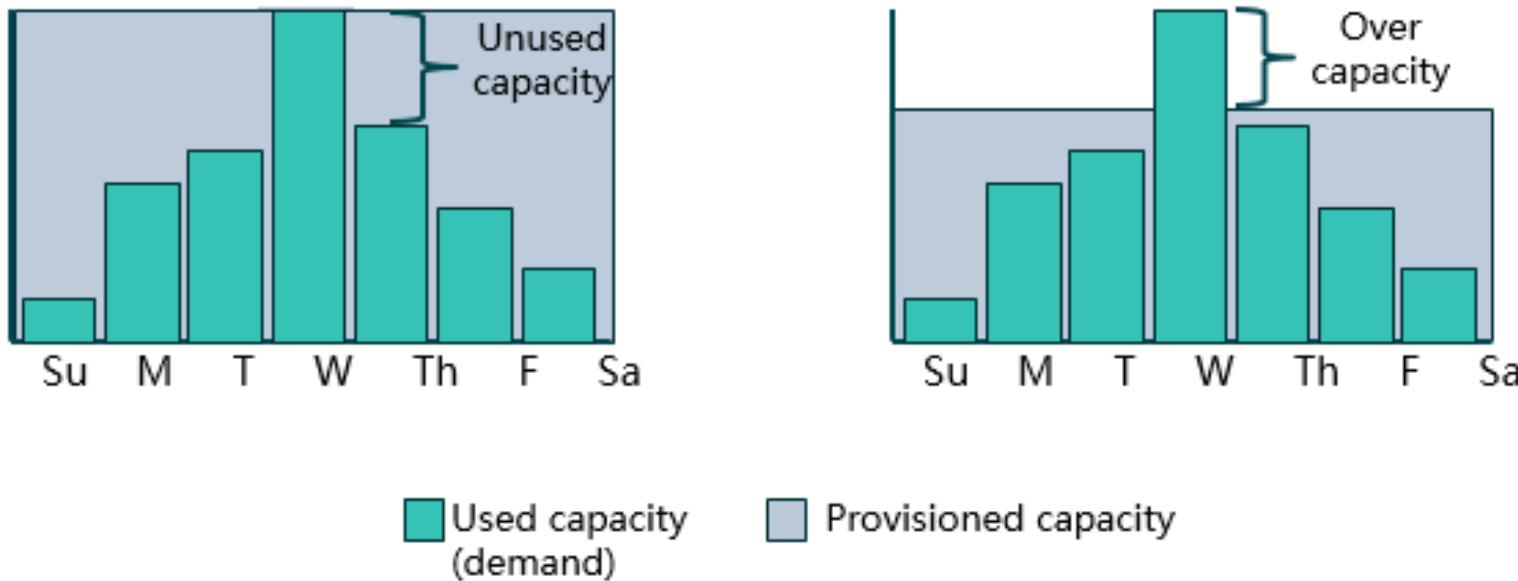
► Additional configuration

Lecture's Agenda

- Elastic Load Balancing
- Amazon CloudWatch
- **Amazon EC2 Auto Scaling**
- Certification Exams

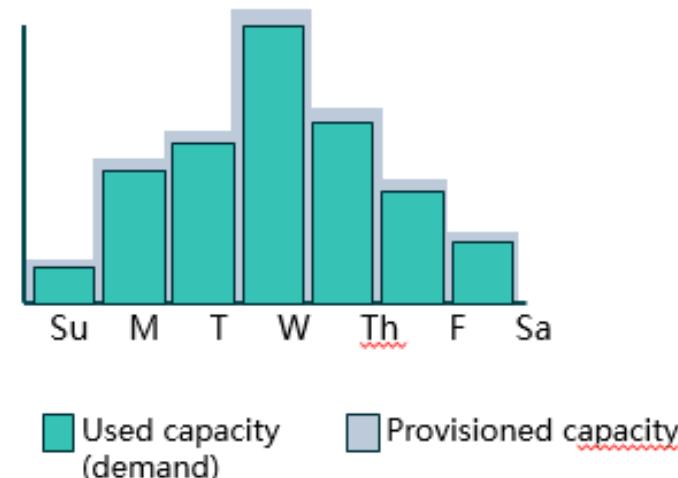


Why is Scaling Important?



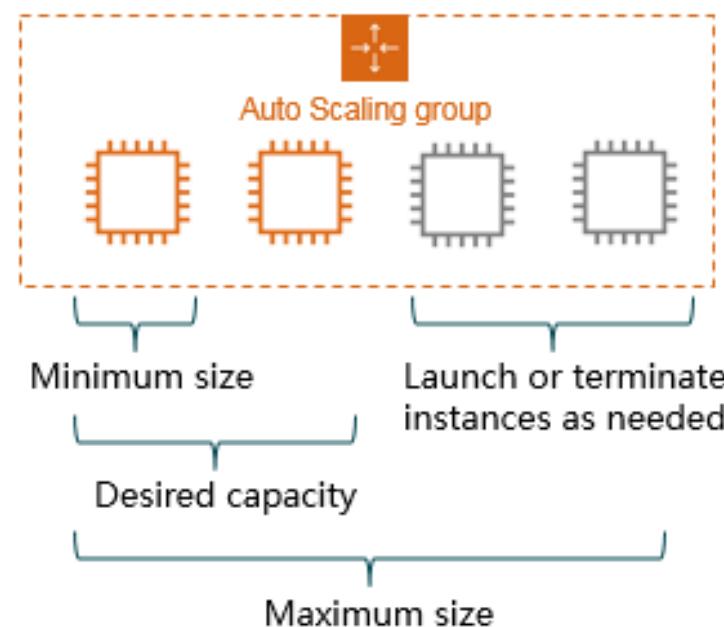
Amazon EC2 Auto Scaling

- Helps to maintain application availability
- Enables to automatically add or remove EC2 instances according to the conditions defined by the customer
- Detects impaired EC2 instances and unhealthy applications, and replaces the instances without user intervention
- Provides several scaling options
 - Manual
 - Scheduled
 - Dynamic
 - On-demand
 - Predictive

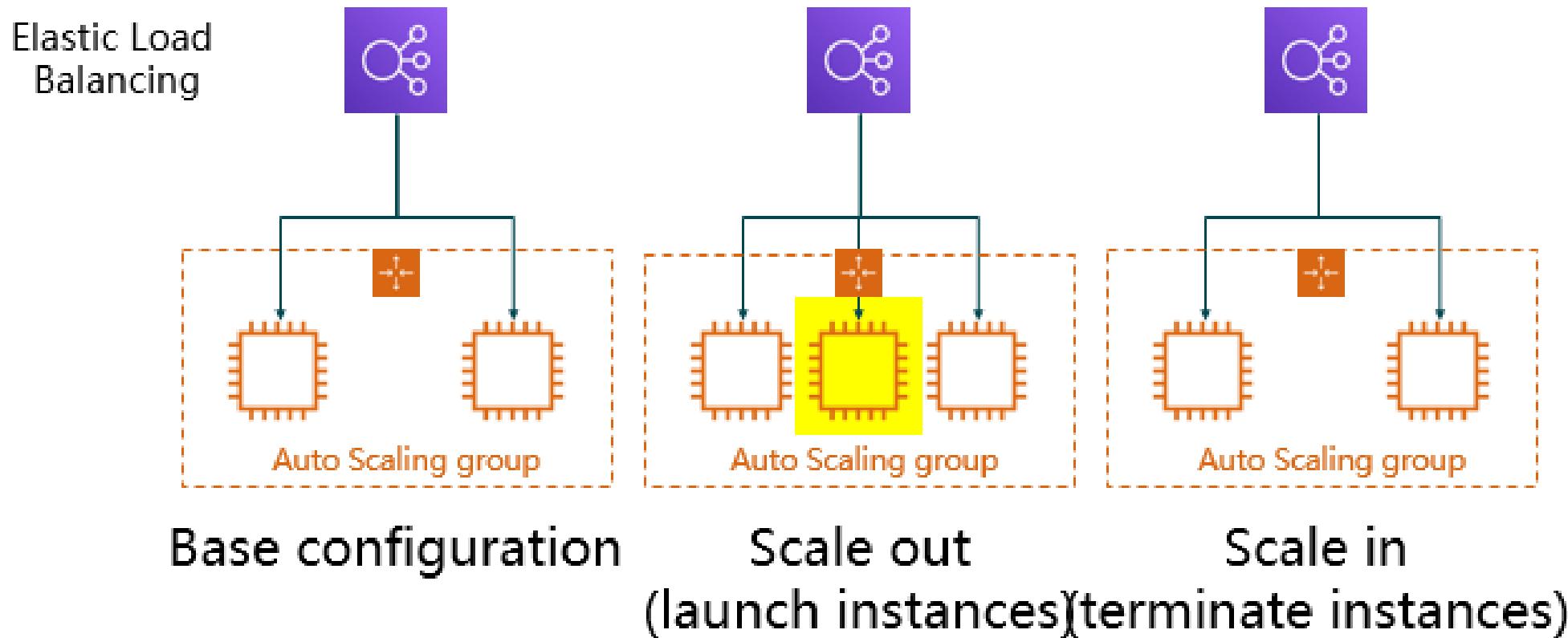


Auto Scaling Groups

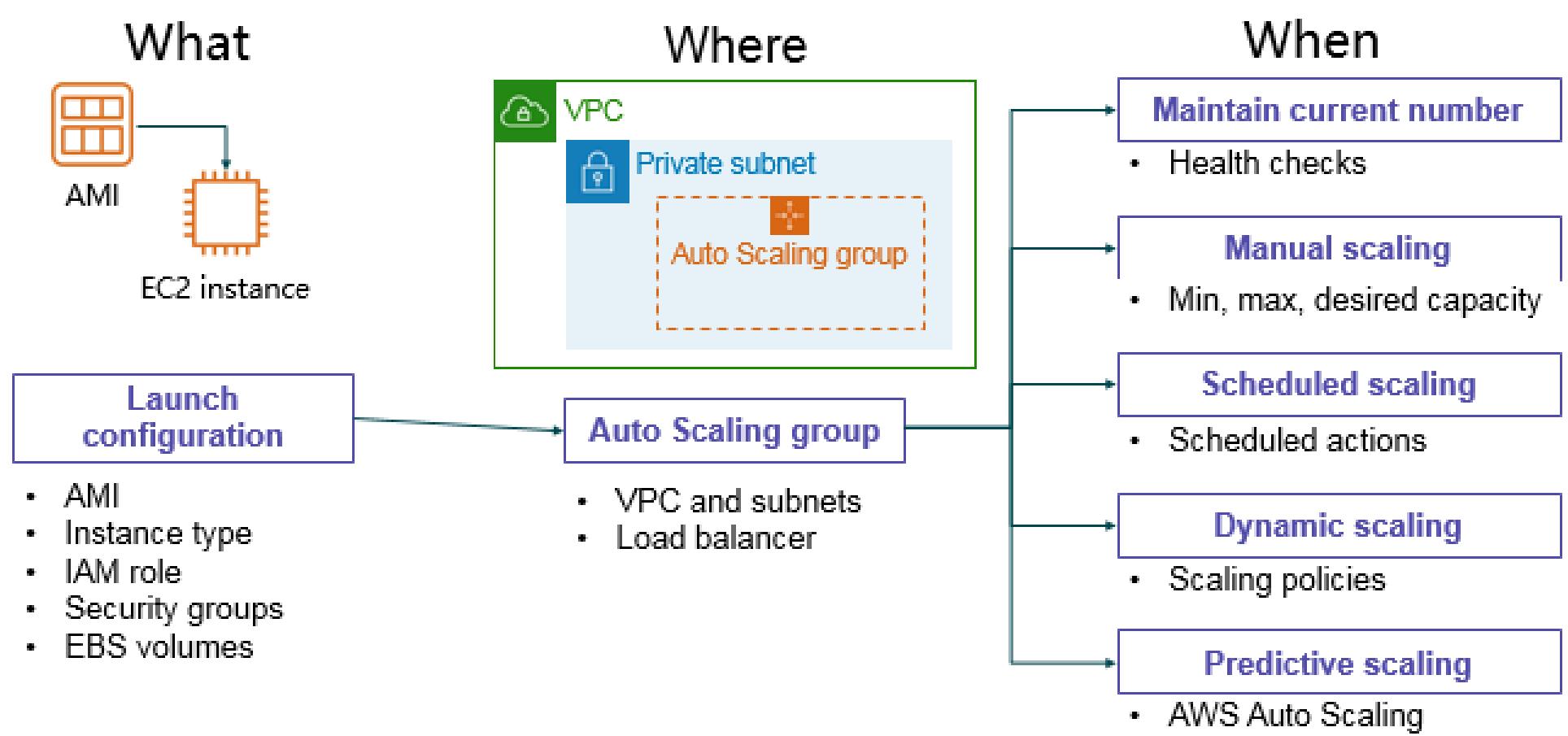
- “An **Auto Scaling group** is a collection of EC2 instances that are treated as a logical grouping for the purposes of automatic scaling and management.”



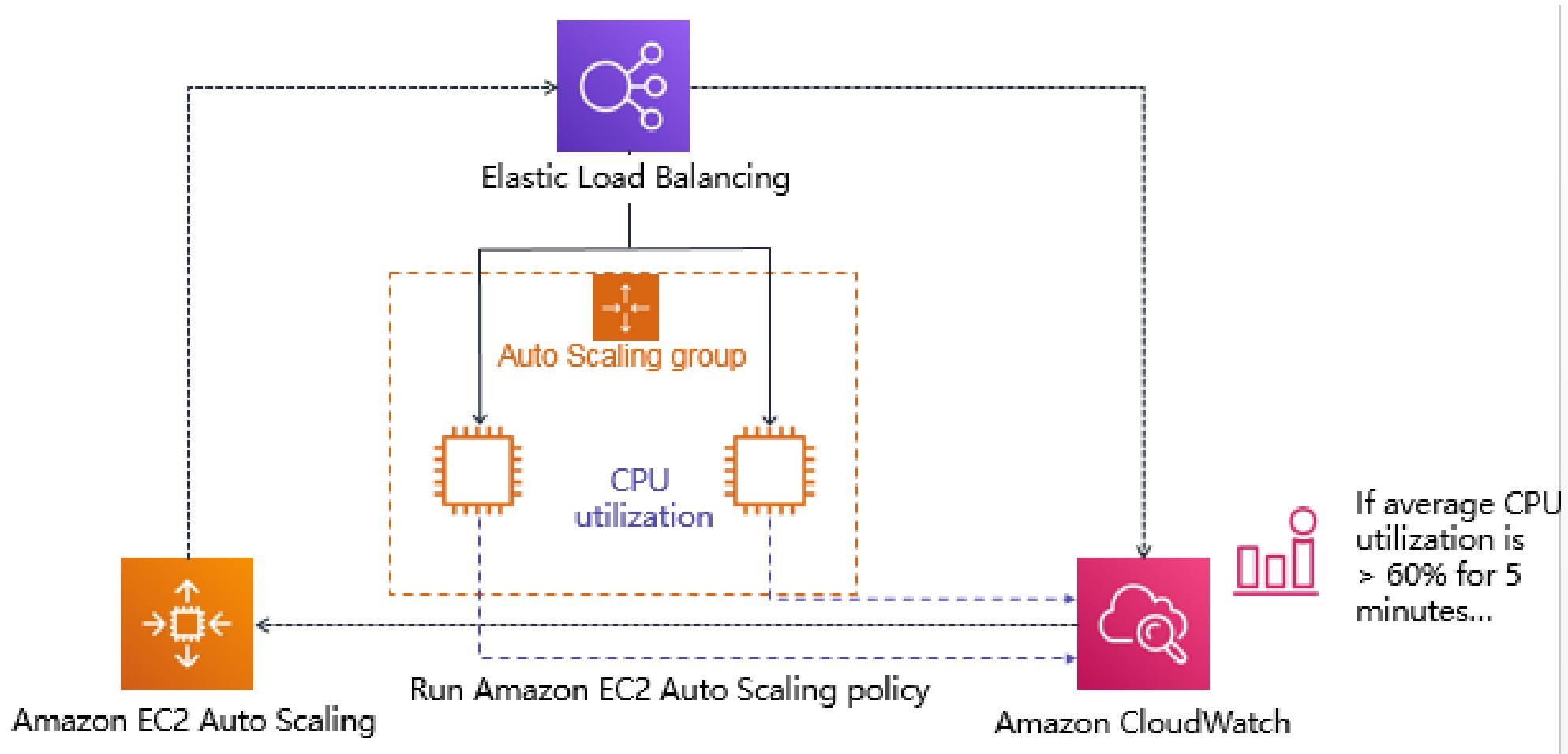
Scaling Out vs. Scaling In



How Amazon EC2 Auto Scaling Works

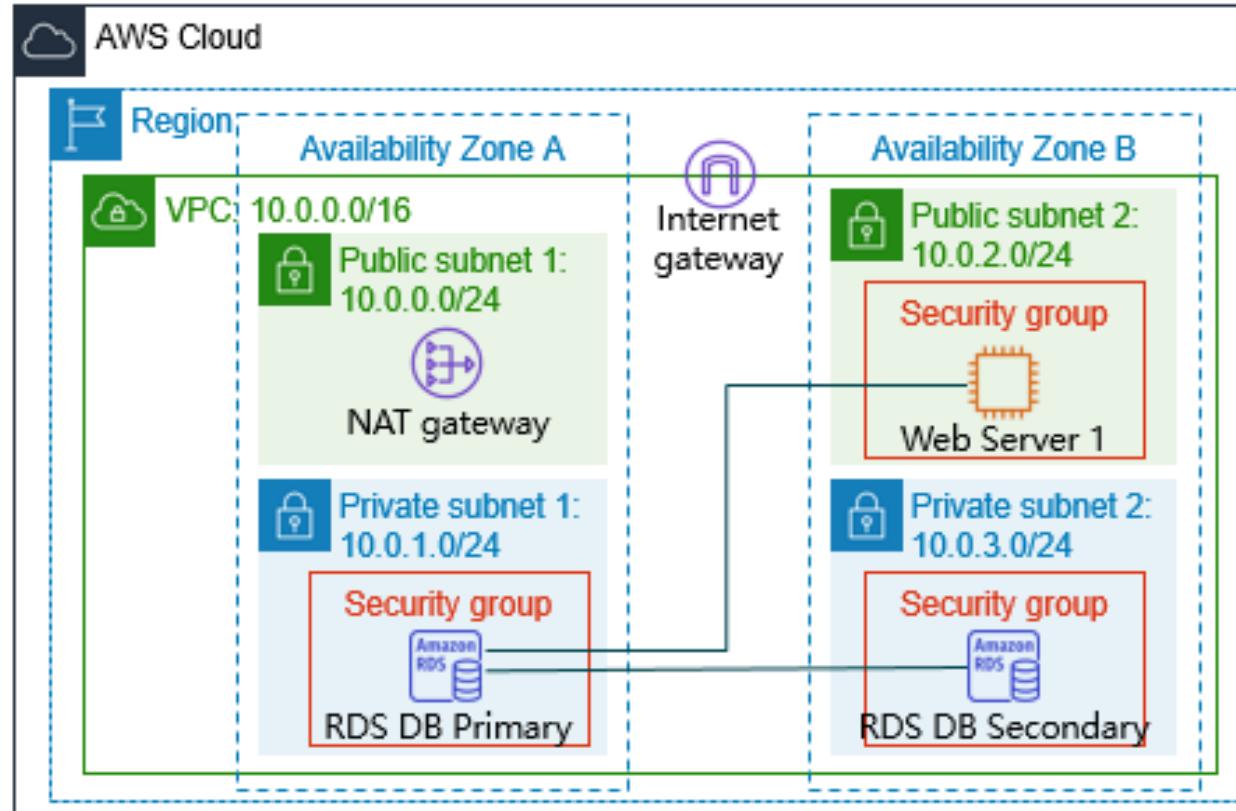


Implementing Dynamic Scaling



Lab 6: Scale and Load Balance Your Architecture

Lab Scenario: In this lab, you will use **Elastic Load Balancing** and **Amazon EC2 Auto Scaling** to load balance and scale your infrastructure. You will start with the given infrastructure.



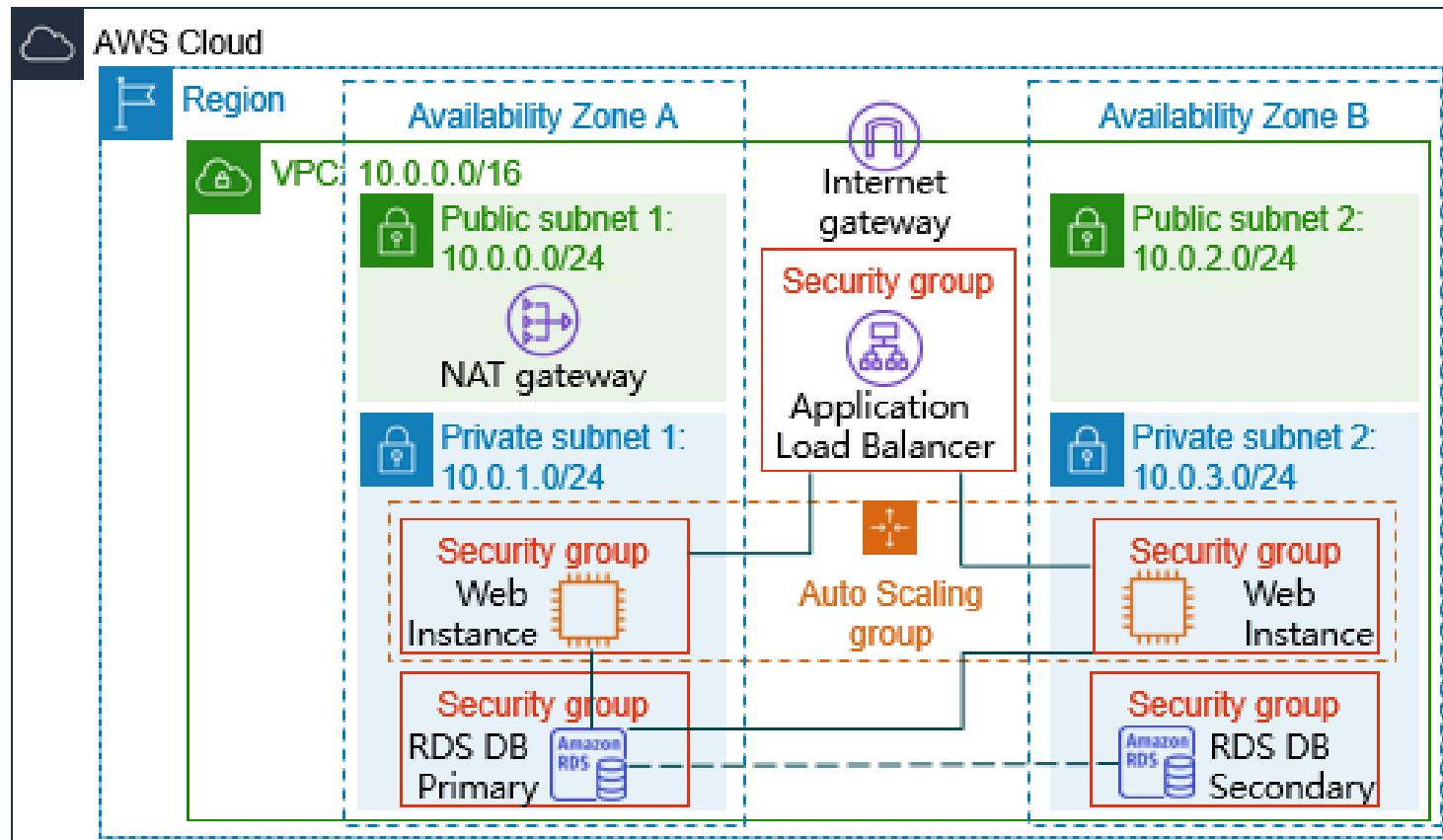
Lab 6: Scale and Load Balance Your Architecture (Cont.)

Lab Tasks:

- Task 1 – Create an AMI from a running instance
- Task 2 – Create an Application Load Balancer
- Task 3 – Create a launch configuration and an Auto Scaling group
- Task 4 – Automatically scale new instances within a private subnet
- Task 5 – Create Amazon CloudWatch alarms and monitor performance of your infrastructure

Lab 6: Scale and Load Balance Your Architecture (Cont.)

Final Product:



Lecture's Agenda

- Elastic Load Balancing
- Amazon CloudWatch
- Amazon EC2 Auto Scaling
- **Certification Exams**



AWS Certification Exams

*This course helps
prepare you for
the AWS **Cloud
Practitioner**
certification exam*

Available AWS Certifications



Professional

Two years of comprehensive experience designing, operating, and troubleshooting solutions using the AWS Cloud

Associate

One year of experience solving problems and implementing solutions using the AWS Cloud

Foundational

Six months of fundamental AWS Cloud and industry knowledge



Specialty

Technical AWS Cloud experience in the Specialty domain as specified in the [exam guide](#)



Additional Resources

- Details about the exam—including how to register for it
 - <https://aws.amazon.com/certification/certified-cloud-practitioner/>
- Practice Questions Sites
 - <https://www.examtopics.com/exams/amazon/aws-certified-cloud-practitioner/>
 - <https://www.daypo.com/buscar.php?t=CLF-C01>

Questions?