

AWS Academy Cloud Foundations

Módulo 10: Monitoreo y escalado automático



Temas

- Elastic Load Balancing
- Amazon CloudWatch
- Amazon EC2 Auto Scaling

Actividades

- Actividad de Elastic Load Balancing
- Actividad de Amazon CloudWatch

Laboratorio

- Ajuste de la escala y balanceo de la carga de su arquitectura



**Revisión de
conocimientos**

Después de completar este módulo, debería ser capaz de lo siguiente:

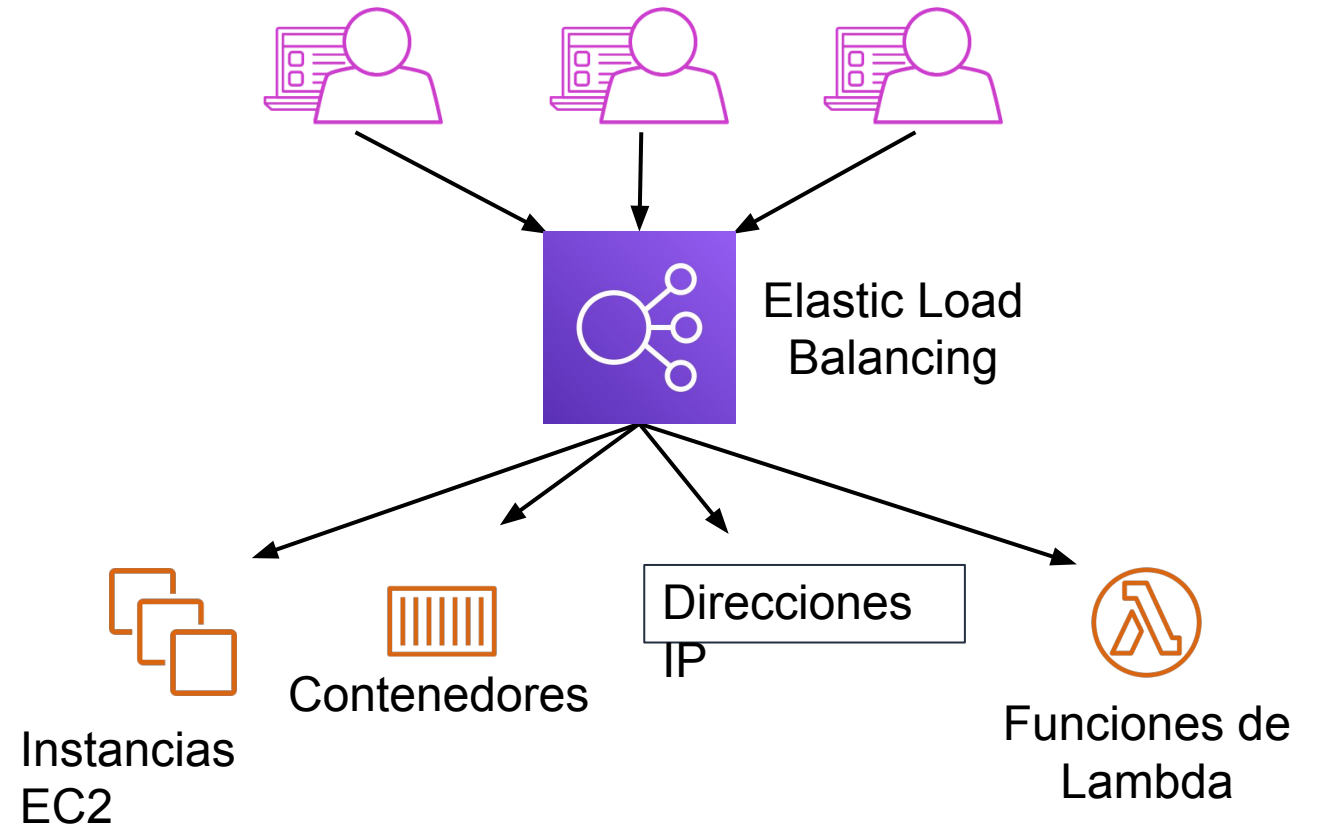
- Indicar cómo distribuir el tráfico entre las instancias de Amazon Elastic Compute Cloud (Amazon EC2) usando Elastic Load Balancing
- Identificar cómo Amazon CloudWatch permite monitorear recursos y aplicaciones de AWS en tiempo real
- Explicar cómo Amazon EC2 Auto Scaling lanza y publica servidores en respuesta a los cambios en las cargas de trabajo
- Realizar tareas de escalado y balanceo de carga para mejorar una arquitectura

Módulo 10: Monitoreo y escalado automático

Sección 1: Elastic Load Balancing

Elastic Load Balancing

- Distribuye el tráfico entrante de las aplicaciones o de la red entre varios destinos en una única zona de disponibilidad o en varias zonas de disponibilidad.
- Escala el balanceador de carga a medida que el tráfico dirigido a la aplicación cambia con el tiempo.



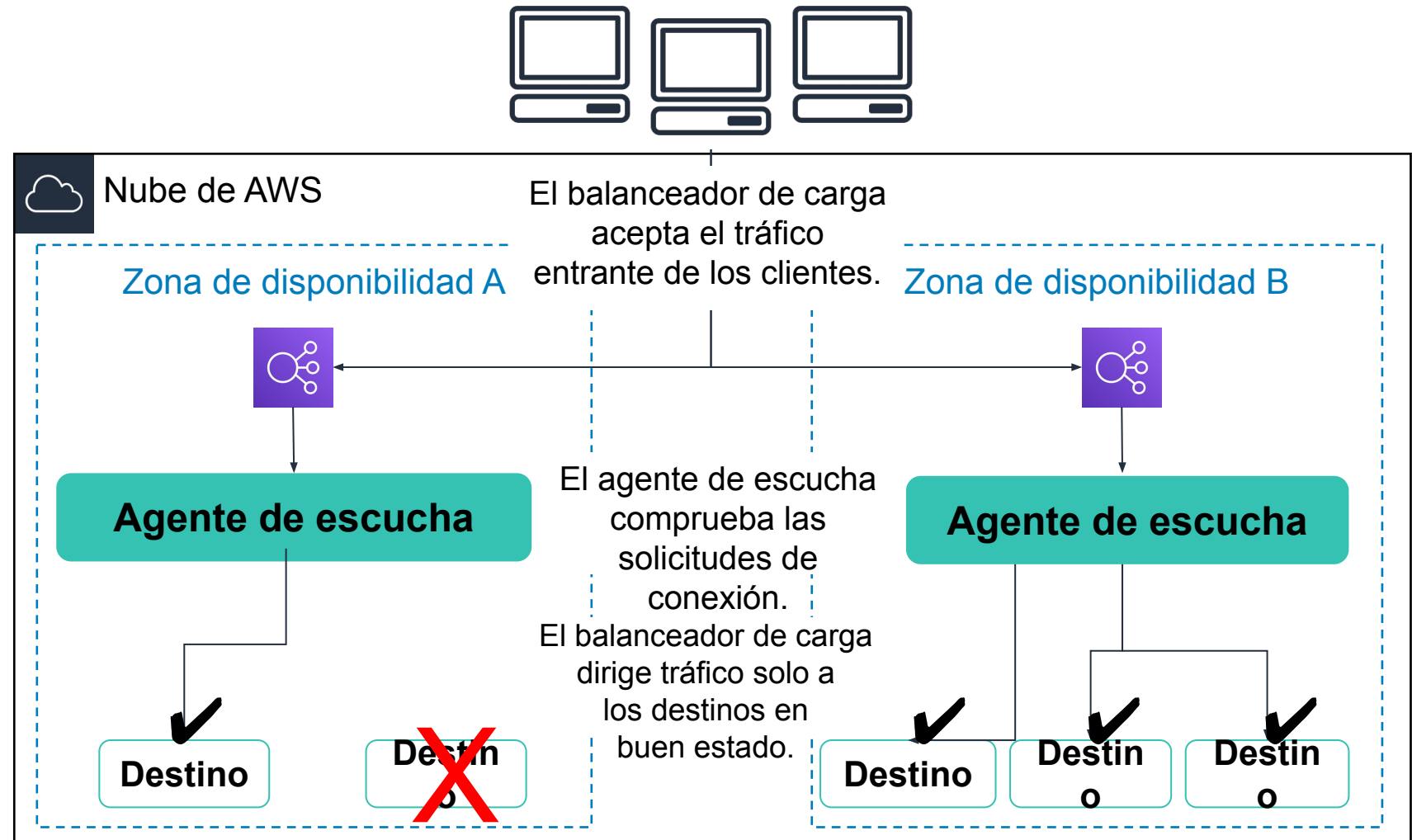
Tipos de balanceadores de carga

Balanceador de carga de aplicaciones	Balanceador de carga de red	Balanceador de carga clásico (generación anterior)
<ul style="list-style-type: none">• Balanceo de carga del tráfico HTTP y HTTPS	<ul style="list-style-type: none">• Balanceo de carga del tráfico TCP, UDP y TLS donde se requiere un rendimiento extremo	<ul style="list-style-type: none">• Balanceo de carga del tráfico HTTP, HTTPS, TCP y SSL
<ul style="list-style-type: none">• Dirige el tráfico a los destinos en función del contenido de la solicitud.• Proporciona direccionamiento de solicitudes avanzadas a su entrega en arquitecturas modernas de aplicaciones, como los microservicios y los contenedores.	<ul style="list-style-type: none">• Dirige el tráfico a los destinos en función de los datos del protocolo IP.• Puede gestionar millones de solicitudes por segundo y, a la vez, mantener latencias muy bajas.• Está optimizado para gestionar patrones de tráfico repentinos y volátiles.	<ul style="list-style-type: none">• Balanceo de carga entre varias instancias EC2
<ul style="list-style-type: none">• Funciona en la capa de aplicación (capa 7 del modelo OSI).	<ul style="list-style-type: none">• Funciona en la capa de transporte (capa 4 del modelo OSI).	<ul style="list-style-type: none">• Funciona en las capas de aplicación y de transporte.

Cómo funciona Elastic Load Balancing

- Con los balanceadores de carga de aplicaciones y los balanceadores de carga de red, **los destinos se registran en grupos de destinos** y el tráfico se dirige hacia los grupos de destinos.
- Con los balanceadores de carga clásicos, **las instancias se registran con el balanceador de carga**.

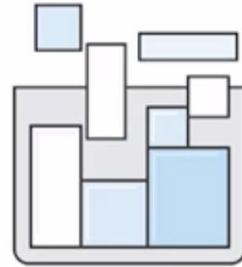
El balanceador de carga lleva a cabo comprobaciones de estado para monitorear el estado de los destinos registrados.



Casos de uso de Elastic Load Balancing



Aplicaciones de alta
disponibilidad
y tolerantes a errores



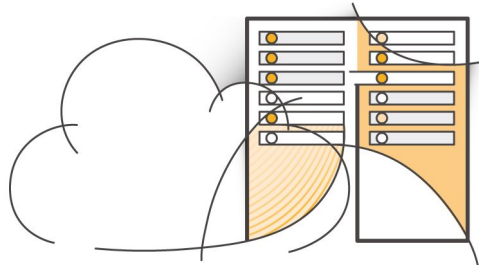
Aplicaciones en
contenedores



Elasticidad y
escalabilidad



Nube virtual
privada
(VPC)



Entornos híbridos



Invocación de las
funciones de Lambda
a través de HTTP(S)

Actividad: Elastic Load Balancing

Debe admitir el tráfico a una aplicación en contenedores.

Balanceador de carga de aplicaciones

El tráfico TCP es extremadamente irregular e impredecible.

Balanceador de carga de red

Necesita balanceo de carga simple con múltiples protocolos.

Balanceador de carga clásico

Debe admitir una dirección IP estática o elástica o un destino IP fuera de la VPC.

Balanceador de carga de red

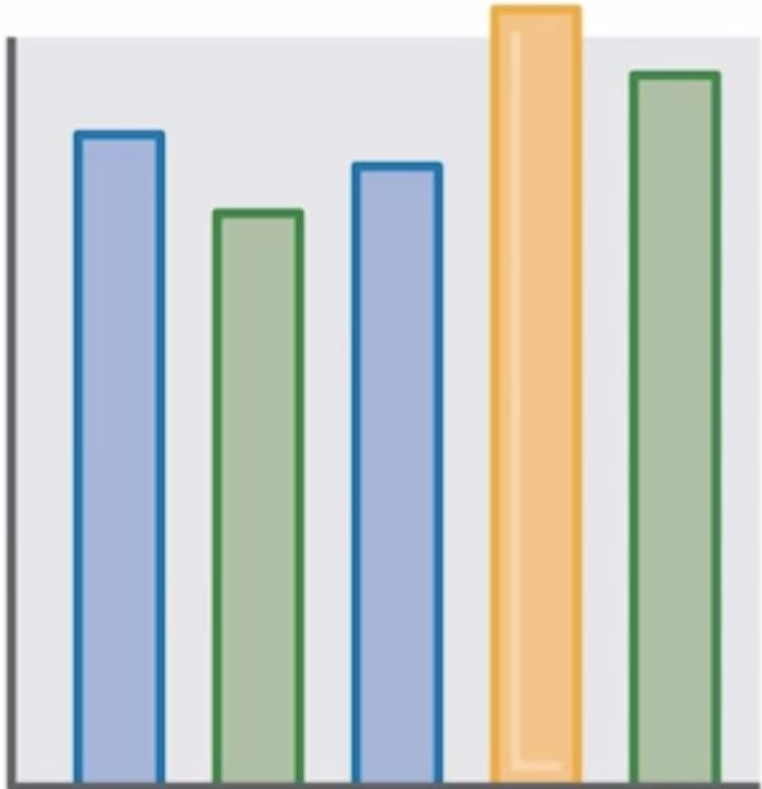
Necesita un balanceador de carga que pueda gestionar millones de solicitudes por segundo mientras mantiene latencias bajas.

Balanceador de carga de red

Debe admitir solicitudes HTTPS.

Balanceador de carga de aplicaciones

Monitoreo del balanceador de carga



- **Métricas de Amazon CloudWatch:** estas métricas sirven para verificar que el sistema funcione según lo previsto y, además, este servicio crea una alarma para iniciar una acción si alguna métrica pasa a estar fuera de un rango aceptable.
- **Registros de acceso:** estos registros guardan información detallada acerca de las solicitudes enviadas a su balanceador de carga.
- **Registros de AWS CloudTrail:** estos registros guardan información acerca del quién, qué, cuándo y dónde de las interacciones de la API en los servicios de AWS.

Aprendizajes clave de la sección 1



- Elastic Load Balancing distribuye el tráfico entrante de las aplicaciones o de la red entre varios destinos en una o más zonas de disponibilidad.
- Elastic Load Balancing admite tres tipos de balanceadores de carga:
 - Balanceador de carga de aplicaciones
 - Balanceador de carga de red
 - Balanceador de carga clásico
- ELB ofrece comprobaciones de estado, seguridad y monitoreo de las instancias.

Módulo 10: Monitoreo y escalado automático

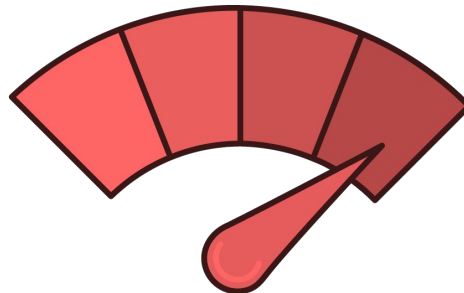
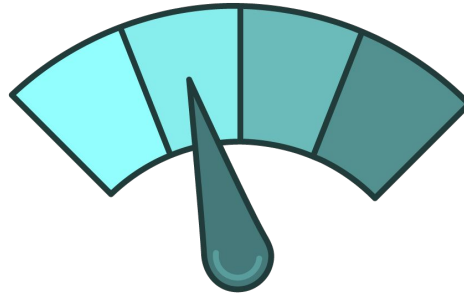
Sección 2: Amazon CloudWatch

Para utilizar AWS de manera eficiente, necesita información acerca de sus recursos de AWS:

- ¿Cómo se sabe cuándo se debe **lanzar más instancias de Amazon EC2**?
- ¿Se está viendo afectado el **rendimiento o la disponibilidad de su aplicación** por insuficiencia de capacidad?
- En realidad, ¿cuánto **se está usando** de su infraestructura?



Amazon CloudWatch



- Monitoreo:
 - Recursos de AWS
 - Aplicaciones que se ejecutan en AWS
- Recopilación y seguimiento:
 - Métricas estándar
 - Métricas personalizadas
- Alarmas:
 - Enviar notificaciones a un tema de Amazon SNS
 - Efectuar acciones de Amazon EC2 Auto Scaling o Amazon EC2
- Eventos:
 - Definir reglas para que coincidan con los cambios en el entorno de AWS y dirigir estos eventos a uno o más flujos o funciones de destino para su procesamiento

Las alarmas de CloudWatch

- Cree alarmas en función de lo siguiente:
 - Límite estático
 - Detección de anomalías
 - Expresión matemática de métricas
- Especifique lo siguiente:
 - Espacio de nombres
 - Métrica
 - Estadística
 - Periodo
 - Condiciones
 - Configuración adicional
 - Acciones

Statistic

Q Average X

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

Actividad: Amazon CloudWatch



Amazon
EC2



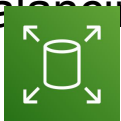
Amazon
RDS



Amazon
S3



Elastic Load
Balancing



Amazon Elastic
Block Store

Si el uso promedio de la CPU es $> 60\%$ durante 5 minutos...

Correcto

.

Si la cantidad de conexiones simultáneas es > 10 durante 1 minuto...

Correcto

.

Si el tamaño máximo del bucket en bytes es de alrededor de 3 durante 1 día...

Incorrecto. *Alrededor* no es una opción de límite. Debe especificar un límite de $>$, \geq , \leq o $<$.

Si la cantidad de hosts en buen estado es < 5 durante 10 minutos...

Correcto

.

Si el volumen de operaciones de lectura es > 1000 durante 10 segundos...

Incorrecto. Debe especificar una estadística (por ejemplo, el *volumen promedio*).

Aprendizajes clave de la sección 2

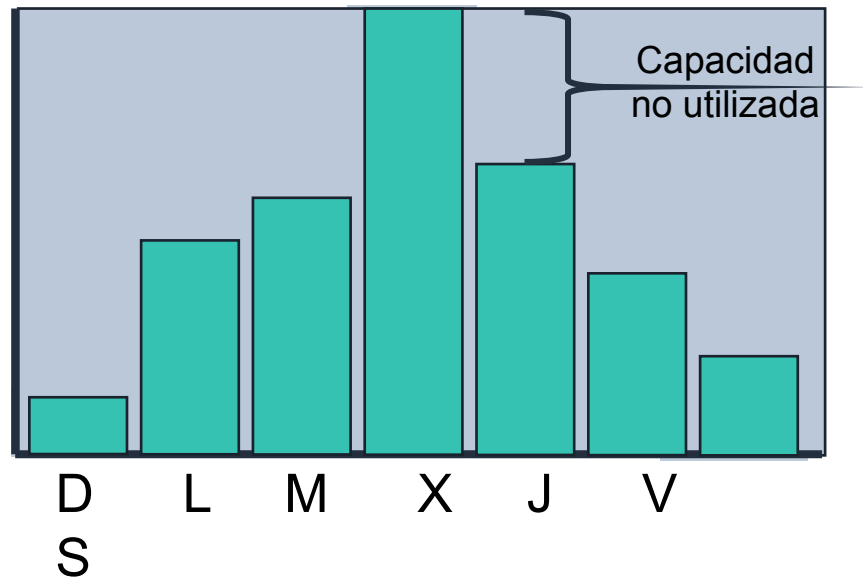


- Amazon CloudWatch lo ayuda a monitorear sus recursos de AWS y las aplicaciones que ejecuta en AWS en tiempo real.
- CloudWatch le permite realizar lo siguiente:
 - Recopilar y hacer un seguimiento de las métricas estándar y personalizadas.
 - Establecer alarmas para enviar notificaciones automáticas a los temas de SNS o efectuar acciones de Amazon EC2 Auto Scaling o Amazon EC2.
 - Definir reglas que coincidan con los cambios en su entorno de AWS y dirigir estos eventos a los destinos para su procesamiento.

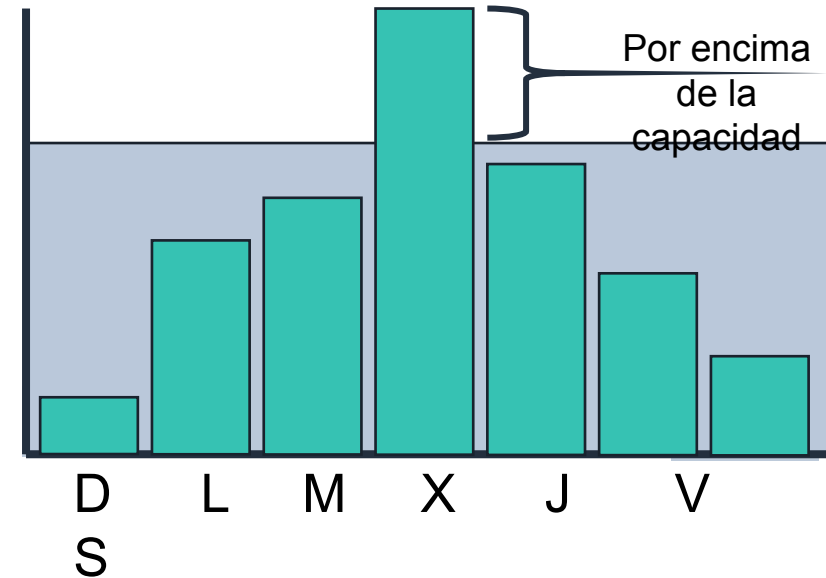
Módulo 10: Monitoreo y escalado automático

Sección 3: Amazon EC2 Auto Scaling

¿Por qué es importante el escalado?

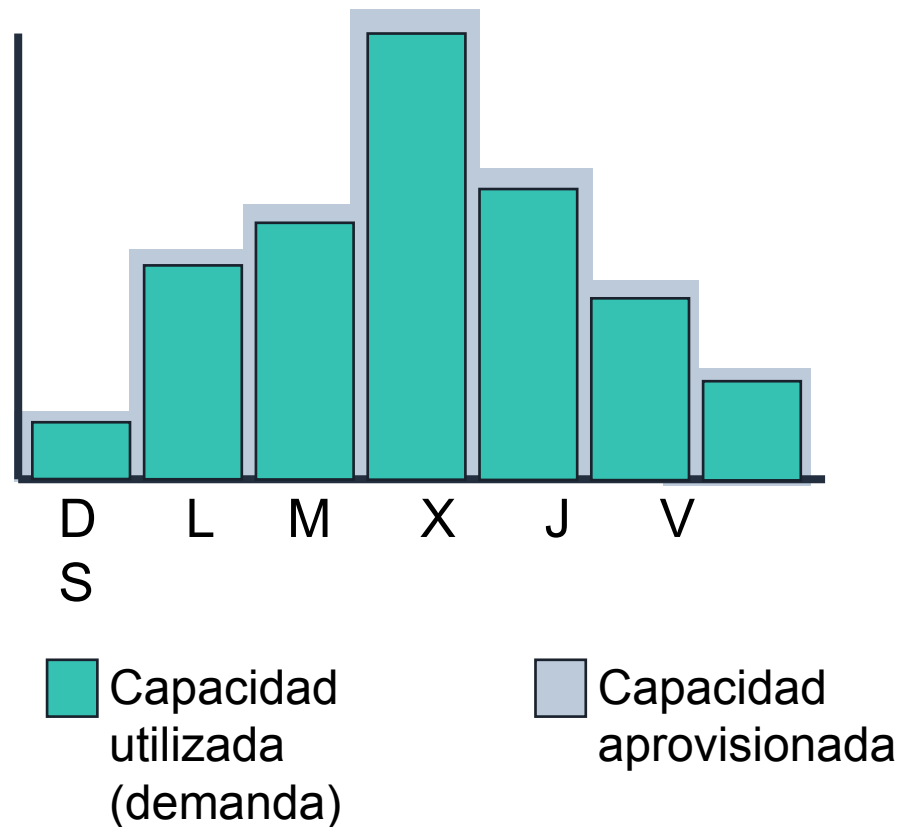


Capacidad utilizada (demanda)



Capacidad aprovisionada

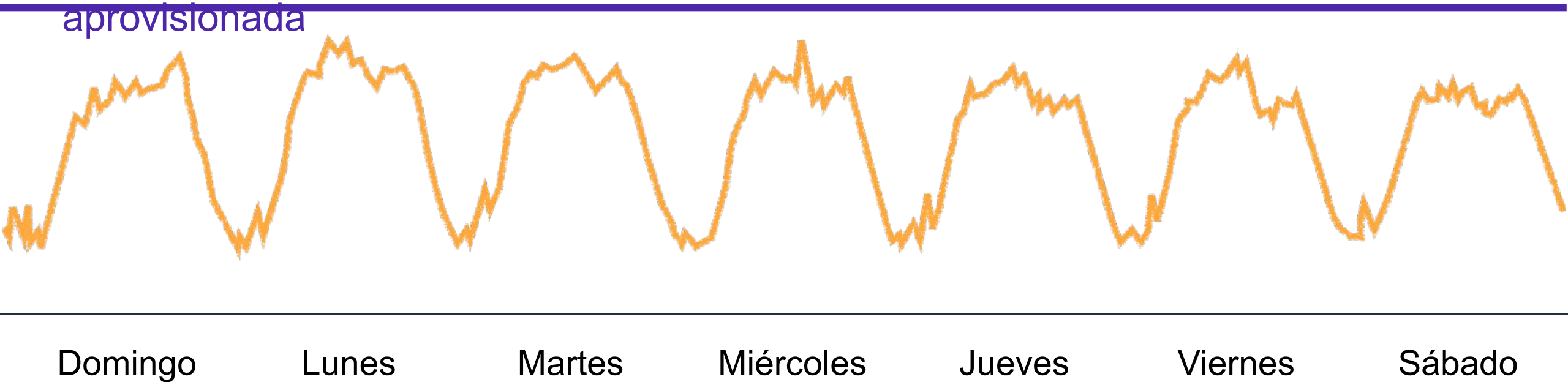
Amazon EC2 Auto Scaling



- Lo ayuda a mantener la disponibilidad de las aplicaciones
- Le permite agregar o eliminar automáticamente instancias EC2 de acuerdo con las condiciones que defina
- Detecta las instancias EC2 dañadas y las aplicaciones en mal estado, y reemplaza las instancias sin su intervención
- Ofrece varias opciones de escalado: manual, programado, dinámico o bajo demanda y predictivo

Tráfico semanal habitual en Amazon.com

Capacidad
aprovisionada



Tráfico que llega a Amazon.com en noviembre

Capacidad
aprovisionada

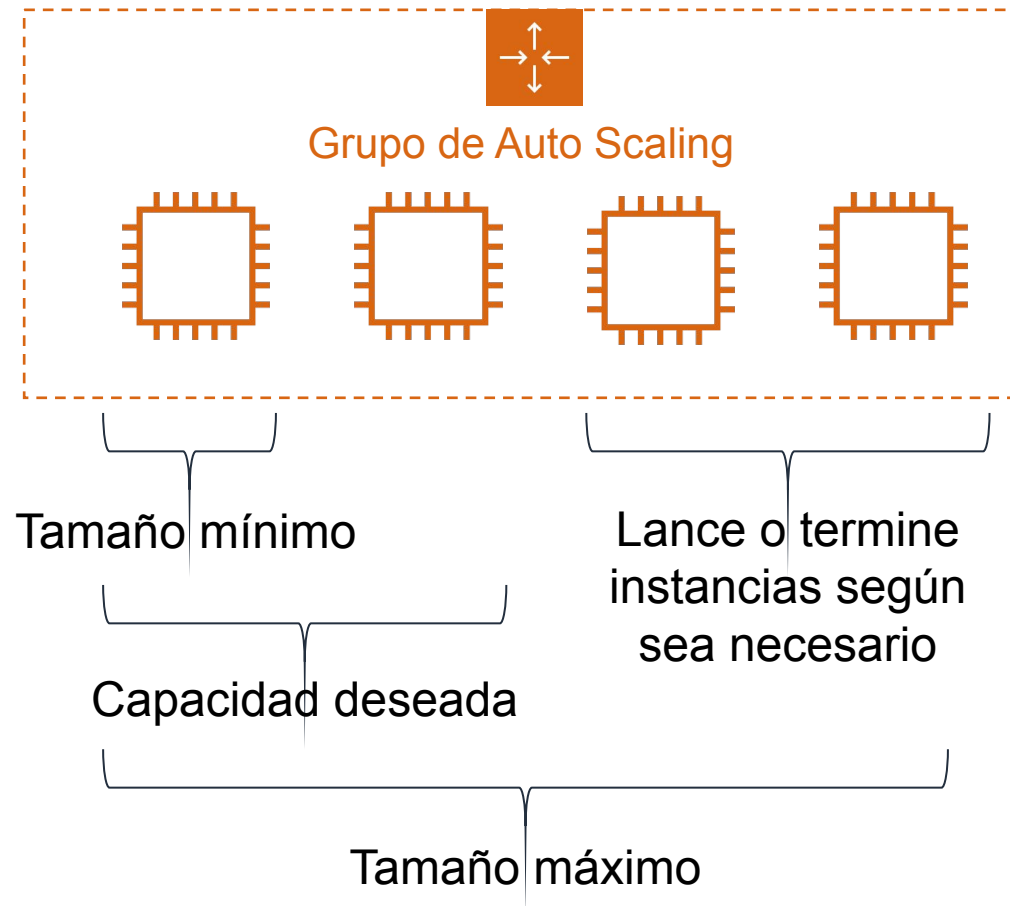
El desafío reside en acertar la cantidad desconocida de la capacidad de cómputo que se precisa.

76
por
cie
nto

24
por
cie

Noviembr
e

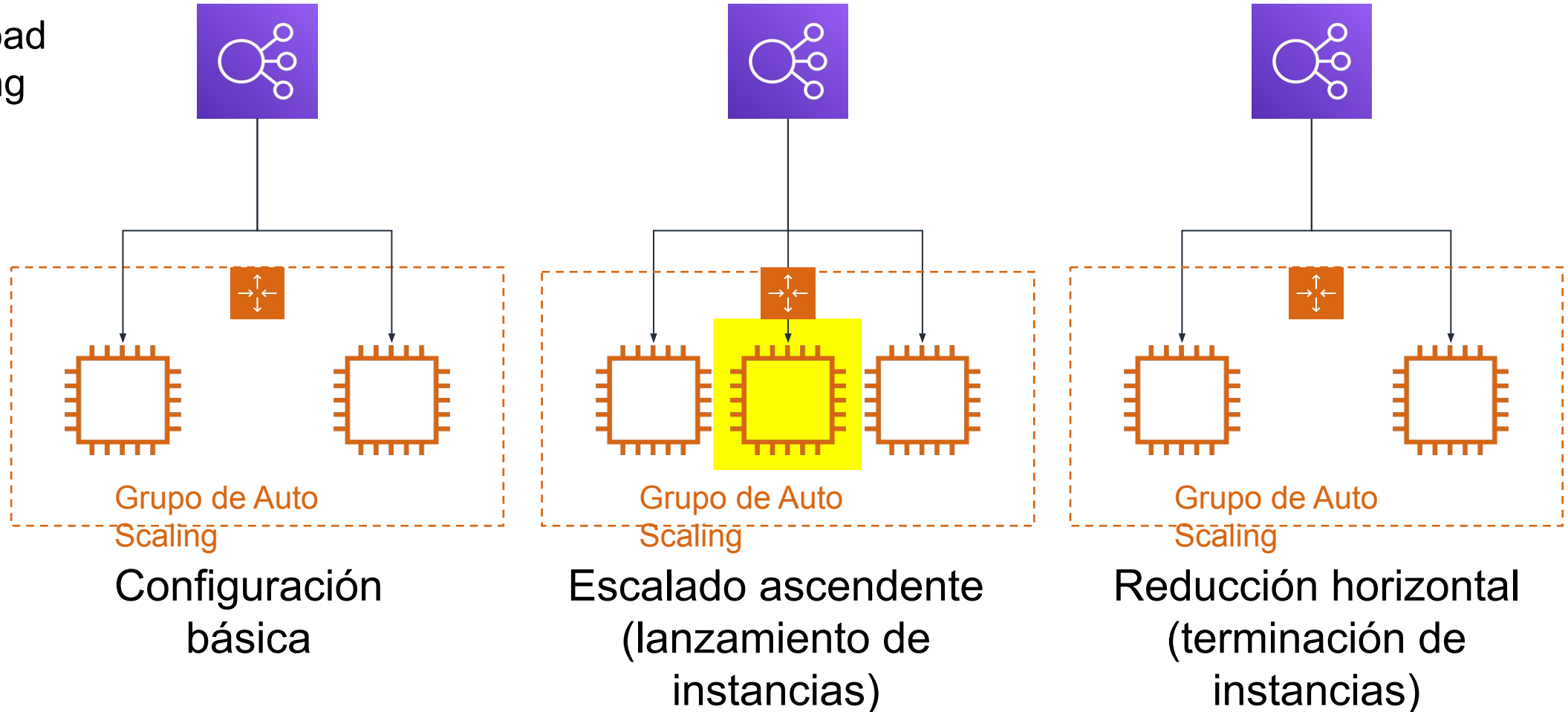
Grupos de Auto Scaling



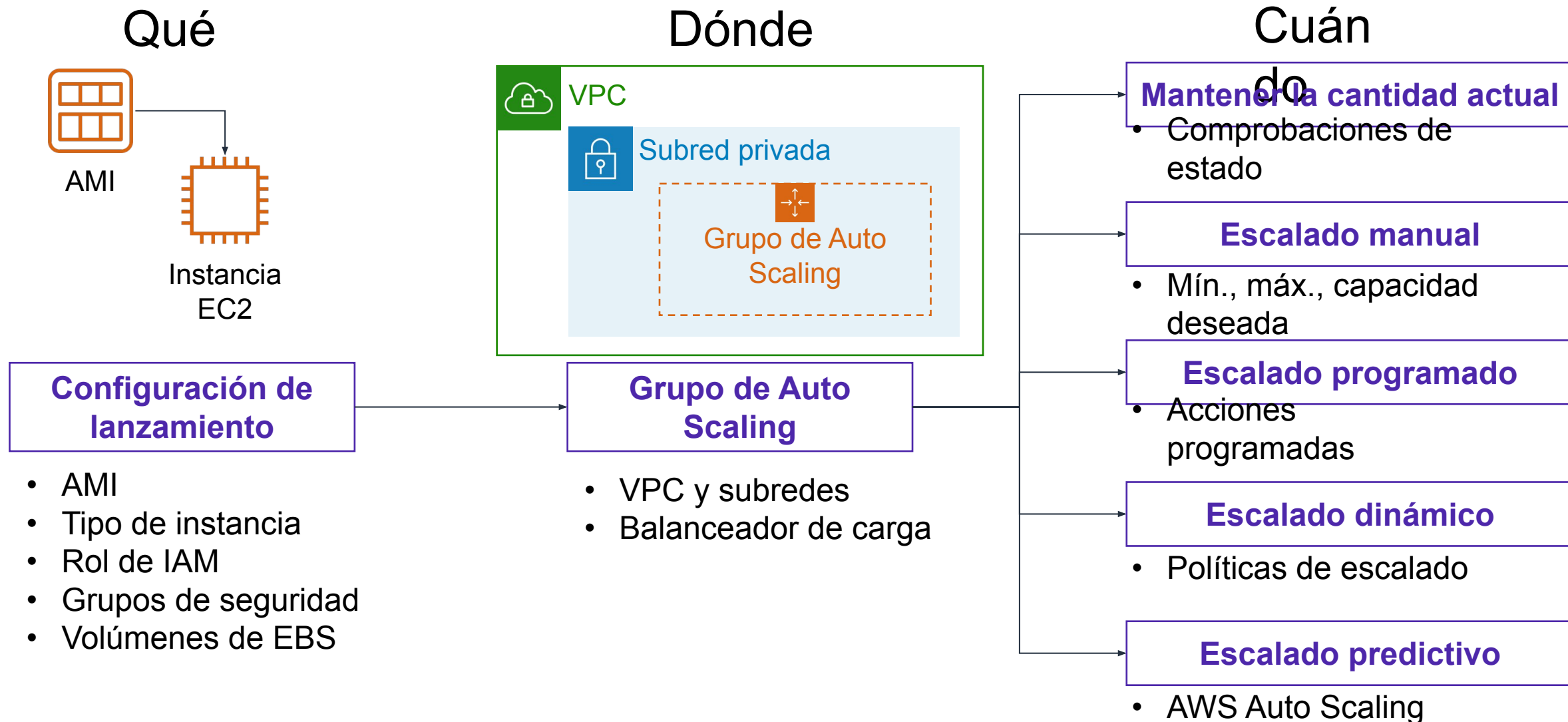
Un **grupo de Auto Scaling** consiste en una colección de instancias EC2 que se tratan como una agrupación lógica a efectos de la administración y el escalado automático.

Comparación entre el escalado ascendente y el escalado descendente

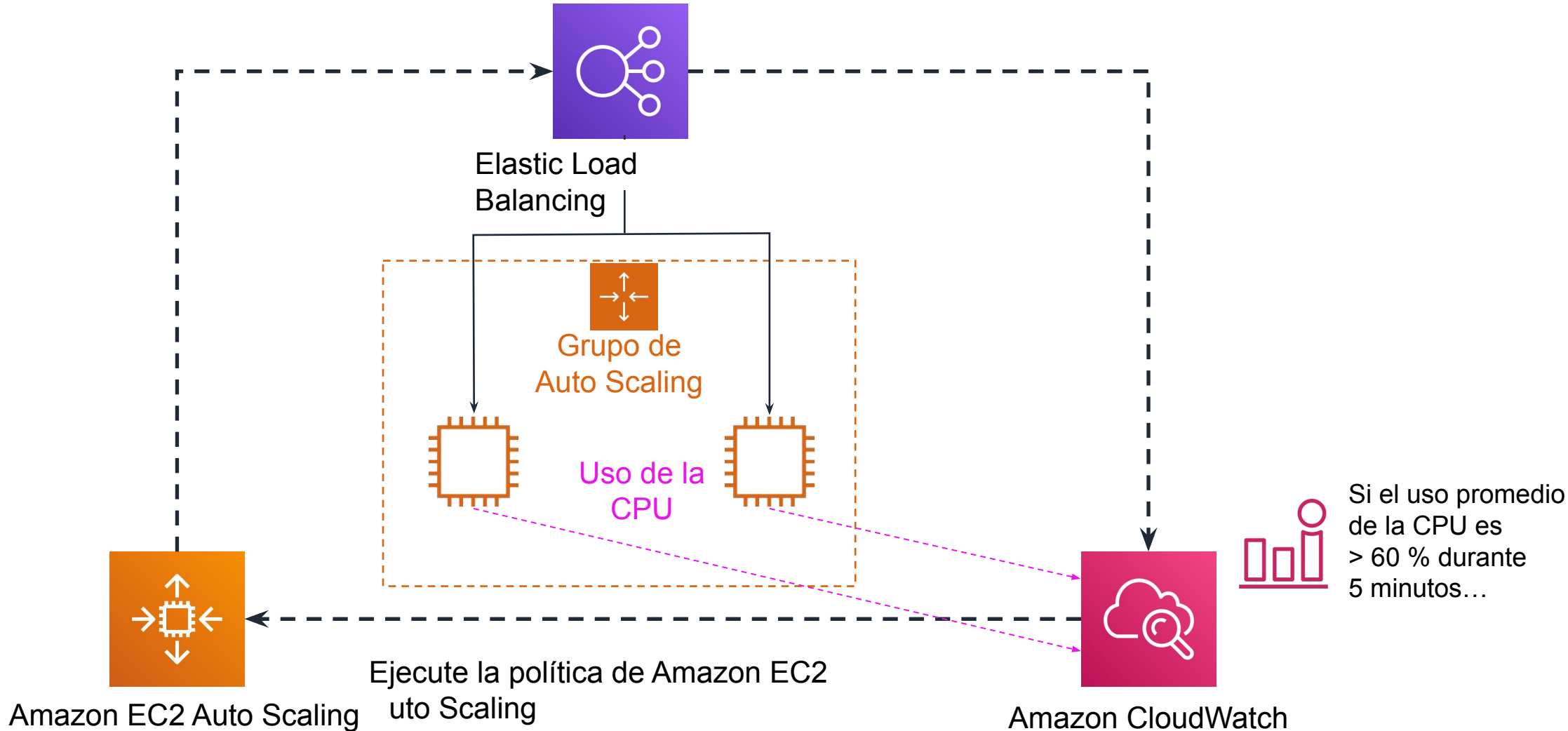
Elastic Load
Balancing

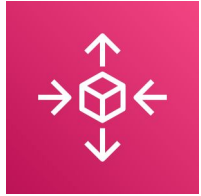


Cómo funciona Amazon EC2 Auto Scaling



Implementación del escalado dinámico





AWS Auto Scaling

- Monitorea sus aplicaciones y ajusta automáticamente la capacidad para mantener un rendimiento estable y predecible al menor costo posible.
- Proporciona una interfaz de usuario simple y potente que le permite crear planes de escalado para los siguientes recursos:
 - Instancias de Amazon EC2 y flotas de spot
 - Tareas de Amazon Elastic Container Service (Amazon ECS)
 - Tablas e índices de Amazon DynamoDB
 - Réplicas de Amazon Aurora

Aprendizajes clave de la sección 3

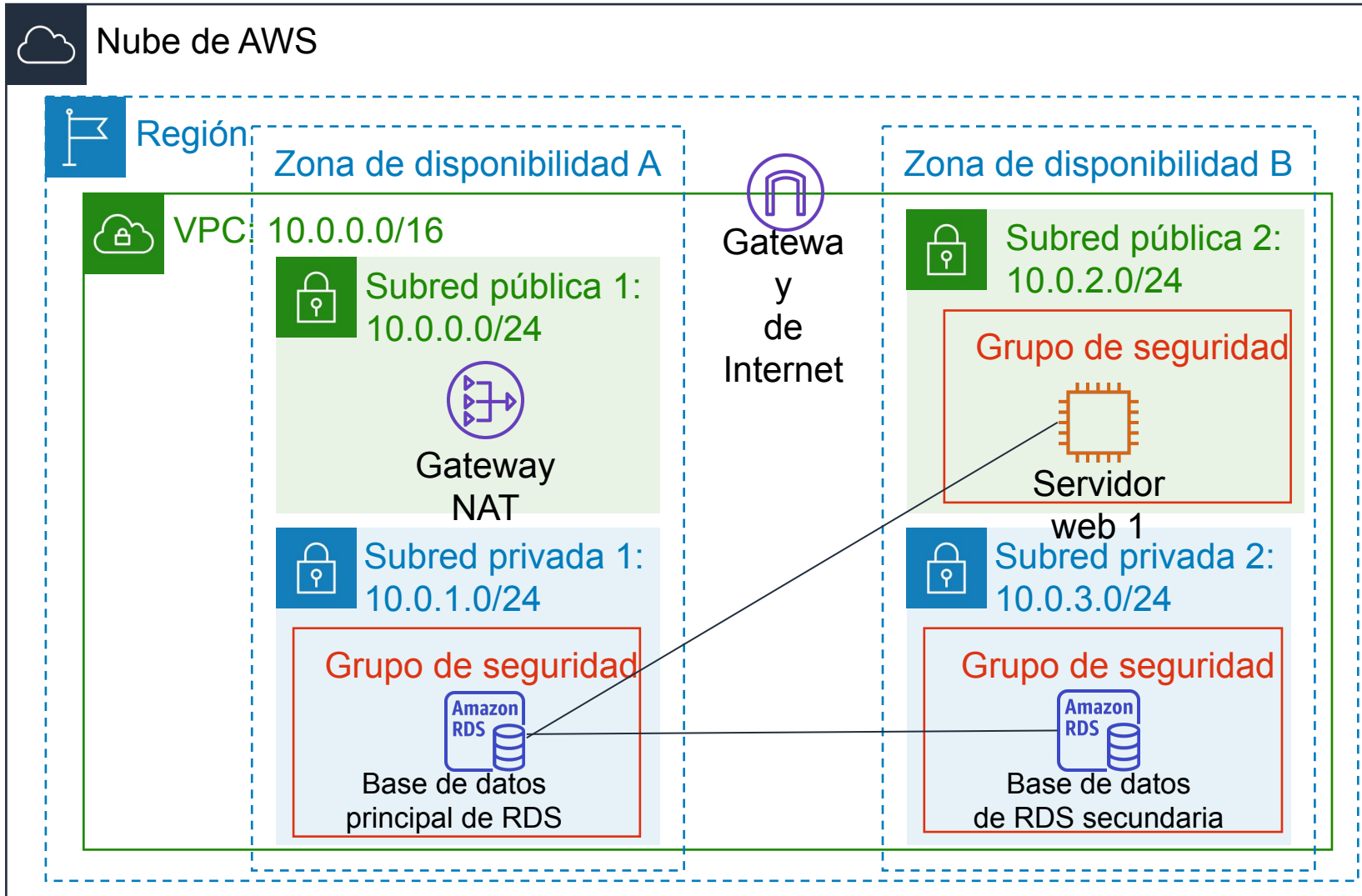


- El escalado le permite responder rápidamente a los cambios en las necesidades de recursos.
- Amazon EC2 Auto Scaling mantiene la disponibilidad de las aplicaciones agregando o eliminando instancias EC2 de manera automática.
- Un grupo de Auto Scaling es una colección de instancias EC2.
- Una configuración de lanzamiento es una plantilla de configuración de instancias.
- El escalado dinámico utiliza Amazon EC2 Auto Scaling, CloudWatch y Elastic Load Balancing.
- AWS Auto Scaling es un servicio independiente de Amazon EC2 Auto Scaling.

Laboratorio 6: Ajuste de la escala y balanceo de la carga de su arquitectura

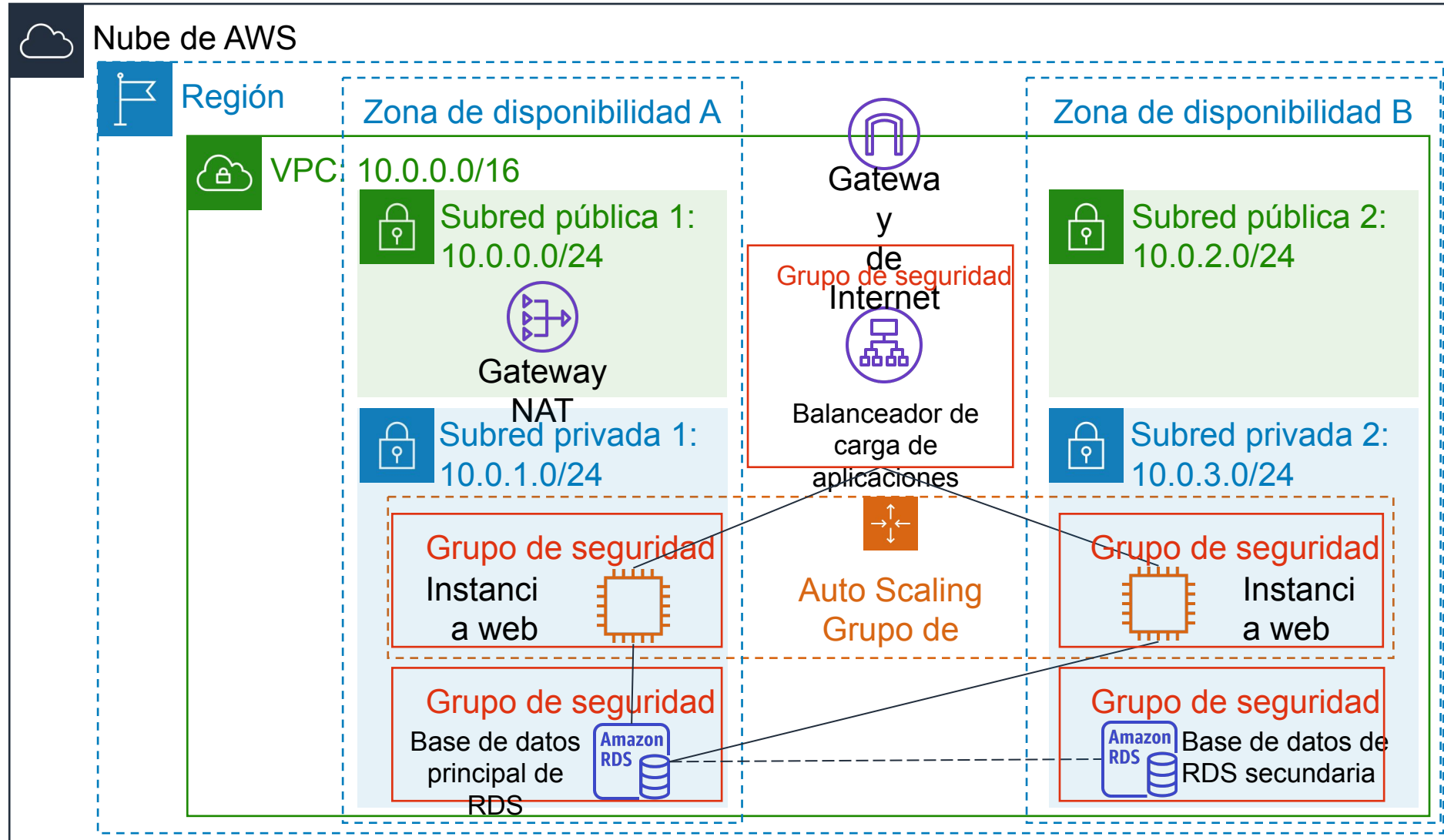


Laboratorio 6: situación



- Cree una imagen de Amazon Machine (AMI) a partir de una instancia en ejecución.
- Cree un balanceador de carga de aplicaciones.
- Cree una configuración de lanzamiento y un grupo de Auto Scaling.
- Escale automáticamente instancias nuevas dentro de una subred privada.
- Cree alarmas de Amazon CloudWatch y monitoree el rendimiento de la infraestructura.

Laboratorio 6: producto final





Aprox. 30 minutos



Comience el Laboratorio 6: Ajuste de la escala y balanceo de la carga de su arquitectura

Análisis posterior del laboratorio: Aprendizajes clave



Módulo 10: Monitoreo y escalado automático

Conclusión del módulo

En resumen, en este módulo, aprendió a hacer lo siguiente:

- Indicar cómo distribuir el tráfico entre las instancias de Amazon Elastic Compute Cloud (Amazon EC2) usando Elastic Load Balancing
- Identificar cómo Amazon CloudWatch le permite monitorear los recursos y las aplicaciones de AWS en tiempo real
- Explicar cómo Amazon EC2 Auto Scaling lanza y publica servidores en respuesta a los cambios en las cargas de trabajo
- Efectuar tareas de escalado y balanceo de carga para mejorar una arquitectura

Complete la revisión de conocimientos



Pregunta del examen de muestra

¿Qué servicio utilizaría para enviar alertas basadas en las alarmas de Amazon CloudWatch?

- A. Amazon Simple Notification Service
- B. AWS CloudTrail
- C. AWS Trusted Advisor
- D. Amazon Route 53

Gracias

© 2019, Amazon Web Services, Inc. o sus empresas afiliadas. Todos los derechos reservados. Este contenido no puede reproducirse ni redistribuirse, total ni parcialmente, sin el permiso previo por escrito de Amazon Web Services, Inc. Queda prohibida la copia, el préstamo o la venta de carácter comercial. Envíenos sus correcciones o comentarios relacionados con el curso a: aws-course-feedback@amazon.com. Si tiene cualquier otra duda, contacte con nosotros en: <https://aws.amazon.com/contact-us/aws-training/>. Todas las marcas comerciales pertenecen a sus propietarios.

