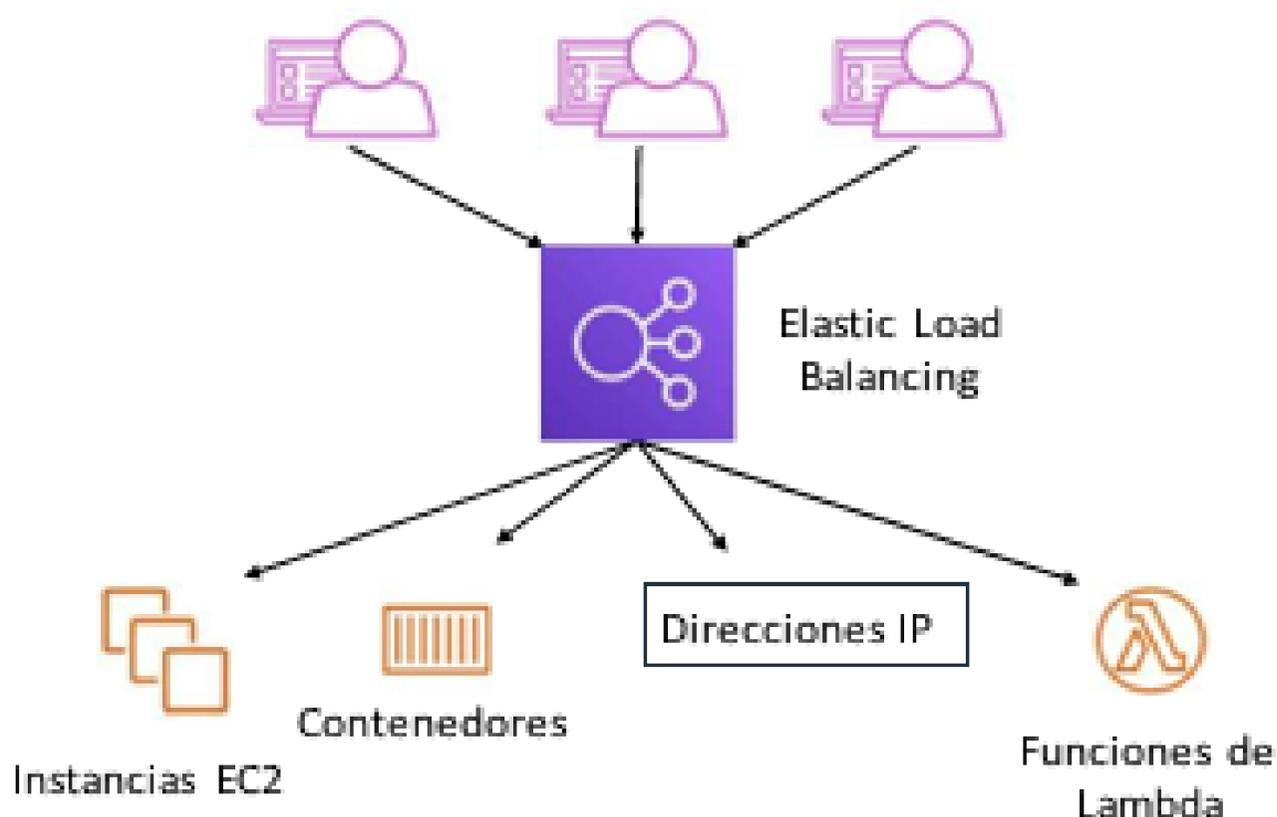


# LECCIÓN 1: INTRODUCCIÓN A LOS SERVICIOS DE AWS PARA ESCALABILIDAD Y MONITOREO ELASTIC LOAD BALANCING





**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.**

Los sitios web modernos de mucho tráfico atienden cientos de miles, si no millones, de solicitudes simultáneas de los usuarios o los clientes y, luego, devuelven el texto, las imágenes, los videos o los datos de las aplicaciones correctos de una manera rápida y confiable. Por lo general, se necesitan servidores adicionales para satisfacer estos altos volúmenes.

Elastic Load Balancing es un servicio de AWS que distribuye el tráfico entrante de las aplicaciones o la red entre varios destinos, como las instancias de Amazon Elastic Compute Cloud (Amazon EC2), los contenedores, las direcciones de protocolo de Internet (IP) y las funciones de Lambda, en una única zona de disponibilidad o en varias. Elastic Load Balancing escala el balanceador de carga a medida que el tráfico dirigido a la aplicación cambia con el tiempo. Puede escalar automáticamente a la mayoría de las cargas de trabajo.

## TIPOS DE BALANCEADORES DE CARGA

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

### Elastic Load Balancing está disponible en tres tipos:

**El balancedor de carga de red** opera en el nivel de transporte de la red (capa 4 del modelo OSI) y dirige las conexiones a los destinos (instancias EC2, microservicios y contenedores) en función de los datos del protocolo IP. Funciona bien para balancear la carga del tráfico del protocolo de control de transmisión (TCP) y del protocolo de datagramas de usuario (UDP). El balancedor de carga de red es capaz de gestionar millones de solicitudes por segundo mientras mantiene latencias ultrabajas. El balancedor de carga de red está optimizado para gestionar patrones de tráfico de red repentinos y volátiles.

**El balancedor de carga clásico** proporciona balancedo de carga básico en varias instancias EC2 y opera tanto en el nivel de aplicación como en el de transporte de red. Un balancedor de carga clásico admite el balancedo de carga de las aplicaciones que utilizan HTTP, HTTPS, TCP y SSL. El balancedor de carga clásico es una implementación más antigua. Cuando sea posible, AWS recomienda utilizar un balancedor de carga de aplicaciones o un balancedor de carga de red dedicados.

**El balanceador de carga de aplicaciones** opera en el nivel de la aplicación (capa 7 del modelo de interconexión de sistemas abiertos u OSI). Dirige el tráfico a los destinos (instancias de Amazon Elastic Compute Cloud [Amazon EC2], contenedores, direcciones de protocolo de Internet [IP] y funciones de Lambda) en función del contenido de la solicitud. Es ideal para el balanceo de carga avanzado del tráfico del protocolo de transferencia de hipertexto (HTTP) y del HTTP seguro (HTTPS). El balanceador de carga de aplicaciones brinda direccionamiento de solicitudes avanzadas a su entrega en arquitecturas modernas de las aplicaciones, como los microservicios y las aplicaciones basadas en contenedores. El balanceador de carga de aplicaciones simplifica y mejora la seguridad de sus aplicaciones al garantizar que se utilicen en todo momento los protocolos y cifrados de la capa de conexión segura (SSL) o de la seguridad de la capa de transporte (TLS) más recientes.

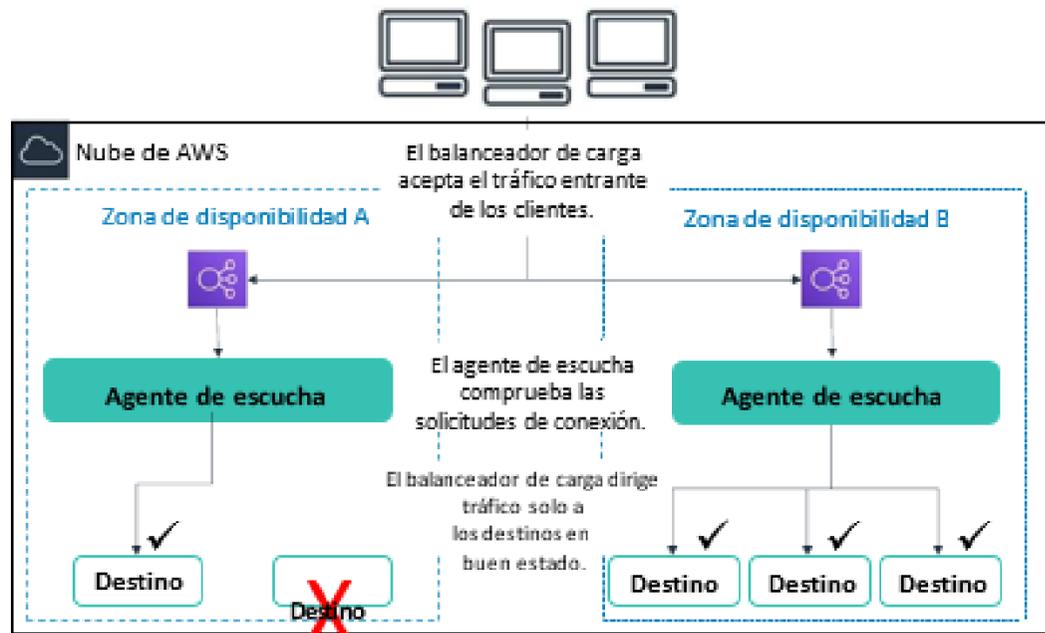


Para obtener más información acerca de las diferencias entre los tres tipos de balanceadores de carga, consulte [Comparaciones de productos](#) en la [página de características](#) de Elastic Load Balancing.

## 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.



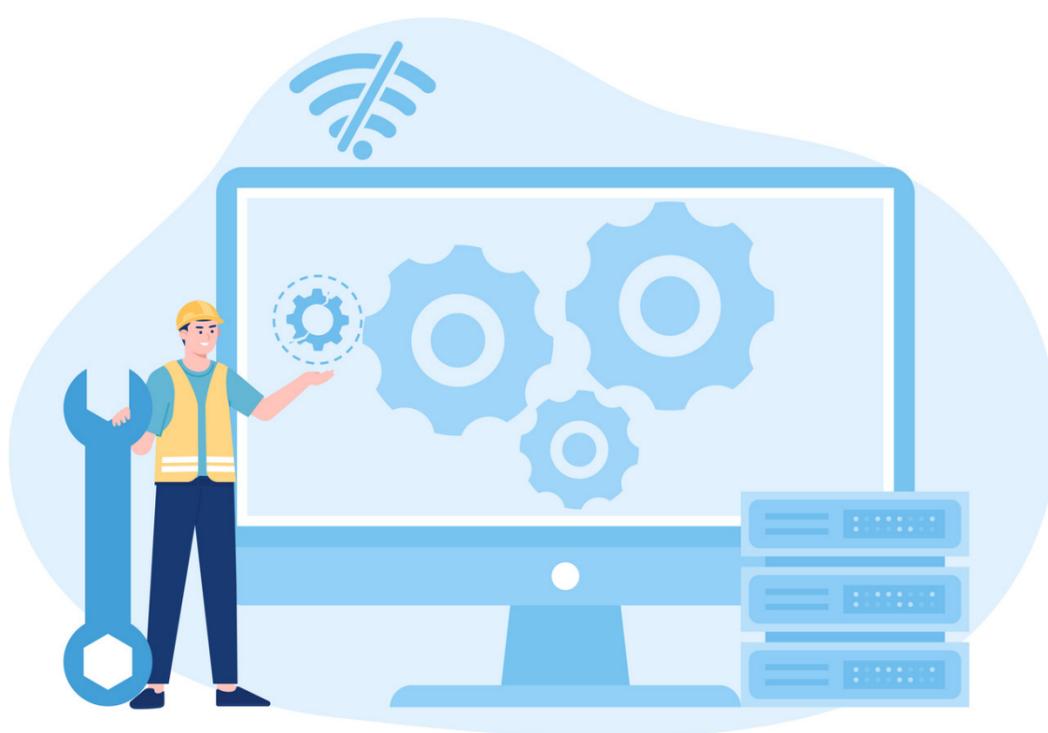
Un balanceador de carga acepta el tráfico entrante de los clientes y dirige las solicitudes a sus destinos registrados (como, por ejemplo, las instancias EC2) en una o más zonas de disponibilidad.

Puede configurar el balanceador de carga para que acepte el tráfico entrante especificando uno o más agentes de escucha. Un agente de escucha es un proceso que verifica las solicitudes de conexión. Se configura con un protocolo y un número de puerto para las conexiones de los clientes al balanceador de carga. De modo similar, se configura con un protocolo y un número de puerto para las conexiones del balanceador de carga a los destinos.

También puede configurar el balanceador de carga para que realice comprobaciones de estado, que se utilizan para monitorear el estado de los destinos registrados de manera que el balanceador de carga solo envíe solicitudes a las instancias que estén en buen estado.

Cuando el balanceador de carga detecta que un destino no está en buen estado, deja de enviar tráfico a ese destino. Luego, cuando detecta que está nuevamente en buen estado, reanuda el tráfico hacia ese destino.

Hay una diferencia clave en el modo en que se configuran los tipos de balanceador de carga. 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.



## CASOS DE USO DE ELASTIC LOAD BALANCING

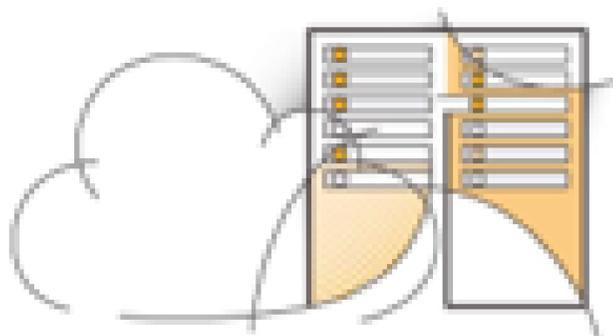
APLICACIONES DE ALTA  
DISPONIBILIDAD  
YTOLERANTES A ERRORES



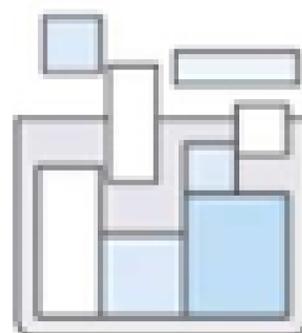
NUBE VIRTUAL  
PRIVADA (VPC)



ENTORNOS HÍBRIDOS



APLICACIONES EN  
CONTENEDORES



ELASTICIDAD Y  
ESCALABILIDAD



INVOCACIÓN DE LAS  
FUNCIONES DE LAMBDA A  
TRAVÉS DE HTTP(S)



Hay muchos motivos para utilizar un balanceador de carga, entre los que se incluyen los siguientes:

## **Lograr alta disponibilidad y mejor tolerancia a errores para sus aplicaciones:**

Elastic Load Balancing balancea el tráfico en los destinos en buen estado en varias zonas de disponibilidad. Si uno o varios de los destinos de una sola zona de disponibilidad presentan errores, Elastic Load Balancing dirigirá el tráfico a los destinos en buen estado de otras zonas de disponibilidad. Una vez que los destinos vuelven a un estado correcto, el balanceador de carga reanuda de forma automática el tráfico dirigido hacia ellos.

## **Balancear automáticamente la carga de las aplicaciones en contenedores:**

Gracias a la compatibilidad mejorada de los contenedores con Elastic Load Balancing, ahora puede balancear la carga entre varios puertos de la misma instancia EC2. También puede aprovechar la integración profunda con Amazon Elastic Container Service (Amazon ECS), que ofrece contenedores completamente administrados. Solo debe registrar un servicio con un balanceador de carga y Amazon ECS administrará de forma transparente el registro y la anulación del registro de los contenedores de Docker. El balanceador de carga detecta el puerto de manera automática y se reconfigura dinámicamente.

## Escalar automáticamente sus aplicaciones:

Elastic Load Balancing trabaja con Amazon CloudWatch y Amazon EC2 Auto Scaling para ayudarlo a escalar las aplicaciones según las demandas de sus clientes. Las alarmas de Amazon CloudWatch pueden activar el escalado automático para su flota de instancias EC2 cuando la latencia de cualquiera de sus instancias EC2 supere un límite preconfigurado. Luego, Amazon EC2 Auto Scaling aprovisiona nuevas instancias y, así, sus aplicaciones estarán listas para atender la siguiente solicitud del cliente. El balanceador de carga registrará la instancia EC2 y dirigirá el tráfico hacia ella según sea necesario.

## Utilizar Elastic Load Balancing en su nube virtual privada (VPC):

Puede utilizar Elastic Load Balancing para crear un punto de entrada público en su VPC o para dirigir el tráfico de solicitudes entre las capas de su aplicación dentro de la VPC. Puede asignar grupos de seguridad al balanceador de carga a fin de controlar qué puertos están abiertos para una lista de fuentes permitidas. Como Elastic Load Balancing trabaja con su VPC, todas las listas existentes de control de acceso a la red (ACL a la red) y las tablas de direccionamiento continúan ofreciendo controles de red adicionales. En el momento de crear un balanceador de carga en su VPC, puede especificar si será público (opción predeterminada) o interno. Si selecciona interno, no necesitará disponer de una gateway de Internet para llegar al balanceador de carga y sus direcciones IP privadas se utilizarán en el registro de su sistema de nombres de dominio (DNS).

## Habilitar el balanceo de carga híbrido:

Elastic Load Balancing le permite balancear la carga entre los recursos en las instalaciones y en AWS con el mismo balanceador de carga. Por ejemplo, si debe distribuir el tráfico de las aplicaciones entre los recursos de AWS y los que están en las instalaciones, puede registrar todos los recursos en el mismo grupo de destino y asociarlo con el balanceador de carga. De manera alternativa, puede usar el balanceo de cargas ponderado basado en DNS entre los recursos de AWS y en las instalaciones mediante el uso de dos balanceadores de carga, uno para los recursos de AWS y el otro para los recursos en las instalaciones. También puede utilizar el balanceo de carga híbrido para beneficiar aplicaciones independientes en las que una se encuentra en una VPC y la otra en las instalaciones. Coloque los destinos de la VPC en un grupo de destino y los destinos en las instalaciones en otro grupo y, a continuación, utilice el direccionamiento basado en el contenido para dirigir el tráfico a cada grupo de destinos.

## Invocar las funciones de Lambda a través de HTTP(S):

Elastic Load Balancing admite la invocación de funciones de Lambda para atender solicitudes HTTP(S). Esto permite a los usuarios acceder a aplicaciones sin servidor desde cualquier cliente HTTP, incluidos los navegadores web. Puede registrar las funciones de Lambda como destinos y utilizar el soporte para las reglas de direccionamiento basado en el contenido en los balanceadores de carga de aplicaciones a fin de dirigir las solicitudes a distintas funciones de Lambda. Puede utilizar el balanceador de carga de aplicaciones como un punto de enlace HTTP común para las aplicaciones que utilizan servidores e informática sin servidor. Puede crear un sitio web completo utilizando funciones de Lambda o combinar instancias EC2, contenedores, servidores en las instalaciones y funciones de Lambda para crear aplicaciones.

## ACTIVIDAD: ELASTIC LOAD BALANCING

Debe admitir el tráfico a una aplicación en contenedores.	Balancedador de carga de aplicaciones
El tráfico TCP es extremadamente irregular e impredecible.	Balancedador de carga de red
Necesita balanceo de carga simple con múltiples protocolos.	Balancedador de carga clásico
Debe admitir una dirección IP estática o elástica o un destino IP fuera de la VPC.	Balancedador de carga de red
Necesita un balancedador de carga que pueda gestionar millones de solicitudes por segundo mientras mantiene latencias bajas.	Balancedador de carga de red
Debe admitir solicitudes HTTPS.	Balancedador de carga de aplicaciones

Para esta actividad, mencione el balancedador de carga que usaría para la situación presentada.

## 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 balancedador 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.



Puede utilizar las siguientes características para monitorear los balanceadores de carga, analizar los patrones de tráfico y solucionar los problemas de los balanceadores de carga y los destinos:

### **Métricas de Amazon CloudWatch:**

Elastic Load Balancing publica puntos de datos en Amazon CloudWatch para los balanceadores de carga y los destinos. CloudWatch le permite recuperar las estadísticas acerca de esos puntos de datos en conjuntos ordenados de datos de serie temporal, denominados métricas. Puede utilizar las métricas para comprobar si el sistema funciona según lo esperado. Por ejemplo, puede crear una alarma de CloudWatch para monitorear una métrica determinada e iniciar una acción (como, por ejemplo, enviar una notificación a una dirección de email) si la métrica no está comprendida dentro del rango que considera aceptable.

### **Registros de acceso:**

Puede utilizar registros de acceso para registrar información detallada acerca de las solicitudes que se realizaron al balanceador de carga y almacenarla como archivos de registro en Amazon Simple Storage Service (Amazon S3). Puede utilizar estos registros de acceso para analizar los patrones de tráfico y solucionar los problemas en los destinos o en las aplicaciones de backend.

### **Registros de AWS CloudTrail:**

Puede utilizar AWS CloudTrail para registrar información detallada acerca de las llamadas que se realizaron a la interfaz de programación de aplicaciones (API) de Elastic Load Balancing y almacenarla como archivos de registro en Amazon S3. Puede utilizar estos registros de CloudTrail para determinar quién realizó la llamada, qué llamadas se efectuaron, cuándo se realizó la llamada, la dirección IP de origen de donde procedió la llamada, etcétera.

## ESTOS SON ALGUNOS DE LOS APRENDIZAJES CLAVE DE ESTA LECCIÓN:

- **Elastic Load Balancing distribuye el tráfico entrante de las aplicaciones o de la red entre varios destinos (como las instancias de Amazon EC2, los contenedores, las direcciones IP y las funciones de Lambda) 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**
- **Elastic Load Balancing ofrece varias herramientas de monitoreo para el registro y monitoreo continuos destinados a las auditorías y los análisis.**

