閘道負載平衡器

Elastic Load Balancing

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Elastic Load Balancing: 閘道負載平衡器

Copyright © 2025 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon 的商標和商業外觀不得用於任何非 Amazon 的產品或服務，也不能以任何可能造成客戶混淆、任何貶低或使 Amazon 名譽受損的方式使用 Amazon 的商標和商業外觀。所有其他非 Amazon 擁有的商標均為其各自擁有者的財產，這些擁有者可能附屬於 Amazon，或與 Amazon 有合作關係，亦或受到 Amazon 贊助。

Table of Contents

什麼是 Gateway Load Balancer？ 1
Gateway Load Balancer 概觀 1
應用裝置廠商 2
開始使用 2
定價 2
開始使用 3
概要 3
路由 5
先決條件 6
步驟 1：建立 Gateway Load Balancer 6
步驟 2：建立 Gateway Load Balancer 端點服務 7
步驟 3：建立 Gateway Load Balancer 端點 8
步驟 4：設定路由 9
使用 CLI 入門 11
概要 11
路由 5
先決條件 14
步驟 1：建立 Gateway Load Balancer 並註冊目標 14
步驟 2：建立 Gateway Load Balancer 端點 16
步驟 3：設定路由 17
閘道負載平衡器 19
負載平衡器狀態 19
IP 地址類型 20
可用區域 20
閒置逾時 21
負載平衡器屬性 21
網路 ACL 22
非對稱流程 22
網路最大傳輸單位 (MTU) 22
建立負載平衡器 22
先決條件 23
建立負載平衡器 23
重要的後續步驟 24
更新 IP 地址類型 24

編輯負載平衡器屬性 24
刪除保護 25
跨區域負載平衡 25
標記負載平衡器 26
刪除負載平衡器 27
LCU 保留 27
請求保留 28
更新或終止保留 29
監控保留 30
接聽程式 31
接聽程式屬性 31
更新接聽程式目標群組 31
更新閒置逾時 32
目標群組 33
路由組態 33
Target type (目標類型) 34
已登記的目標 34
目標群組屬性 35
建立目標群組 36
設定運作狀態檢查 37
運作狀態檢查設定 37
目標運作狀態 38
運作狀態檢查原因代碼 39
目標失敗案例 40
檢查目標的運作狀態 41
修改運作狀態檢查設定 42
編輯目標群組屬性 42
目標容錯移轉 42
取消登記的延遲 44
流程黏性 44
登記目標 45
考量事項 46
目標安全群組 46
網路 ACL 46
依執行個體 ID 註冊目標 46
依 IP 地址註冊目標 47

取消註冊目標 48
標記目標群組 48
刪除目標群組 49
監控負載平衡器 50
CloudWatch 指標 50
Gateway Load Balancer 指標 51
Gateway Load Balancer 的指標維度 54
查看 Gateway Load Balancer 的 CloudWatch 指標 55
配額 57
負載平衡器 57
目標群組 57
Load Balancer容量單位 58
文件歷史紀錄 59
..... lxi

什麼是 Gateway Load Balancer？

Elastic Load Balancing 會自動將傳入流量分配到一或多個可用區域中的多個目標。其會監控已註冊目標的運作狀態，並且僅將流量路由至運作狀態良好的目標。當傳入流量隨著時間發生變化，Elastic Load Balancing 會擴展您的負載平衡器。他可以自動擴展以因應絕大多數的工作負載。

Elastic Load Balancing 支援下列負載平衡器：Application Load Balancer、Network Load Balancer、Gateway Load Balancer 和 Classic Load Balancer。您可以選取最符合您需要的負載平衡器類型。本指南主要探討 Gateway Load Balancer。如需其他負載平衡器的詳細資訊，請參閱 Application Load Balancer 使用者指南、Network Load Balancer 使用者指南和 Classic Load Balancer 使用者指南。

Gateway Load Balancer 概觀

Gateway Load Balancer 可讓您部署、擴展與管理虛擬應用裝置，如防火牆、入侵偵測與預防系統，以及深層封包檢查系統。它結合透明網路閘道 (亦即，所有流量的單一進入和出口點) 並分配流量，同時依據需求擴展您的虛擬應用裝置。

Gateway Load Balancer 在開放系統互相連線 (OSI) 模型的第三層，網路層運作。它會接聽所有連接埠的全部 IP 封包，並轉送流量至接聽程式規則中指定的目標群組。它使用 5 個元組 （預設）、3 個元組或 2 個元組維持特定目標設備的流程黏性。Gateway Load Balancer 及其註冊的虛擬應用裝置執行個體會在連接埠 6081 上使用 GENEVE 通訊協定交換應用程式流量。

Gateway Load Balancer 使用閘道負載平衡器端點，以便跨 VPC 邊界安全交換流量。Gateway Load Balancer 端點是一種 VPC 端點，它在服務提供者 VPC 的虛擬應用裝置和服務消費者 VPC 的應用程式伺服器之間提供私有連線。您要在與虛擬應用裝置相同的 VPC 中部署 Gateway Load Balancer。您要為 Gateway Load Balancer 將虛擬應用裝置註冊到目標群組。

使用路由表來設定往返 Gateway Load Balancer 端點的流量。流量從服務消費者 VPC 透過 Gateway Load Balancer 端點流向服務提供者 VPC 中的 Gateway Load Balancer，然後傳回服務消費者 VPC。您必須在不同的子網路中建立 Gateway Load Balancer 端點和應用程式伺服器。這可讓您將 Gateway Load Balancer 端點設定為應用程式子網路的路由表中的下一個躍點。

如需詳細資訊，請參閱 AWS PrivateLink 指南中的透過 AWS PrivateLink存取虛擬應用裝置。

應用裝置廠商

您有責任從應用裝置供應商選擇和鑑定軟體。您必須信任應用裝置軟體，才能檢查或修改來自負載平衡器的流量。列為 Elastic Load Balancing 合作夥伴的設備供應商已與其設備軟體整合並取得資格 AWS。您可以對此清單中廠商的應用裝置軟體提供更高的信任度。但是， AWS 不保證這些廠商的軟體安全性或可靠性。

開始使用

若要使用 建立 Gateway Load Balancer AWS Management Console，請參閱 開始使用。若要使用 建立 Gateway Load Balancer AWS Command Line Interface，請參閱 使用 CLI 入門。

定價

使用負載平衡器時，您只需按實際用量付費。如需詳細資訊，請參閱 Elastic Load Balancing 定價。

開始使用 Gateway Load Balancer

Gateway Load Balancer 可讓您輕鬆部署、擴展與管理第三方虛擬應用裝置，如安全應用裝置。

在本教學課程中，我們將使用 Gateway Load Balancer 和 Gateway Load Balancer 端點來實作檢查系統。

目錄

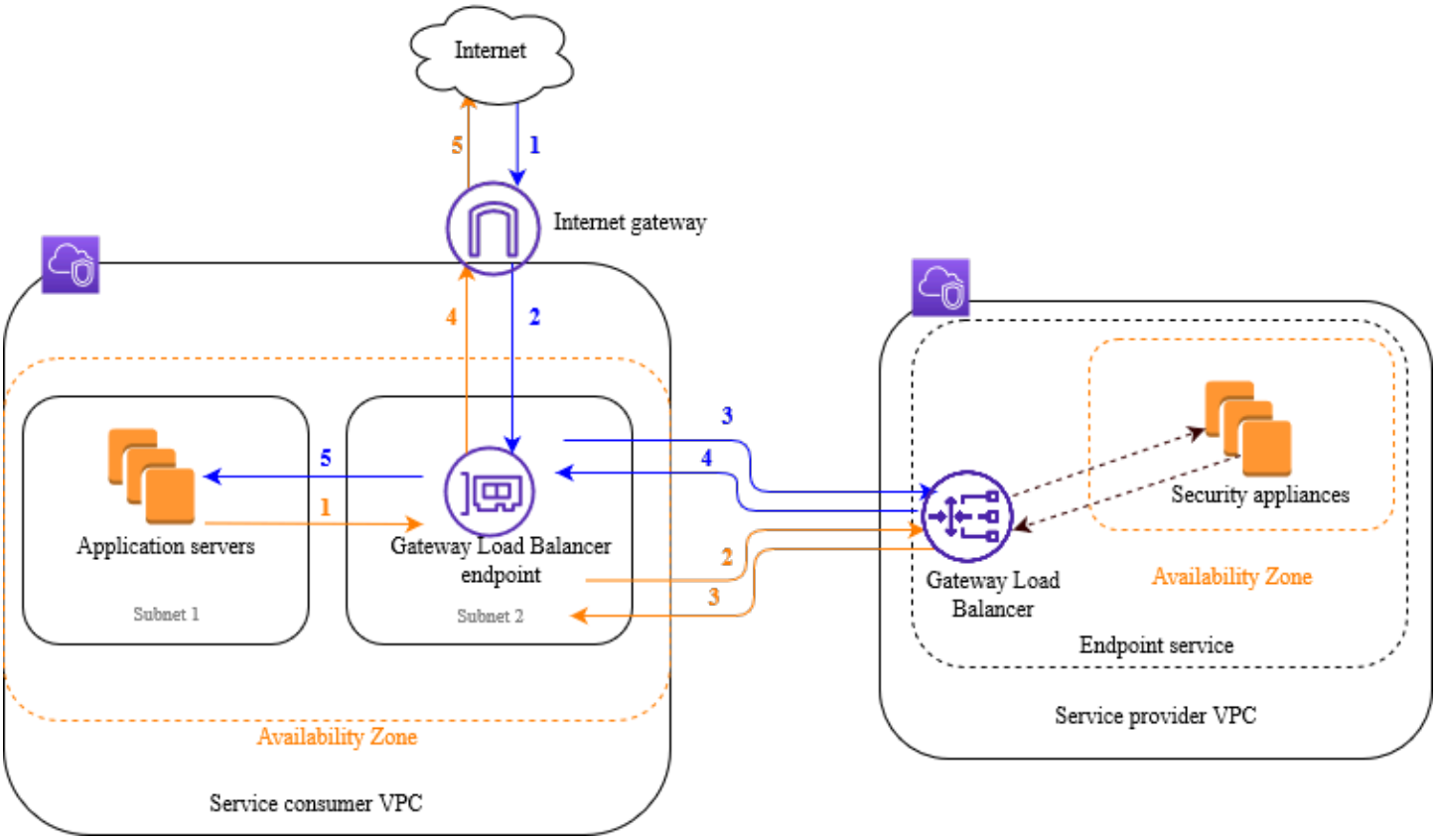
- 概要
- 先決條件
- 步驟 1：建立 Gateway Load Balancer
- 步驟 2：建立 Gateway Load Balancer 端點服務
- 步驟 3：建立 Gateway Load Balancer 端點
- 步驟 4：設定路由

概要

Gateway Load Balancer 端點是一種 VPC 端點，它在服務提供者 VPC 的虛擬應用裝置和服務消費者 VPC 的應用程式伺服器之間提供私有連線。在與虛擬應用裝置相同的 VPC 中部署 Gateway Load Balancer。這些虛擬應用裝置作為 Gateway Load Balancer 的目標群組註冊。

應用程式伺服器在服務消費者 VPC 中的一個子網路 (目的地子網路) 中執行，而 Gateway Load Balancer 端點位於相同 VPC 的另一個子網路中。所有透過網際網路閘道進入服務消費者 VPC 的流量會先路由至 Gateway Load Balancer 端點，然後再路由至目的地子網路。

同樣，離開應用程式伺服器 (目的地子網路) 的所有流量會路由至 Gateway Load Balancer 端點，然後再路由回網際網路。下列網路圖是如何使用 Gateway Load Balancer 端點存取端點服務的視覺化表示。



下面的帶編號項目突出顯示和解釋在前述網路圖中顯示的元素。

從網際網路到應用程式的流量 (藍色箭頭)：

1. 流量透過網際網路閘道進入服務消費者 VPC。
2. 作為傳入路由的結果，流量傳送至 Gateway Load Balancer 端點。
3. 流量傳送至 Gateway Load Balancer，後者將此流量分發至其中一個安全應用裝置。
4. 流量會在由安全應用裝置檢查之後傳回到 Gateway Load Balancer 端點。
5. 流量傳送至應用程式伺服器 (目的地子網路)。

從應用程式到網際網路的流量 (橙色箭頭)：

1. 作為在應用程式伺服器子網路上設定的預設路由的結果，流量傳送至 Gateway Load Balancer 端點。
2. 流量傳送至 Gateway Load Balancer，後者將此流量分發至其中一個安全應用裝置。
3. 流量會在由安全應用裝置檢查之後傳回到 Gateway Load Balancer 端點。
4. 根據路由表組態，將流量傳送至網際網路閘道。

5. 流量會傳回網際網路。

路由

網際網路閘道的路由表必須具有相應條目，即將目的地為應用程式伺服器的流量傳送至 Gateway Load Balancer 端點。若要指定 Gateway Load Balancer 端點，請使用 VPC 端點的 ID。如下範例顯示雙堆疊組態的路由。

目的地	目標
VPC A IPv4 CIDR	區域
VPC A IPv6 CIDR	區域
### 1 IPv4 CIDR	*vpc-endpoint-id*
### 1 IPv6 CIDR	*vpc-endpoint-id*

帶應用程式伺服器的子網路的路由表必須具有相應條目，即將應用程式伺服器傳出的所有流量路由至 Gateway Load Balancer 端點。

目的地	目標
VPC A IPv4 CIDR	區域
VPC A IPv6 CIDR	區域
0.0.0.0/0	*vpc-endpoint-id*
::/0	*vpc-endpoint-id*

帶 Gateway Load Balancer 端點的子網路的路由表必須將從檢查傳回的流量路由至其最終目的地。對於源自網際網路的流量，本機路由會確保流量到達應用程式伺服器。對於源自應用程式伺服器的流量，請新增相應條目，即將所有流量路由至網際網路閘道。

目的地	目標
VPC A IPv4 CIDR	區域
VPC A IPv6 CIDR	區域
0.0.0.0/0	*internet-gateway-id*
::/0	*internet-gateway-id*

先決條件

- 確定服務消費者 VPC 對於包含應用程式伺服器的每個可用區域至少有兩個子網路。一個子網路用於 Gateway Load Balancer 端點，另一個子網用於應用程式伺服器。
- Gateway Load Balancer 和目標可以位於相同的子網路中。
- 您無法使用從其他帳戶共用的子網路來部署 Gateway Load Balancer。
- 在服務消費者 VPC 中的每個安全應用裝置子網路內啟動至少一個安全應用裝置執行個體。這些執行個體的安全群組必須允許連接埠 6081 上的 UDP 流量。

步驟 1：建立 Gateway Load Balancer

使用下列程序建立您的負載平衡器、接聽程式和目標群組。

使用主控台建立負載平衡器、接聽程式和目標群組

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選擇 Create load balancer (建立負載平衡器)。
4. 在 Gateway Load Balancer 下，選擇建立。
5. 基本組態
   a. 針對 Load balancer name (負載平衡器名稱)，輸入負載平衡器的名稱。
   b. 針對 IP 地址類型，選擇 IPv4 以僅支援 IPv4 地址，或選擇 Dualstack 以同時支援 IPv4 和 IPv6 地址。
6. 網路映射

a. 對於 VPC，請選取服務提供者 VPC。

b. 對於映射，選取您在其中啟動安全應用裝置執行個體的所有可用區域，以及每個可用區域選取一個子網路。

7. IP 接聽程式路由

a. 對於預設動作，請選取要接收流量的現有目標群組。此目標群組必須使用 GENEVE 通訊協定。

如果您沒有目標群組，請選擇 建立目標群組，這會在瀏覽器中開啟新索引標籤。選擇目標類型，輸入此目標群組的名稱，並且保持使用 GENEVE 通訊協定。選取具有安全應用裝置執行個體的 VPC。視需要修改運作狀態檢查設定，並新增您需要的任何標籤。選擇下一步。您可以立即向目標群組註冊安全應用裝置執行個體，或在完成此程序後註冊。選擇建立目標群組，然後返回上一個瀏覽器索引標籤。

b. (選擇性) 展開接聽程式標籤，然後新增您需要的標籤。

8. (選擇性) 展開負載平衡器標籤，然後新增您需要的標籤。

9. 選擇 Create load balancer (建立負載平衡器)。

步驟 2：建立 Gateway Load Balancer 端點服務

使用下列程序，利用 Gateway Load Balancer 建立端點服務。

建立 Gateway Load Balancer 端點服務

1. 在 https://console.aws.amazon.com/vpc/ 開啟 Amazon VPC 主控台。

2. 在導覽窗格中，選擇 Endpoints Services (端點服務)。

3. 選擇建立端點服務並執行如下動作：

a. 針對 Load balancer type (負載平衡器類型)，選取 Gateway (閘道)。

b. 針對 Available load balancers (可用的負載平衡器)，請選取您的 Gateway Load Balancer。

c. 對於要求接受端點，選取要求接受，以要求手動接受對端點服務的連線請求。否則，系統會自動接受這些請求。

d. 針對 Supported IP address types (支援的 IP 地址類型)，執行下列其中一個操作：

- 選取 IPv4 - 啟用端點服務以接受 IPv4 請求。
- 選取 IPv6 - 啟用端點服務以接受 IPv6 請求。

- 選取 IPv4 和 IPv6 - 啟用端點服務以接受 IPv4 和 IPv6 請求。

e. (選用) 若要新增標籤，請選擇 Add new tag (新增標籤)，然後輸入標籤鍵和標籤值。

f. 選擇建立。注意服務名稱；在建立端點時，您將需要此名稱。

4. 選取新端點服務，然後選擇動作、允許主體。輸入服務消費者的 ARN，允許這些服務消費者建立服務的端點。服務消費者可以是使用者、IAM 角色或 AWS 帳戶。選擇 Allow principals (允許委託人)。

步驟 3：建立 Gateway Load Balancer 端點

使用下列程序建立連線至 Gateway Load Balancer 端點服務的 Gateway Load Balancer 端點。Gateway Load Balancer 端點是區域性的。建議您為每個區域建立一個 Gateway Load Balancer 端點。如需詳細資訊，請參閱 AWS PrivateLink 指南中的透過 AWS PrivateLink存取虛擬應用裝置。

建立閘道負載平衡器端點

1. 在 https://console.aws.amazon.com/vpc/ 開啟 Amazon VPC 主控台。
2. 在導覽窗格中選擇 Endpoints (端點)。
3. 選擇建立端點，然後執行下列動作：

   a. 針對 Service category (服務類別) 中，選擇 Other endpoint services (其他端點服務)。

   b. 對於服務名稱，請輸入之前註明的服務名稱，然後選擇驗證服務。

   c. 對於 VPC，選取服務消費者 VPC。

   d. 對於子網路，選取 Gateway Load Balancer 端點的子網路。

      注意：建立閘道Load Balancer端點時，您只能在每個可用區域內選取一個子網路。

   e. 針對 IP address type (IP 地址類型)，從下列選項中選擇：

      - IPv4 - 將 IPv4 地址指派給您的端點網路介面。只有當所有選取的子網都具有 IPv4 地址範圍時，才支援此選項。
      - IPv6 - 將 IPv6 地址指派給您的端點網路介面。只有當所有選取的子網都是 IPv6 子網時，才支援此選項。
      - Dualstack - 將 IPv4 和 IPv6 地址指派給您的端點網路介面。只有當所有選取的子網都具有 IPv4 和 IPv6 地址範圍時，才支援此選項。

   f. (選用) 若要新增標籤，請選擇 Add new tag (新增標籤)，然後輸入標籤鍵和標籤值。

   g. 選擇建立端點。起始狀態為 `pending acceptance`。

若要接受端點連線請求，請使用下列程序。

1. 在導覽窗格中，選擇 Endpoints Services (端點服務)。
2. 選取端點服務。
3. 從 Endpoint connections (端點連線) 標籤中，選取端點連線。
4. 若要接受連線請求，請選擇 Actions (動作)、Accept endpoint connection request (接受端點連線請求)。出現確認提示時，請輸入 **accept**，然後選擇 Accept (接受)。

步驟 4：設定路由

為服務消費者 VPC 設定下列路由表。如此可讓安全應用裝置針對傳送至應用程式伺服器的傳入流量執行安全檢查。

設定路由

1. 在 https://console.aws.amazon.com/vpc/ 開啟 Amazon VPC 主控台。
2. 在導覽窗格中，選擇 Route tables (路由表)。
3. 選取網際網路閘道路由表並執行以下操作：
   a. 選擇 Actions (動作)、Edit routes (編輯路由)。
   b. 選擇 Add route (新增路由)。針對 Destination (目的地)，請輸入應用程式伺服器子網的 IPv4 CIDR 區塊。針對 Target (目標)，請選取 VPC 端點。
   c. 如果您支援 IPv6，請選擇 Add route (新增路由)。針對 Destination (目的地)，請輸入應用程式伺服器子網的 IPv6 CIDR 區塊。針對 Target (目標)，請選取 VPC 端點。
   d. 選擇儲存變更。
4. 為具有應用程式伺服器的子網選取路由表並執行以下操作：
   a. 選擇 Actions (動作)、Edit routes (編輯路由)。
   b. 選擇 Add route (新增路由)。針對 Destination (目標)，輸入 **0.0.0.0/0**。針對 Target (目標)，請選取 VPC 端點。
   c. 如果您支援 IPv6，請選擇 Add route (新增路由)。針對 Destination (目標)，輸入 **::/0**。針對 Target (目標)，請選取 VPC 端點。
   d. 選擇儲存變更。
5. 選取具有 Gateway Load Balancer 端點之子網路的路由表，並執行以下操作：

a. 選擇 Actions (動作)、Edit routes (編輯路由)。
b. 選擇 Add route (新增路由)。針對 Destination (目標)，輸入 **`0.0.0.0/0`**。針對 Target (目標)，請選取網際網路閘道。
c. 如果您支援 IPv6，請選擇 Add route (新增路由)。針對 Destination (目標)，輸入 **`::/0`**。針對 Target (目標)，請選取網際網路閘道。
d. 選擇儲存變更。

使用 開始使用 Gateway Load Balancer AWS CLI

Gateway Load Balancer 可讓您輕鬆部署、擴展與管理第三方虛擬應用裝置，如安全應用裝置。

在本教學課程中，我們將使用 Gateway Load Balancer 和 Gateway Load Balancer 端點來實作檢查系統。

目錄

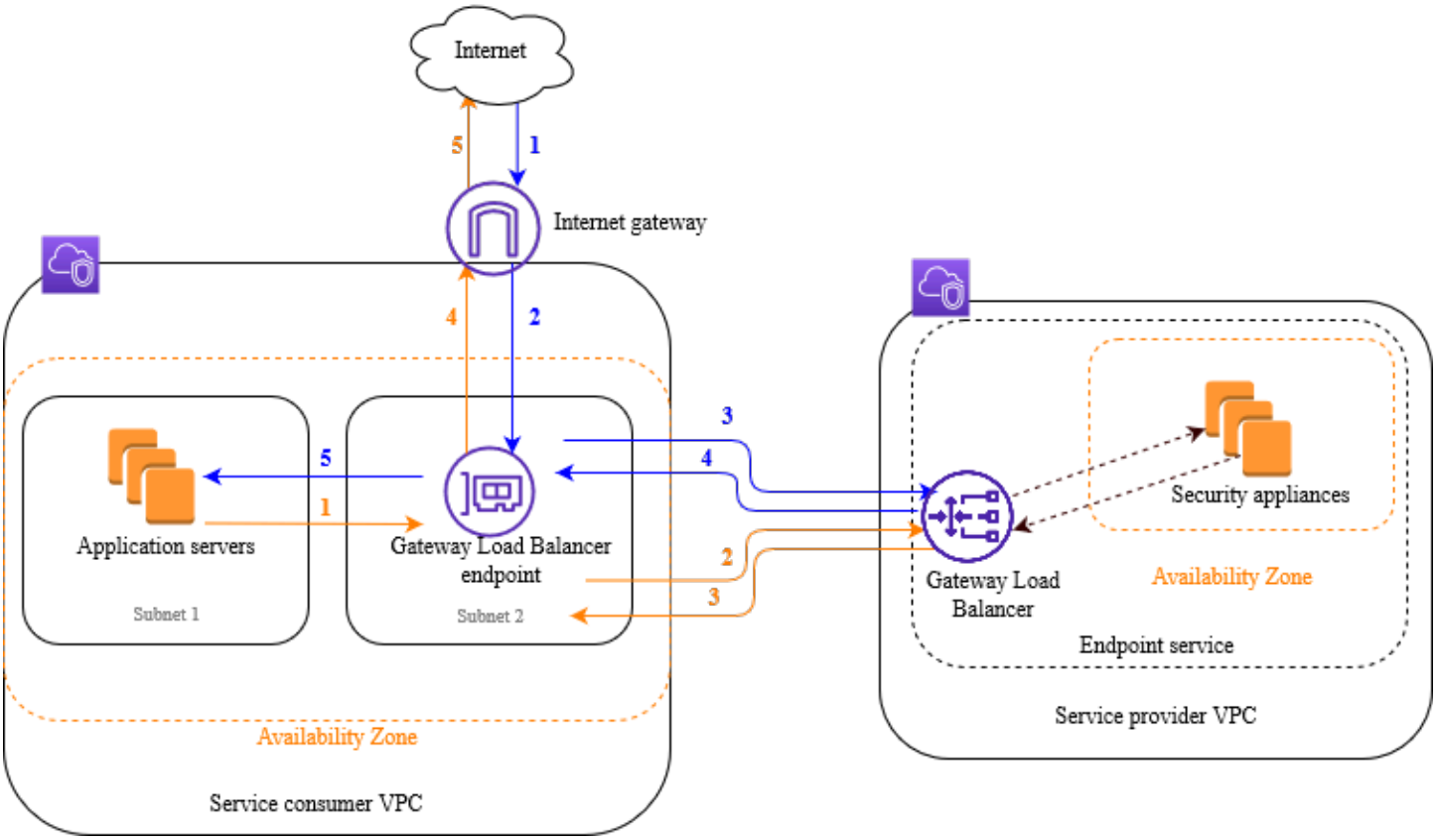
- 概要
- 先決條件
- 步驟 1：建立 Gateway Load Balancer 並註冊目標
- 步驟 2：建立 Gateway Load Balancer 端點
- 步驟 3：設定路由

概要

Gateway Load Balancer 端點是一種 VPC 端點，它在服務提供者 VPC 的虛擬應用裝置和服務消費者 VPC 的應用程式伺服器之間提供私有連線。在與虛擬應用裝置相同的 VPC 中部署 Gateway Load Balancer。這些虛擬應用裝置作為 Gateway Load Balancer 的目標群組註冊。

應用程式伺服器在服務消費者 VPC 中的一個子網路 (目的地子網路) 中執行，而 Gateway Load Balancer 端點位於相同 VPC 的另一個子網路中。所有透過網際網路閘道進入服務消費者 VPC 的流量會先路由至 Gateway Load Balancer 端點，然後再路由至目的地子網路。

同樣，離開應用程式伺服器 (目的地子網路) 的所有流量會路由至 Gateway Load Balancer 端點，然後再路由回網際網路。下列網路圖是如何使用 Gateway Load Balancer 端點存取端點服務的視覺化表示。



下面的帶編號項目突出顯示和解釋在前述網路圖中顯示的元素。

從網際網路到應用程式的流量 (藍色箭頭)：

1. 流量透過網際網路閘道進入服務消費者 VPC。
2. 作為傳入路由的結果，流量傳送至 Gateway Load Balancer 端點。
3. 流量傳送至 Gateway Load Balancer，後者將此流量分發至其中一個安全應用裝置。
4. 流量會在由安全應用裝置檢查之後傳回到 Gateway Load Balancer 端點。
5. 流量傳送至應用程式伺服器 (目的地子網路)。

從應用程式到網際網路的流量 (橙色箭頭)：

1. 作為在應用程式伺服器子網路上設定的預設路由的結果，流量傳送至 Gateway Load Balancer 端點。
2. 流量傳送至 Gateway Load Balancer，後者將此流量分發至其中一個安全應用裝置。
3. 流量會在由安全應用裝置檢查之後傳回到 Gateway Load Balancer 端點。
4. 根據路由表組態，將流量傳送至網際網路閘道。

5. 流量會傳回網際網路。

路由

網際網路閘道的路由表必須具有相應條目，即將目的地為應用程式伺服器的流量傳送至 Gateway Load Balancer 端點。若要指定 Gateway Load Balancer 端點，請使用 VPC 端點的 ID。如下範例顯示雙堆疊組態的路由。

目的地	目標
VPC A IPv4 CIDR	區域
VPC A IPv6 CIDR	區域
### 1 IPv4 CIDR	*vpc-endpoint-id*
### 1 IPv6 CIDR	*vpc-endpoint-id*

帶應用程式伺服器的子網路的路由表必須具有相應條目，即將應用程式伺服器傳出的所有流量路由至 Gateway Load Balancer 端點。

目的地	目標
VPC A IPv4 CIDR	區域
VPC A IPv6 CIDR	區域
0.0.0.0/0	*vpc-endpoint-id*
::/0	*vpc-endpoint-id*

帶 Gateway Load Balancer 端點的子網路的路由表必須將從檢查傳回的流量路由至其最終目的地。對於源自網際網路的流量，本機路由會確保流量到達應用程式伺服器。對於源自應用程式伺服器的流量，請新增相應條目，即將所有流量路由至網際網路閘道。

目的地	目標
VPC A IPv4 CIDR	區域
VPC A IPv6 CIDR	區域
0.0.0.0/0	*internet-gateway-id*
::/0	*internet-gateway-id*

先決條件

- AWS CLI 如果您使用的版本不支援 Gateway Load Balancer，請安裝 AWS CLI 或更新至目前版本的 。如需詳細資訊，請參閱「AWS Command Line Interface 使用者指南」中的安裝 AWS CLI。
- 確定服務消費者 VPC 對於包含應用程式伺服器的每個可用區域至少有兩個子網路。一個子網路用於 Gateway Load Balancer 端點，另一個子網用於應用程式伺服器。
- 確定服務消費者 VPC 對於包含安全應用裝置執行個體的每個可用區域至少有兩個子網路。一個子網路用於 Gateway Load Balancer，另一個子網路用於執行個體。
- 在服務消費者 VPC 中的每個安全應用裝置子網路內啟動至少一個安全應用裝置執行個體。這些執行個體的安全群組必須允許連接埠 6081 上的 UDP 流量。

步驟 1：建立 Gateway Load Balancer 並註冊目標

使用下列程序建立負載平衡器、接聽程式和目標群組，並將您的安全應用裝置執行個體註冊為目標。

建立 Gateway Load Balancer 並註冊目標

1. 使用 create-load-balancer 命令來建立類型為 `gateway` 的負載平衡器。您可以為在其中啟動安全應用裝置執行個體的每個可用區域指定一個子網路。

   ```
   aws elbv2 create-load-balancer --name my-load-balancer --type gateway --
   subnets provider-subnet-id
   ```

   預設為僅支援 IPv4 地址。要同時支援 IPv4 和 IPv6 地址，請新增 `--ip-address-type dualstack` 選項。

其輸出將包含負載平衡器的 Amazon Resource Name (ARN)，其格式顯示在如下範例中。

```
arn:aws:elasticloadbalancing:us-east-2:123456789012:loadbalancer/gwy/my-load-
balancer/1234567890123456
```

2. 使用 create-target-group 命令建立目標群組，並指定在其中啟動執行個體的服務提供者 VPC。

```
aws elbv2 create-target-group --name my-targets --protocol GENEVE --port 6081 --
vpc-id provider-vpc-id
```

其輸出將包含目標群組的 ARN，格式如下。

```
arn:aws:elasticloadbalancing:us-east-2:123456789012:targetgroup/my-
targets/0123456789012345
```

3. 使用 register-targets 命令向目標群組註冊您的執行個體。

```
aws elbv2 register-targets --target-group-arn targetgroup-arn --targets
 Id=i-1234567890abcdef0 Id=i-0abcdef1234567890
```

4. 使用 create-listener 命令為您的負載平衡器建立具有預設規則以轉送請求至目標群組的接聽程式。

```
aws elbv2 create-listener --load-balancer-arn loadbalancer-arn --default-actions
 Type=forward,TargetGroupArn=targetgroup-arn
```

其輸出將包含接聽程式的 ARN，格式如下。

```
arn:aws:elasticloadbalancing:us-east-2:123456789012:listener/gwy/my-load-
balancer/1234567890123456/abc1234567890123
```

5. (選用) 您可以使用如下 describe-target-health 命令驗證目標群組已註冊目標的運作狀態。

```
aws elbv2 describe-target-health --target-group-arn targetgroup-arn
```

步驟 2：建立 Gateway Load Balancer 端點

使用下列程序建立 Gateway Load Balancer 端點。Gateway Load Balancer 端點是區域性的。建議您為每個區域建立一個 Gateway Load Balancer 端點。如需詳細資訊，請參閱透過 AWS PrivateLink存取虛擬應用裝置。

建立閘道負載平衡器端點

1. 使用 create-vpc-endpoint-service-configuration 命令，透過 Gateway Load Balancer 建立端點服務組態。

```
aws ec2 create-vpc-endpoint-service-configuration --gateway-load-balancer-arns loadbalancer-arn --no-acceptance-required
```

   要同時支援 IPv4 和 IPv6 地址，請新增 `--supported-ip-address-types ipv4 ipv6` 選項。

   輸出包含服務 ID (例如，vpce-svc-12345678901234567) 和服務名稱 (例如，com.amazonaws.vpce.us-east-2.vpce-svc-12345678901234567)。

2. 使用 modify-vpc-endpoint-service-permissions 命令允許服務消費者建立服務端點。服務消費者可以是使用者、IAM 角色或 AWS 帳戶。下列範例會新增指定 的許可 AWS 帳戶。

```
aws ec2 modify-vpc-endpoint-service-permissions --service-id vpce-svc-12345678901234567 --add-allowed-principals arn:aws:iam::123456789012:root
```

3. 使用 create-vpc-endpoint 命令為您的服務建立 Gateway Load Balancer 端點。

```
aws ec2 create-vpc-endpoint --vpc-endpoint-type GatewayLoadBalancer --service-name com.amazonaws.vpce.us-east-2.vpce-svc-12345678901234567 --vpc-id consumer-vpc-id --subnet-ids consumer-subnet-id
```

   要同時支援 IPv4 和 IPv6 地址，請新增 `--ip-address-type dualstack` 選項。

   輸出包含 Gateway Load Balancer 端點的 ID (例如，vpce-01234567890abcdef)。

步驟 3：設定路由

為服務消費者 VPC 設定下列路由表。如此可讓安全應用裝置針對傳送至應用程式伺服器的傳入流量執行安全檢查。

設定路由

1. 使用 create-route 命令為網際網路閘道向路由表新增條目，該閘道將目的地為應用程式伺服器的流量路由至 Gateway Load Balancer 端點。

```
aws ec2 create-route --route-table-id gateway-rtb --destination-cidr-block Subnet 1
 IPv4 CIDR --vpc-endpoint-id vpce-01234567890abcdef
```

   如果您支援 IPv6，則新增如下路由。

```
aws ec2 create-route --route-table-id gateway-rtb --destination-cidr-block Subnet 1
 IPv6 CIDR --vpc-endpoint-id vpce-01234567890abcdef
```

2. 使用 create-route 命令為帶應用程式伺服器的子網路向路由表新增相應條目，該子網路將應用程式伺服器傳出的所有流量路由至 Gateway Load Balancer 端點。

```
aws ec2 create-route --route-table-id application-rtb --destination-cidr-block
 0.0.0.0/0 --vpc-endpoint-id vpce-01234567890abcdef
```

   如果您支援 IPv6，則新增如下路由。

```
aws ec2 create-route --route-table-id application-rtb --destination-cidr-block ::/0
 --vpc-endpoint-id vpce-01234567890abcdef
```

3. 使用 create-route 命令，為具有 Gateway Load Balancer 端點的子網路向路由表新增相應條目，該子網路會將源自應用程式伺服器的所有流量路由至網際網路閘道。

```
aws ec2 create-route --route-table-id endpoint-rtb --destination-cidr-block
 0.0.0.0/0 --gateway-id igw-01234567890abcdef
```

   如果您支援 IPv6，則新增如下路由。

```
aws ec2 create-route --route-table-id endpoint-rtb --destination-cidr-block ::/0 --
gateway-id igw-01234567890abcdef
```

4. 針對每個區域中的每個應用程式子網路路由表重複此步驟。

閘道負載平衡器

使用 Gateway Load Balancer 來部署和管理支援 GENEVE 通訊協定的虛擬應用裝置機群。

Gateway Load Balancer 在開放系統互相連線 (OSI) 模型的第三層運作。它會接聽所有連接埠的全部 IP 封包，並使用連接埠 6081 上的 GENEVE 通訊協定轉送流量至接聽程式規則中指定的目標群組。

您可以依據需求的變動，為負載平衡器新增或移除目標，而不會中斷整體的請求流程。當應用程式的流量隨著時間發生變化，Elastic Load Balancing 會擴展您的負載平衡器。Elastic Load Balancing 能夠自動擴展以因應絕大多數的工作負載。

目錄

- 負載平衡器狀態
- IP 地址類型
- 可用區域
- 閒置逾時
- 負載平衡器屬性
- 網路 ACL
- 非對稱流程
- 網路最大傳輸單位 (MTU)
- 建立 Gateway Load Balancer
- 更新 Gateway Load Balancer 的 IP 地址類型
- 編輯 Gateway Load Balancer 的屬性
- 標記閘道Load Balancer
- 刪除 Gateway Load Balancer
- Gateway Load Balancer 的容量保留

負載平衡器狀態

Gateway Load Balancer 可以是以下其中一個狀態：

`provisioning`

正在設定 Gateway Load Balancer。

`active`

Gateway Load Balancer 已設定完成並準備好路由流量。

`failed`

無法設定 Gateway Load Balancer。

IP 地址類型

您可以設定應用程式伺服器可用來存取 Gateway Load Balancer 的 IP 地址類型。

Gateway Load Balancer 支援下列 IP 地址類型：

`ipv4`

僅支援 IPv4。

`dualstack`

同時支援 IPv4 和 IPv6。

考量事項

- 您為負載平衡器指定的 Virtual Private Cloud (VPC) 和子網路必須具有相關聯的 IPv6 CIDR 區塊。
- 服務消費者 VPC 中的子網路路由表必須路由 IPv6 流量，而這些子網路的網路 ACL 必須允許 IPv6 流量。
- Gateway Load Balancer 使用 IPv4 GENEVE 標頭封裝 IPv4 和 IPv6 用戶端流量，並將其傳送至應用裝置。應用裝置會使用 IPv4 GENEVE 標頭封裝 IPv4 和 IPv6 用戶端流量，並將其傳回至 Gateway Load Balancer。

如需 IP 地址類型的詳細資訊，請參閱 更新 Gateway Load Balancer 的 IP 地址類型。

可用區域

建立 Gateway Load Balancer 時，您可以啟用一或多個可用區域，並指定與每個區域對應的子網路。當您啟用多個可用區域時，即使可用區域無法使用，它會確保負載平衡器仍可繼續路由流量。您指定的子網路必須分別至少有 8 個可用的 IP 地址。建立負載平衡器後，無法移除子網路。若要移除子網路，您必須建立新的負載平衡器。

閒置逾時

對於透過 Gateway Load Balancer 提出的每個 TCP 請求，會追蹤該連線的狀態。若在比閒置逾時更長的時間內沒有由用戶端或目標透過連線傳送的資料，連線將關閉。在閒置逾時期間過後，負載平衡器會將下一個 TCP SYN 視為新的流程，並將其路由至新的目標。不過，閒置逾時期間過後傳送的資料封包會遭到捨棄。

TCP 流程的預設閒置逾時值為 350 秒，但可以更新為介於 60-6000 秒之間的任何值。用戶端或目標可以使用 TCP 保持連線封包來重設閒置逾時。

> **黏性限制**
>
> 只有在使用 5 個元組黏性時，才能更新 Gateway Load Balancer 閒置逾時。使用 3 元組或 2 元組黏性時，會使用預設閒置逾時值。如需詳細資訊，請參閱流程黏性

儘管 UDP 為無連線，負載平衡器會根據來源與目的地 IP 地址及連接埠來維護 UDP 流程狀態。這可確保持續傳送屬於相同流程的封包至相同目標。在閒置逾時期間經過之後，負載平衡器會將傳入的 UDP 封包視為新流程，並將其路由至新目標。Elastic Load Balancing 會將 UDP 流量的閒置逾時值設為 120 秒。無法對此進行變更。

EC2 執行個體必須在 30 秒內回應新的請求，才能建立傳回路徑。

如需詳細資訊，請參閱更新閒置逾時。

負載平衡器屬性

以下是 Gateway Load Balancer 的負載平衡器屬性：

`deletion_protection.enabled`

表示是否已啟用刪除保護。預設值為 `false`。

`load_balancing.cross_zone.enabled`

表示是否已啟用跨區域負載平衡。預設值為 `false`。

如需詳細資訊，請參閱編輯負載平衡器屬性。

網路 ACL

如果應用程式伺服器和 Gateway Load Balancer 端點位於相同的子網路中，則會評估從應用程式伺服器到 Gateway Load Balancer 端點的流量 NACL 規則。

非對稱流程

當負載平衡器處理初始流程封包且回應流程封包未透過負載平衡器路由時，Gateway Load Balancer 支援非對稱流程。不建議使用非對稱路由，因為這會導致網路效能降低。當負載平衡器不處理初始流程封包，但回應流程封包透過負載平衡器路由時，Gateway Load Balancer 不支援非對稱流程。

網路最大傳輸單位 (MTU)

最大傳輸單位 (MTU) 是允許傳輸通過網路的最大資料封包大小。Gateway Load Balancer 界面 MTU 支援最多 8,500 位元組的封包。若到達 Gateway Load Balancer 界面的封包大於 8500 位元組，則該封包會遭捨棄。

Gateway Load Balancer 用 GENEVE 標頭封裝 IP 流量，並將其轉送至應用裝置。GENEVE 封裝程序會將 68 個位元組新增至原始封包。因此，若要支援最多 8，500 個位元組的封包，請確保您設備的 MTU 設定支援至少 8，568 個位元組的封包。

Gateway Load Balancer 不支援 IP 分段。此外，Gateway Load Balancer 不會產生 ICMP 訊息「目的地無法存取：需要分段和設定 DF」。因此，不支援路徑 MTU 探索 (PMTUD)。

建立 Gateway Load Balancer

Gateway Load Balancer 會從用戶端取得請求，然後分布到目標群組的目標，例如，EC2 執行個體。

若要使用 建立 Gateway Load Balancer AWS Management Console，請完成下列任務。或者，若要使用 建立 Gateway Load Balancer AWS CLI，請參閱 使用 CLI 入門。

任務

- 先決條件
- 建立負載平衡器
- 重要的後續步驟

先決條件

開始之前，請確保您 Gateway Load Balancer 的虛擬私有雲端 (VPC) 在目標所在的每個可用區域中至少有一個公有子網路。

建立負載平衡器

使用下列程序建立您的 Gateway Load Balancer。提供您負載平衡器的基本組態資訊，例如名稱和 IP 地址類型。然後提供您網路的相關資訊，以及將流量路由至您目標群組的接聽程式資訊。Gateway Load Balancer 需要使用 GENEVE 通訊協定的目標群組。

使用主控台建立負載平衡器和接聽程式

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選擇 Create load balancer (建立負載平衡器)。
4. 在 Gateway Load Balancer 下，選擇建立。
5. 基本組態
   a. 針對 Load balancer name (負載平衡器名稱)，輸入負載平衡器的名稱。例如 **my-glb**。Gateway Load Balancer 的名稱必須在該區域的負載平衡器集內是唯一的。該名稱最多可包含 32 個字元，只能包含英數字元和連字號，並且不得以連字號開頭或結尾。
   b. 針對 IP 地址類型，選擇 IPv4 以僅支援 IPv4 地址，或選擇 Dualstack 同時支援 IPv4 和 IPv6 地址。
6. 網路映射
   a. 對於 VPC，請選取服務提供者 VPC。
   b. 對於映射，選取您在其中啟動安全應用裝置執行個體的所有可用區域，以及對應的公有子網路。
7. IP 接聽程式路由
   a. 對於預設動作，請選取要接收流量的目標群組。如果您沒有目標群組，請選擇建立目標群組。如需詳細資訊，請參閱建立目標群組。
   b. (選擇性) 展開接聽程式標籤，然後新增您需要的標籤。
8. (選擇性) 展開負載平衡器標籤，然後新增您需要的標籤。
9. 複查您的組態，然後選擇建立負載平衡器。

重要的後續步驟

建立負載平衡器後，請確認您的 EC2 執行個體已通過初始運作狀態檢查。若要測試負載平衡器，則必須建立 Gateway Load Balancer 端點並更新路由表，以使 Gateway Load Balancer 端點成為下一個躍點。這些組態皆在 Amazon VPC 主控台內設定。如需詳細資訊，請參閱開始使用教學課程。

更新 Gateway Load Balancer 的 IP 地址類型

您可以設定 Gateway Load Balancer，讓應用程式伺服器只能使用 IPv4 地址或使用 IPv4 和 IPv6 地址（雙堆疊）存取負載平衡器。負載平衡器會根據目標群組的 IP 地址類型與目標進行通訊。如需詳細資訊，請參閱IP 地址類型。

使用主控台更新 IP 地址類型

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選取負載平衡器。
4. 選擇 Actions (動作)、Edit IP address type (編輯 IP 地址類型)。
5. 針對 IP address type (IP 地址類型)，選擇 ipv4 只支援 IPv4 地址，或選擇 dualstack 同時支援 IPv4 和 IPv6 地址。
6. 選擇儲存。

使用 更新 IP 地址類型 AWS CLI

使用 set-ip-address-type 命令。

編輯 Gateway Load Balancer 的屬性

建立 Gateway Load Balancer 之後，您可以編輯其負載平衡器屬性。

負載平衡器屬性

- 刪除保護
- 跨區域負載平衡

刪除保護

為避免您的 Gateway Load Balancer 上遭意外刪除，您可以啟用刪除保護。根據預設，刪除保護是停用的。

如果您為 Gateway Load Balancer 啟用刪除保護，則必須先停用才可刪除 Gateway Load Balancer。

使用主控台來啟用刪除保護

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選取 Gateway Load Balancer。
4. 選擇動作、編輯屬性。
5. 在編輯負載平衡器屬性頁面上，選取刪除保護的啟用，然後選擇儲存。

使用主控台來停用刪除保護

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選取 Gateway Load Balancer。
4. 選擇動作、編輯屬性。
5. 在編輯負載平衡器屬性頁面上，清除刪除保護的啟用，然後選擇儲存。

使用 啟用或停用刪除保護 AWS CLI

以 屬性來使用 modify-load-balancer-attributes`deletion_protection.enabled` 命令。

跨區域負載平衡

預設情況下，每個負載平衡器節點只會將流量分布到其可用區域中的登錄目標。若您啟用跨區域負載平衡功能，每個 Gateway Load Balancer 節點會將流量分布至所有可用區域內已登錄的目標。如需詳細資訊，請參閱 Elastic Load Balancing 使用者指南中的跨區域負載平衡。

使用主控台啟用跨區域負載平衡

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。

3. 選取 Gateway Load Balancer。
4. 選擇動作、編輯屬性。
5. 在編輯負載平衡器屬性頁面上，選取跨區域負載平衡的啟用，然後選擇儲存。

使用 啟用跨區域負載平衡 AWS CLI

以 屬性來使用 modify-load-balancer-attributes`load_balancing.cross_zone.enabled` 命令。

標記閘道Load Balancer

標籤可幫助您以不同的方式來將負載平衡器分類，例如，根據目的、擁有者或環境。

您可以在每個負載平衡器中加入多個標籤。每個 Gateway Load Balancer 的標籤索引鍵必須是唯一的。如果所新增的標籤，其索引鍵已經與負載平衡器相關聯，則此動作會更新該標籤的值。

使用標籤完成負載平衡器使用後，可將其自 Gateway Load Balancer 中移除。

限制

- 每一資源標籤數上限：50
- 索引鍵長度上限：127 個 Unicode 字元
- 數值長度上限：255 個 Unicode 字元
- 標籤金鑰與值皆區分大小寫。允許的字元包括可用 UTF-8 表示的英文字母、空格和數字，還有以下特殊字元：+ - = . _ : / @。不可使用結尾或前方空格。
- 請勿在標籤名稱或值中使用 `aws:`字首，因為它已保留供 AWS 使用。您不可編輯或刪除具此字首的標籤名稱或值。具此字首的標籤，不算在受資源限制的標籤計數內。

使用主控台來更新 Gateway Load Balancer 的標籤

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選取 Gateway Load Balancer。
4. 選擇 Tags (標籤)、Add/Edit Tags (新增/編輯標籤)，然後執行一項或多項下列動作：
   a. 若要更新標籤，請編輯 Key (索引鍵) 和 Value (值) 的值。
   b. 若要新增新標籤，請選擇 Create Tag (建立標籤)。對於索引鍵和值，輸入相應的值。
   c. 若要刪除標籤，請選擇標籤旁的刪除圖示 (X)。

5. 完成更新標籤的作業時，請選擇 Save (儲存)。

使用 更新 Gateway Load Balancer 的標籤 AWS CLI

使用 add-tags 和 remove-tags 指令。

刪除 Gateway Load Balancer

在您的 Gateway Load Balancer 可用後，將會根據持續執行時間收取一小時或不足一小時的費用。當您不再需要 Gateway Load Balancer 時，可以將它刪除。刪除 Gateway Load Balancer 後，便會停止收取費用。

如果 Gateway Load Balancer 正由其他服務使用，則無法刪除負載平衡器。例如，如果 Gateway Load Balancer 與 VPC 端點服務相關聯，您必須先刪除端點服務組態，才能刪除相關聯的 Gateway Load Balancer。

若刪除 Gateway Load Balancer，接聽程式也會一併刪除。刪除 Gateway Load Balancer 不會影響其註冊目標。例如，您的 EC2 執行個體將繼續執行，且仍會登錄到他們的目標群組。若要刪除您的目標群組，請參閱刪除閘道Load Balancer的目標群組。

使用主控台刪除 Gateway Load Balancer

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選取 Gateway Load Balancer。
4. 選擇 動作、刪除。
5. 出現確認提示時，選擇 Yes, Delete (是，刪除)。

使用 刪除閘道Load Balancer AWS CLI

使用 delete-load-balancer 指令。

Gateway Load Balancer 的容量保留

負載平衡器容量單位 (LCU) 保留可讓您為負載平衡器保留靜態最小容量。Gateway Load Balancer 會自動擴展，以支援偵測到的工作負載並滿足容量需求。設定最小容量時，您的負載平衡器會根據收到的流量繼續向上或向下擴展，但也會防止容量低於設定的最小容量。

考慮在下列情況下使用 LCU 保留：

- 您即將發生的事件會有突然、異常的高流量，並希望確保您的負載平衡器能夠在事件期間支援突增的流量尖峰。
- 由於工作負載的性質很短，您的尖峰流量無法預測。
- 您要將負載平衡器設定為在特定開始時間加入或遷移服務，並且需要從高容量開始，而不是等待自動擴展生效。
- 您需要維持最低容量，以符合服務水準協議或合規要求。
- 您正在負載平衡器之間遷移工作負載，並想要設定目的地以符合來源的規模。

估算您需要的容量

判斷負載平衡器應預留的容量時，建議您執行負載測試或檢閱代表您預期近期流量的歷史工作負載資料。使用 Elastic Load Balancing 主控台，您可以根據檢閱的流量預估需要預留多少容量。

或者，您可以參考 CloudWatch 指標 ProcessedBytes 來判斷正確的容量層級。負載平衡器的容量保留在 LCUs中，每個 LCU 等於 2.2Mbps。您可以使用 PeakBytesPerSecond 指標來查看負載平衡器上的每分鐘輸送量流量上限，然後使用 2.2Mbps 的轉換率將該輸送量轉換為 LCUs，等於 1 個 LCU。

如果您沒有歷史工作負載資料可供參考，且無法執行負載測試，您可以使用 LCU 保留計算器預估所需的容量。LCU 保留計算器會根據 AWS 觀察到的歷史工作負載使用資料，可能不會代表您的特定工作負載。如需詳細資訊，請參閱Load Balancer容量單位保留計算器。

LCU 保留配額

您的帳戶具有與 LCUs 相關的配額。如需詳細資訊，請參閱the section called “Load Balancer容量單位”。

請求 Gateway Load Balancer 的Load Balancer容量單位保留

在使用 LCU 保留之前，請檢閱下列項目：

- LCU 保留僅支援 Gateway Load Balancer 的預留輸送量容量。請求 LCU 保留時，請使用 1 LCUs 到 2.2 Mbps 的轉換率，將您的容量需求從 Mbps 轉換為 LCU。
- 容量會保留在區域層級，並平均分散到各個可用區域。在開啟 LCU 保留之前，請確認每個可用區域中有足夠的平均分佈目標。
- LCU 保留請求是以先到先得的方式完成，且取決於當時區域的可用容量。大多數請求通常會在一小時內完成，但最多可能需要幾個小時。

- 若要更新現有的保留，先前的請求必須佈建或失敗。您可以視需要增加預留容量，但每天只能減少兩次預留容量。

請求 LCU 保留

此程序中的步驟說明如何在負載平衡器上請求 LCU 保留。

使用主控台請求 LCU 保留

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格上選擇 Load Balancers (負載平衡器)。
3. 選取負載平衡器名稱。
4. 在容量索引標籤上，選擇編輯 LCU 保留。
5. 選取歷史參考型估算，然後從下拉式清單中選取負載平衡器。
6. 選取參考期間，以檢視建議的預留 LCU 層級。
7. 如果您沒有歷史參考工作負載，您可以選擇手動估算，然後輸入要保留LCUs 數量。
8. 選擇儲存。

使用 請求 LCU 保留 AWS CLI

使用 modify-capacity-reservation 命令。

更新或終止 Gateway Load Balancer 的Load Balancer容量單位保留

更新或終止 LCU 保留

此程序中的步驟說明如何更新或終止負載平衡器上的 LCU 保留。

使用主控台更新或終止 LCU 保留

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格上選擇 Load Balancers (負載平衡器)。
3. 選取負載平衡器名稱。
4. 在容量索引標籤上，確認已佈建保留的狀態。
   a. 若要更新 LCU 保留，請選擇編輯 LCU 保留。
   b. 若要終止 LCU 保留，請選擇取消容量。

使用 更新或終止 LCU 保留 AWS CLI

使用 modify-capacity-reservation 命令。

監控 Gateway Load Balancer 的Load Balancer容量單位保留

保留狀態

LCU 保留有四個可用狀態：

- 待定 - 指出其正在進行佈建的保留。
- 已佈建 - 表示預留容量已就緒且可供使用。
- 失敗 - 表示目前無法完成請求。
- 重新平衡 - 表示已新增可用區域，且負載平衡器正在重新平衡容量。

預留 LCU

若要判斷預留 LCU 使用率，您可以將每分鐘 PeakBytesPerSecond 指標與每小時總和 (ReservedLCUs進行比較。若要將每分鐘位元組數轉換為每小時 LCU，請使用 （每分鐘位元組數）*8/60/ (10^6)/2.2。

監控預留容量

此程序中的步驟說明如何檢查負載平衡器上 LCU 保留的狀態。

使用主控台檢視 LCU 保留的狀態

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格上選擇 Load Balancers (負載平衡器)。
3. 選取負載平衡器名稱。
4. 在容量索引標籤上，您可以檢視保留狀態和預留 LCU 值。

使用 監控 LCU 保留的狀態 AWS CLI

使用 describe-capacity-reservation 命令。

Gateway Load Balancer 的接聽程式

建立 Gateway Load Balancer 時，您可以新增接聽程式。接聽程式是檢查連線請求的程序。

Gateway Load Balancer 的接聽程式會接聽所有連接埠上的所有 IP 封包。為 Gateway Load Balancer 建立接聽程式時，您無法指定通訊協定或連接埠。

當您建立接聽程式後，可指定路由請求的規則。此規則將轉發請求到指定的目標群組。您可以更新接聽程式規則，將請求轉送至不同的目標群組。

接聽程式屬性

以下是 Gateway Load Balancer 的接聽程式屬性：

`tcp.idle_timeout.seconds`

tcp 閒置逾時值，以秒為單位。有效範圍為 60-6000 秒。預設值為 350 秒。

如需詳細資訊，請參閱更新閒置逾時。

更新 Gateway Load Balancer 接聽程式的目標群組

當您建立接聽程式後，可指定路由請求的規則。此規則將轉發請求到指定的目標群組。您可以更新接聽程式規則，將請求轉送至不同的目標群組。

使用主控台更新您的接聽程式

1. 在 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選取負載平衡器，然後選擇 Listeners (接聽程式)。
4. 選擇編輯接聽程式。
5. 對於轉送至目標群組，選擇目標群組。
6. 選擇 Save (儲存)。

使用 更新您的接聽程式 AWS CLI

使用 modify-listener 命令。

更新 Gateway Load Balancer 接聽程式的 TCP 閒置逾時

對於透過 Gateway Load Balancer 提出的每個 TCP 請求，會追蹤該連線的狀態。若在比閒置逾時更長的時間內沒有由用戶端或目標透過連線傳送的資料，連線將關閉。TCP 流程的預設閒置逾時值為 350 秒，但可以更新為介於 60-6000 秒之間的任何值。

使用主控台更新 TCP 閒置逾時

1. 在 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 下方，選擇 Load Balancers (負載平衡器)。
3. 選取 Gateway Load Balancer。
4. 在接聽程式索引標籤上，選擇動作、檢視接聽程式詳細資訊。
5. 在接聽程式詳細資訊頁面的屬性索引標籤中，選取編輯。
6. 在編輯接聽程式屬性頁面上的接聽程式屬性區段中，輸入 TCP 閒置逾時的值。
7. 選擇 Save changes (儲存變更)

使用 更新 TCP 閒置逾時 AWS CLI

使用 modify-listener-attributes 命令搭配 `tcp.idle_timeout.seconds` 屬性。

Gateway Load Balancer 的目標群組

每個目標群組會用來將請求轉送到一個或多個註冊的目標。當您建立接聽程式時，可以為其預設動作指定一個目標群組。流量會轉送至接聽程式規則中指定的目標群組。您可以針對不同類型的請求，建立不同的目標群組。

您可以針對每個目標群組，指定 Gateway Load Balancer 的運作狀態檢查設定。除非您在建立目標群組時覆寫這些設定，或是在之後修改設定，否則每個目標群組都會使用預設的運作狀態檢查設定。當您在接聽程式的規則中指定目標群組後，Gateway Load Balancer 會針對自己已啟用可用區域中的目標群組，持續地監控透過該目標群組註冊的所有目標，以了解目標的運作狀態。Gateway Load Balancer 會將請求路由至運作狀態良好的已註冊目標。如需詳細資訊，請參閱Gateway Load Balancer 目標群組的運作狀態檢查。

目錄

- 路由組態
- Target type (目標類型)
- 已登記的目標
- 目標群組屬性
- 為您的 Gateway Load Balancer 建立目標群組
- Gateway Load Balancer 目標群組的運作狀態檢查
- 編輯 Gateway Load Balancer 的目標群組屬性
- 註冊 Gateway Load Balancer 的目標
- 為您的 Gateway Load Balancer 標記目標群組
- 刪除閘道Load Balancer的目標群組

路由組態

Gateway Load Balancer 的目標群組支援下列的通訊協定和連接埠：

- 通訊協定：GENEVE
- 連接埠：6,081

Gateway Load Balancer 使用 GENEVE 封裝原始封包。GENEVE 標頭使用 Type-Length-Value (TLV) 格式來存放資訊，並使用選項類別 0x0108。設備必須解封裝 TLV 對，才能處理原始封包。如需詳細資訊，請參閱下列部落格文章：整合您的設備與閘道負載平衡器。

Target type (目標類型)

在建立目標群組時，您會指定其目標類型，這會決定您指定其目標的方式。目標群組建立之後，您就無法更改其目標類型。

下列是可能的目標類型：

`instance`

以執行個體 ID 來指定目標。

`ip`

以 IP 地址來指定目標。

如果目標類型是 `ip`，您可以從下列其中一個 CIDR 區塊指定 IP 地址：

- 目標群組 VPC 的子網路
- 10.0.0.0/8 (RFC 1918)
- 100.64.0.0/10 (RFC 6598)
- 172.16.0.0/12 (RFC 1918)
- 192.168.0.0/16 (RFC 1918)

> **Important**
> 您無法指定可公開路由傳送的 IP 地址。

已登記的目標

您的 Gateway Load Balancer 可做為用戶端的單一聯絡窗口，並將傳入的流量分配到各個運作狀態良好的已登錄目標。在 Gateway Load Balancer 能夠使用的每個可用區域中，每個目標群組都必須擁有至少一個已註冊的目標。您可以利用一個或多個群組來登錄每個目標。

如果需求增加，您可以利用一個或多個目標群組來註冊額外的目標以應付需求。只要註冊程序一完成，Gateway Load Balancer 就會開始將流量轉傳到新註冊的目標。

如果需求減少，或者您需要為目標提供服務，可以從目標群組取消目標的註冊。取消目標的登錄，會將該目標從目標群組中移除，但不會影響到目標。取消目標的註冊之後，Gateway Load Balancer 就會立即停止將流量轉傳到目標。目標會進入 `draining` 狀態，直到處理中的請求已完成。當您準備讓目標再繼續接收流量時，可以將目標登錄到目標群組。

目標群組屬性

您可以對目標群組使用下列屬性：

`deregistration_delay.timeout_seconds`

將取消註冊目標的狀態從 `draining` 變更為 unused 之前，Elastic Load Balancing 要等待的時間量。範圍介於 0 到 3600 秒之間。預設值為 300 秒。

`stickiness.enabled`

指示是否啟用目標群組的可設定流程黏性。可能的值為 `true` 或 `false`。預設值為 false。當屬性設定為時 `false`，會使用 5_tuple。

`stickiness.type`

指示流程黏性的類型。與 Gateway Load Balancer 相關聯的目標群組的可能值為：

- `source_ip_dest_ip`
- `source_ip_dest_ip_proto`

`target_failover.on_deregistration`

指示當取消註冊目標時，Gateway Load Balancer 處理現有流程的方式。可能的值為 `rebalance` 和 `no_rebalance`。預設值為 `no_rebalance`。這兩個屬性 (`target_failover.on_deregistration` 和 `target_failover.on_unhealthy`) 無法單獨設定。這兩個屬性必須設為相同的值。

`target_failover.on_unhealthy`

指示當目標運作狀態不佳時，Gateway Load Balancer 處理現有流程的方式。可能的值為 `rebalance` 和 `no_rebalance`。預設值為 `no_rebalance`。這兩個屬性

(`target_failover.on_deregistration` 和 `target_failover.on_unhealthy`) 無法單獨設定。這兩個屬性必須設為相同的值。

如需詳細資訊，請參閱編輯目標群組屬性。

為您的 Gateway Load Balancer 建立目標群組

您可以使用目標群組為 Gateway Load Balancer 註冊目標。

若要將流量轉傳到目標群組中的目標，請建立接聽程式，並且在接聽程式的預設動作中，指定該目標群組。如需詳細資訊，請參閱接聽程式。

您可以隨時從目標群組新增或移除目標。如需詳細資訊，請參閱登記目標。您也可以修改目標群組的運作狀態檢查設定。如需詳細資訊，請參閱修改運作狀態檢查設定。

使用主控台來建立目標群組

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的負載平衡中，選擇目標群組。
3. 選擇 Create target group (建立目標群組)。
4. 基本組態
   a. 針對選擇目標類型，選取執行個體來依執行個體 ID 指定目標，或選取 IP 地址來依 IP 地址指定目標。
   b. 針對目標群組名稱，輸入目標群組的名稱。此名稱在每個帳戶的每個區域中都必須是唯一的，其長度上限為 32 個字元，並且必須僅包含英數字元或連字號，且開頭或結尾不可以是連字號。
   c. 驗證協議是 `GENEVE`，而端口是 `6081`。不支援其他通訊協定或連接埠。
   d. 對於 VPC，請選取具有要包含在目標群組中的安全應用裝置執行個體的虛擬私有雲端 (VPC)。
5. (選用) 針對運作狀態檢查，視需要修改設定和進階設定。如果運作狀態檢查連續超過運作狀態不佳閾值的次數，負載平衡器會停用該目標。當運作狀態檢查連續超過運作狀態不佳閾值次數時，負載平衡器會重新啟用該目標。如需詳細資訊，請參閱Gateway Load Balancer 目標群組的運作狀態檢查。
6. (選用) 展開標籤，然後新增需要的標籤。
7. 選擇下一步。

8. 對於註冊目標，新增一個或多個目標，如下所示：
   - 如果目標類型為執行個體，請選取一或多個執行個體，輸入一或多個連接埠，然後選擇包含為以下待定的項目。
   - 如果目標類型是 IP 地址，請選取網路，輸入 IP 地址和通訊埠，然後選擇包含為以下待定的項目。
9. 選擇 Create target group (建立目標群組)。

使用 建立目標群組 AWS CLI

使用 create-target-group 指令來建立目標群組、使用 add-tags 指令來標記目標群組、使用 register-targets 指令來新增目標。

Gateway Load Balancer 目標群組的運作狀態檢查

您可以利用一個或多個群組來登錄目標。只要註冊程序一完成，Gateway Load Balancer 就會開始將請求路由到新註冊的目標。註冊程序可能需要幾分鐘的時間才能完成，並開始運作狀態檢查。

Gateway Load Balancer 將定期向每個已註冊目標傳送請求以檢查其狀態。每次運作狀態檢查完成後，Gateway Load Balancer 即會關閉其為執行運作狀態檢查而建立的連線。

運作狀態檢查設定

您將使用以下設定，為目標群組中的目標設定主動的運作狀態檢查。如果運作狀態檢查連續失敗超過 UnhealthyThresholdCount 指定的次數，Gateway Load Balancer 會停用該目標。當運作狀態檢查連續成功超過 HealthyThresholdCount 指定的次數時，Gateway Load Balancer 重新啟用該目標。

設定	描述
HealthCheckProtocol	負載平衡器在目標上執行運作狀態檢查時使用的通訊協定。可能的通訊協定包括 HTTP、HTTPS 和 TCP。預設為 TCP。
HealthCheckPort	Gateway Load Balancer 對目標執行運作狀態檢查時使用的連接埠。範圍介於 1 至 65535 之間。預設值為 80。
HealthCheckPath	【HTTP/HTTPS 運作狀態檢查】 運作狀態檢查目標上的運作狀態檢查路徑。預設為 /.

設定	描述
HealthCheckTimeoutSeconds	以秒為單位的時間量，若目標在此期間內毫無回應即表示運作狀態檢查失敗。範圍介於 2 至 120 之間。預設值為 5。
HealthCheckIntervalSeconds	個別目標每次執行運作狀態檢查的大約間隔時間量，以秒為單位。範圍介於 5 至 300 之間。預設為 10 秒。該值必須大於或等於 HealthCheckTimeoutSeconds。 **Important** Gateway Load Balancer 的運作狀態檢查為分散式，採用共識機制判定目標的運作狀態。因此，您應該期望目標應用裝置在設定的時間間隔內收到數個運作狀態檢查。
HealthyThresholdCount	將運作狀態不佳的目標視為運作狀態良好之前，運作狀態檢查需連續成功的次數。範圍介於 2 至 10 之間。預設值為 5。
UnhealthyThresholdCount	將目標視為運作狀態不佳之前，運作狀態檢查需連續失敗的次數。範圍介於 2 至 10 之間。預設為 2。
Matcher	[HTTP/HTTPS 運作狀態檢查] 檢查來自目標的成功回應時所使用的 HTTP 代碼。此值必須為 200-399。

目標運作狀態

在 Gateway Load Balancer 向目標傳送運作狀態檢查請求之前，您必須向目標群組註冊該目標，由接聽程式規則中指定其目標群組，並確保 Gateway Load Balancer 已啟用該目標的可用區域。

下表說明已註冊目標的運作狀態可能的值。

Value	描述
initial	Gateway Load Balancer 正在註冊目標或對目標執行初始運作狀態檢查。 相關原因代碼： Elb.RegistrationInProgress | Elb.InitialHealthChecking
healthy	目標的運作狀態良好。 相關原因代碼：無
unhealthy	目標未回應運作狀態檢查或未通過運作狀態檢查。 相關原因碼：Target.FailedHealthChecks
unused	目標未向目標群組註冊、未在接聽程式規則中使用目標群組、目標位於未啟用的可用區域，或目標處於已停止或已終止狀態。 相關原因碼：Target.NotRegistered | Target.NotInUse | Target.InvalidState | Target.IpUnusable
draining	目標正在取消註冊，連接耗盡作業進行中。 相關原因碼：Target.DeregistrationInProgress
unavailable	目標健全狀態無法使用。 相關原因碼：Elb.InternalError

運作狀態檢查原因代碼

如果目標的狀態是 Healthy 以外的任何值，API 將傳回問題的原因代碼和描述，而且主控台會顯示同樣的描述。開頭為 Elb 的原因代碼源自 Gateway Load Balancer 端，而開頭為 Target 的原因代碼源自目標端。

原因代碼	描述
`Elb.InitialHealthChecking`	初始運作狀態檢查正進行中
`Elb.InternalError`	運作狀態檢查由於內部錯誤而失敗
`Elb.RegistrationInProgress`	目標註冊正進行中
`Target.DeregistrationInProgress`	目標取消註冊正進行中
`Target.FailedHealthChecks`	運作狀態檢查失敗
`Target.InvalidState`	目標處於停止狀態 目標處於終止狀態 目標處於終止或停止狀態 目標處於無效狀態
`Target.IpUnusable`	IP 地址不能做為目標，因為負載平衡器正在使用它
`Target.NotInUse`	目標群組未設定為接收來自 Gateway Load Balancer 的流量 目標位於 Gateway Load Balancer 未啟用的可用區域
`Target.NotRegistered`	目標未向目標群組註冊

Gateway Load Balancer 目標失敗案例

現有流程：根據預設，現有的流程會移至相同的目標，除非流程逾時或重設，無論目標的運作狀態和註冊狀態為何。這種方法會推進連接耗盡，並容納第三方防火牆，這些防火牆有時由於 CPU 用量過高而無法回應運作狀態檢查。如需詳細資訊，請參閱目標容錯移轉。

新流程：新流程會傳送至運作狀態良好的目標。針對流程做出負載平衡決定後，Gateway Load Balancer 會將流程傳送至相同的目標，即使該目標運作狀態不佳或其他目標運作狀態良好亦是如此。

當所有目標運作狀態不佳時，Gateway Load Balancer 會隨機挑選一個目標，並在流程的生命週期內將流量轉送至該目標，直到其重設或逾時為止。由於流量會轉送至運作狀態不佳的目標，因此流量會被捨棄，直到該目標再次變得運作狀態良好為止。

TLS 1.3：如果目標群組設定有 HTTPS 運作狀態檢查，則其註冊的目標在僅支援 TLS 1.3 時將運作狀態檢查失敗。這些目標必須支援早期版本的 TLS，例如 TLS 1.2。

跨區域負載平衡：依預設，會停用跨可用區域的負載平衡。如果啟用跨區域的負載平衡，則每個 Gateway Load Balancer 都可以看到所有可用區域中的全部目標，而且無論其區域為何，這些目標都會被視為相同。

負載平衡和運作狀態檢查決策在區域之間始終是獨立的。即使啟用跨區域的負載平衡，現有流程和新流程的行為也與上述相同。如需詳細資訊，請參閱 Elastic Load Balancing 使用者指南中的跨區域負載平衡。

檢查目標的運作狀態

您可以檢查已向目標群組註冊的各個目標的運作狀態。

使用主控台檢查目標的運作狀態

1. 在 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的負載平衡中，選擇目標群組。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 在 Targets (目標) 標籤上，Status (狀態) 欄指出各目標的狀態。
5. 如果目標狀態為 `Healthy` 以外的任何值，則狀態詳細資料欄會包含更多資訊。

使用 檢查目標的運作狀態 AWS CLI

使用 describe-target-health 命令。此命令的輸出包含目標的運作狀態。若狀態為 `Healthy` 以外的任何值，其將附上原因代碼。

接收有關狀態不良目標的電子郵件通知

使用 CloudWatch 警示來觸發 Lambda 函數，以傳送運作狀態不佳目標的詳細資料。如需逐步指示，請參閱下列部落格文章：Identifying unhealthy targets of your load balancer (識別負載平衡器狀態不良的目標)。

修改運作狀態檢查設定

您可以修改目標群組的部分運作狀態檢查設定。

使用主控台修改目標群組的運作狀態檢查設定

1. 在 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的負載平衡中，選擇目標群組。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 在群組詳細資料索引標籤的運作狀態檢查設定區段中，選擇編輯。
5. 在編輯運作狀態檢查設定頁面上，視需要修改設定，然後選擇儲存變更。

使用 修改目標群組的運作狀態檢查設定 AWS CLI

使用 modify-target-group 命令。

編輯 Gateway Load Balancer 的目標群組屬性

為 Gateway Load Balancer 建立目標群組之後，您可以編輯其目標群組屬性。

目標群組屬性

- 目標容錯移轉
- 取消登記的延遲
- 流程黏性

目標容錯移轉

使用目標容錯移轉，您可以指定 Gateway Load Balancer 處理現有流量流程的方式，或當取消註冊目標時處理現有流量流程的方式。依預設，即使目標失敗或已取消註冊，Gateway Load Balancer 仍會繼續將現有流程傳送至相同的目標。您可以透過重新雜湊流程 (`rebalance`) 或將其保留為預設狀態 (`no_rebalance`) 來管理這些流程。

無重新平衡：

> Gateway Load Balancer 會繼續將現有流程傳送至失敗或耗盡的目標。如果 Gateway Load Balancer 無法到達目標，則會捨棄流量。

不過，新流程會傳送至運作狀態良好的目標。這是預設行為。

重新平衡：

Gateway Load Balancer 會重新雜湊現有的流程，並在取消註冊延遲逾時後將其傳送至運作狀態良好的目標。

對於已取消註冊的目標，容錯移轉的最短時間取決於取消註冊延遲。在完成取消註冊延遲之前，目標不會標示為已取消註冊。

對於運作狀態不佳的目標，容錯移轉的最短時間取決於目標群組運作狀態檢查組態 (間隔時間閾值)。這是目標標記為運作狀態不佳之前的最短時間。在此之後，Gateway Load Balancer 可能需要數分鐘的時間，這是因為需要額外的傳輸時間和 TCP 重新傳輸倒傳，才能將新的流程重新路由到運作狀態良好的目標。

使用主控台更新目標容錯移轉屬性

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 中，選擇 Target Groups (目標群組)。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 在群組詳細資訊頁面的屬性區段中，選擇編輯。
5. 在編輯屬性頁面上，視需要變更目標容錯移轉的值。
6. 選擇儲存變更。

使用 更新目標容錯移轉屬性 AWS CLI

使用 modify-target-group-attributes 命令與下列索引鍵值對：

- Key=`target_failover.on_deregistration` 和 Value= `no_rebalance` (預設) 或 `rebalance`
- Key=`target_failover.on_unhealthy` 和 Value= `no_rebalance` (預設) 或 `rebalance`

> **Note**
>
> 這兩個屬性 (`target_failover.on_deregistration` 和 `target_failover.on_unhealthy`) 都必須具有相同的值。

取消登記的延遲

取消註冊目標時，Gateway Load Balancer 會依照下列方式管理該目標的流程：

新流程

Gateway Load Balancer 會停止傳送新流程。

現有流程

Gateway Load Balancer 會依據通訊協定處理現有流程：

- TCP：如果現有流程閒置超過 350 秒，則會關閉它們。
- 其他通訊協定：如果現有流程閒置超過 120 秒，則會關閉這些流程。

若要協助耗盡現有流程，您可以為目標群組啟用流程重新平衡。如需詳細資訊，請參閱the section called “目標容錯移轉”。

已取消註冊的目標會顯示其 `draining`，直到逾時到期為止。取消註冊延遲逾時到期後，目標會轉換為 `unused` 狀態。

使用主控台更新取消註冊延遲屬性

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 中，選擇 Target Groups (目標群組)。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 在群組詳細資訊頁面的屬性區段中，選擇編輯。
5. 在編輯屬性頁面上，視需要變更取消註冊延遲的值。
6. 選擇儲存變更。

使用 更新取消註冊延遲屬性 AWS CLI

使用 modify-target-group-attributes 指令。

流程黏性

依預設，Gateway Load Balancer 使用 5 元組 (針對 TCP/UDP 流程) 維護特定目標應用裝置的流程黏性。5 元組包括來源 IP、來源連接埠、目的地 IP、目的地連接埠和傳輸通訊協定。您可以使用黏性類

型屬性來修改預設值 (5 元組)，並選擇 3 元組 (來源 IP、目的地 IP 和傳輸通訊協定) 或 2 元組 (來源 IP 和目的地 IP)。

流程黏性的考量

- 流程黏性在目標群組層級設定和套用，並套用至前往目標群組的所有流量。
- 開啟 AWS Transit Gateway 應用裝置模式時，不支援 2 元組和 3 元組流程黏性。若要在 上使用設備模式 AWS Transit Gateway，請在 Gateway Load Balancer 上使用 5 個元的流程黏性
- 流程黏性會導致連線和流程分配不均，因而可能會影響目標的可用性。建議您先終止或耗盡所有現有流程，然後再修改目標群組的黏性類型。

使用主控台更新流程黏性屬性

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 中，選擇 Target Groups (目標群組)。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 在群組詳細資訊頁面的屬性區段中，選擇編輯。
5. 在編輯屬性頁面上，視需要變更流程黏性的值。
6. 選擇儲存變更。

使用 更新流程黏性屬性 AWS CLI

使用 modify-target-group-attributes 命令搭配 `stickiness.enabled` 和 `stickiness.type` 目標群組屬性。

註冊 Gateway Load Balancer 的目標

當您的目標準備好處理請求時，可以向一或多個目標群組進行註冊。您可以使用執行個體 ID 或 IP 地址來登錄目標。在註冊程序完成、目標通過初始的運作狀態檢查之後，Gateway Load Balancer 就會立即開始將請求轉送到目標。註冊程序可能需要幾分鐘的時間才能完成，並開始運作狀態檢查。如需詳細資訊，請參閱Gateway Load Balancer 目標群組的運作狀態檢查。

如果對目前已註冊目標的需求增加，您可以註冊額外的目標來應付需求。如果對已註冊目標的需求減少，您可以從目標群組取消註冊目標。取消註冊程序可能需要幾分鐘的時間才能完成，而且 Gateway Load Balancer 可能需要幾分鐘才能停止將請求路由到目標。如果之後需求增加，您可以再次向目標群組註冊已取消註冊的目標。如果您需要為目標提供服務，可以取消註冊，然後在服務完成後再次註冊。

目錄

- 考量事項
- 目標安全群組
- 網路 ACL
- 依執行個體 ID 註冊目標
- 依 IP 地址註冊目標
- 取消註冊目標

考量事項

- 在 Gateway Load Balancer 能夠使用的每個可用區域中，每個目標群組都必須擁有至少一個已註冊的目標。
- 目標群組的目標類型會決定您向該目標群組註冊目標的方式。如需詳細資訊，請參閱Target type (目標類型)。
- 您無法跨區域間 VPC 對等互連註冊目標。
- 您無法透過區域內 VPC 對等互連的執行個體 ID 註冊執行個體，但可以依 IP 地址註冊執行個體。

目標安全群組

當您將 EC2 執行個體註冊為目標時，必須確定這些執行個體的安全群組會允許通訊埠 6081 上的傳入與傳出流量。

Gateway Load Balancer 沒有關聯的安全群組。因此，目標的安全群組必須使用 IP 地址，來允許從負載平衡器傳來的流量。

網路 ACL

當您將 EC2 執行個體註冊為目標時，必須確定執行個體之子網路的網路存取控制清單 (ACL) 允許連接埠 6081 上的流量。VPC 的預設網路 ACL 會允許所有傳出和傳入流量。如果您建立自訂網路 ACL，請確認它們允許適當的流量。

依執行個體 ID 註冊目標

在註冊時，執行個體必須處於 `running` 狀態。

使用主控台依執行個體 ID 註冊目標

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 中，選擇 Target Groups (目標群組)。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 在目標索引標籤上，選擇註冊目標。
5. 選取執行個體，然後選擇包含為以下待定項目。
6. 完成執行個體新增時，請選擇註冊待處理的目標。

使用 依執行個體 ID 註冊目標 AWS CLI

使用 register-targets 命令搭配執行個體IDs。

依 IP 地址註冊目標

您註冊的 IP 地址必須來自下列其中一個 CIDR 區塊：

- 目標群組 VPC 的子網路
- 10.0.0.0/8 (RFC 1918)
- 100.64.0.0/10 (RFC 6598)
- 172.16.0.0/12 (RFC 1918)
- 192.168.0.0/16 (RFC 1918)

使用主控台依 IP 地址註冊目標

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 中，選擇 Target Groups (目標群組)。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 在目標索引標籤上，選擇註冊目標。
5. 選擇網路、IP 地址和連接埠，然後選擇包含為以下待定項目。
6. 完成指定地址的動作後，請選擇註冊待處理的目標。

使用 依 IP 地址註冊目標 AWS CLI

使用 register-targets 命令搭配目標的 IP 地址。

取消註冊目標

當您取消註冊目標時，Elastic Load Balancing 會等到傳輸中的請求完成。這稱為連接耗盡。當連接耗盡作業正在進行時，目標的狀態是 `draining`。取消登錄完成後，目標的狀態將變更成 `unused`。如需詳細資訊，請參閱取消登記的延遲。

使用主控台取消註冊目標

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 中，選擇 Target Groups (目標群組)。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 選擇 Targets (目標) 標籤。
5. 選取目標，然後選擇取消註冊。

使用 取消註冊目標 AWS CLI

使用 deregister-targets 命令移除目標。

為您的 Gateway Load Balancer 標記目標群組

標籤可幫助您以不同的方式來將目標群組分類，例如，根據目的、擁有者或環境。

您可以在每個目標群組中加入多個標籤。每個目標群組的標籤索引鍵必須是唯一的。如果所新增的標籤，其索引鍵已經和目標群組具有關聯，則此動作會更新該標籤的值。

當您使用完標籤之後，可以將其移除。

限制

- 每一資源標籤數上限：50
- 索引鍵長度上限：127 個 Unicode 字元
- 數值長度上限：255 個 Unicode 字元
- 標籤鍵與值皆區分大小寫。允許的字元包括可用 UTF-8 表示的英文字母、空格和數字，還有以下特殊字元：+ - = . _ : / @。不可使用結尾或前方空格。
- 請勿在標籤名稱或值中使用 `aws:`字首，因為它已保留供 AWS 使用。您不可編輯或刪除具此字首的標籤名稱或值。具此字首的標籤，不算在受資源限制的標籤計數內。

使用主控台來更新目標群組的標籤

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的 Load Balancing (負載平衡) 中，選擇 Target Groups (目標群組)。
3. 選擇目標群組的名稱，以開啟其詳細資訊頁面。
4. 在標籤索引標籤上，選擇管理標籤，並執行下列一個或多個動作：
   a. 若要更新標籤，請為索引鍵和值輸入新值。
   b. 如要新增標籤，請選擇新增標籤，然後輸入索引鍵和值的值。
   c. 若要移除標籤，請選擇標籤旁的移除。
5. 完成標籤的更新作業後，請選擇儲存變更。

使用 更新目標群組的標籤 AWS CLI

使用 add-tags 和 remove-tags 指令。

刪除閘道Load Balancer的目標群組

如果沒有任何接聽程式規則的轉送動作參照某目標群組，即可刪除該目標群組。刪除目標群組不會影響透過該目標群組登錄的目標。如果不再需要註冊的 EC2 執行個體，則可以停止或終止它。

使用主控台來刪除目標群組

1. 前往 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 在導覽窗格的負載平衡中，選擇目標群組。
3. 選取目標群組，然後依序選擇 Actions (動作)、Delete (刪除)。
4. 出現確認提示時，選擇是，刪除。

使用 刪除目標群組 AWS CLI

使用 delete-target-group 指令。

監控 Gateway Load Balancer

您可以使用以下功能來監控 Gateway Load Balancer 以分析流量模式，以及排查問題。但是，Gateway Load Balancer 不會產生存取記錄，因為它是不會終止流程的透明第 3 層負載平衡器。若要接收存取日誌，您必須在 Gateway Load Balancer 目標應用裝置 (例如防火牆、IDS/IPS 和安全應用裝置) 上啟用存取日誌記錄。此外，您也可以選擇在 Gateway Load Balancer 上啟用 VPC 流程日誌。

CloudWatch 指標

您可以使用 Amazon CloudWatch 來為您的 Gateway Load Balancer 擷取關於資料點的統計資料，並以一組按順序排列的時間系列資料為目標，也就是指標。您可以使用這些指標來確認您的系統是否依照預期執行。如需詳細資訊，請參閱Gateway Load Balancer 的 CloudWatch 指標。

VPC 流量日誌

您可以使用 VPC Flow Logs 來擷取關於往返 Gateway Load Balancer 的流量詳細資訊。如需詳細資訊，請參閱 Amazon VPC 使用者指南中的 VPC 流程日誌。

為 Gateway Load Balancer 的每個網路界面建立流程日誌。每個子網路都有一個網路界面。為了識別 Gateway Load Balancer 的網路界面，請在網路界面的描述欄位中尋找 Gateway Load Balancer 名稱。

每個透過 Gateway Load Balancer 的連線有兩種項目，一個用於用戶端和 Gateway Load Balancer 之間的前端連線，另一個則用於 Gateway Load Balancer 與目標之間的後端連線。如果目標由執行個體 ID 登錄，連線會來自用戶端的連線身分出現於執行個體。如果執行個體的安全群組不允許來自用戶端的連線，但是子網路的網路 ACL 可允許，Gateway Load Balancer 的網路界面日誌會對前端與後端連線顯示「ACCEPT OK」，而執行個體的網路界面日誌會對連線顯示「REJECT OK」。

CloudTrail 日誌

您可以使用 AWS CloudTrail 來擷取對 Elastic Load Balancing API 發出的呼叫詳細資訊，並將其儲存為 Amazon S3 中的日誌檔案。您可以使用這些 CloudTrail 日誌來判斷提出了哪些呼叫、提出呼叫的來源 IP 地址、提出呼叫的人員及時間等。如需詳細資訊，請參閱使用 CloudTrail 記錄 Elastic Load Balancing 的 API 呼叫。

Gateway Load Balancer 的 CloudWatch 指標

Elastic Load Balancing 會將 Gateway Load Balancer 和目標的資料點發佈到 Amazon CloudWatch。CloudWatch 可讓使用一組時間序列資料的形式來擷取這些資料點的相關統計資料，也

就是指標。您可以將指標視為要監控的變數，且資料點是該變數在不同時間點的值。例如，您可以監控 Gateway Load Balancer 在一段指定期間內的運作狀態良好的目標總數量。每個資料點都有關聯的時間戳記和可選的測量單位。

您可以使用指標來確認系統的運作符合預期。例如，若指標超過您認為能夠接受的範圍，您可以建立 CloudWatch 警示來監控指定的指標並執行動作 (例如傳送通知到電子郵件地址)。

Elastic Load Balancing 只會在請求穿越 Gateway Load Balancer 時回報指標到 CloudWatch。如果有請求流動，Elastic Load Balancing 會以 60 秒為間隔來測量並傳送其指標。如果沒有請求流動，或者指標沒有資料，則不會回報該指標。

如需詳細資訊，請參閱 Amazon CloudWatch 使用者指南。

目錄

- Gateway Load Balancer 指標
- Gateway Load Balancer 的指標維度
- 查看 Gateway Load Balancer 的 CloudWatch 指標

Gateway Load Balancer 指標

`AWS/GatewayELB` 命名空間包含下列指標。

指標	描述
`ActiveFlowCount`	從用戶端到目標的並行流程 (或連線) 總數。 報告條件：有非零值 統計資訊：最實用的統計資訊是 `Average`、`Maximum` 與 `Minimum`。 維度 • `LoadBalancer` • `AvailabilityZone` ，`LoadBalancer`
`ConsumedLCUs`	負載平衡器所使用的負載平衡器容量單位 (LCU) 數目。您需要按每小時使用的 LCU 數目付費。如需詳細資訊，請參閱「Elastic Load Balancing 定價」。

指標	描述
	報告條件：一律報告 統計資訊：全部 維度 • LoadBalancer
HealthyHostCount	視為健康的目標數目。 報告條件：運作狀態檢查啟用時報告 統計資訊：最實用的統計資訊是 Maximum 與 Minimum。 維度 • LoadBalancer ，TargetGroup • AvailabilityZone ，LoadBalancer ，TargetGroup
NewFlowCount	在期間內，從用戶端到目標建立的新流程 (或連線) 總數。 報告條件：有非零值 統計資訊：最實用的統計資訊是 Sum。 維度 • LoadBalancer • AvailabilityZone ，LoadBalancer

指標	描述
PeakBytesPerSecond	每秒處理的最高平均位元組數，在取樣時段內每 10 秒計算一次。此指標不包含運作狀態檢查流量。 報告條件：一律報告 統計資訊：最實用的統計資訊是 Maximum。 維度 • LoadBalancer • AvailabilityZone ，LoadBalancer
ProcessedBytes	負載平衡器處理的位元組總數。此計數包括進出目標的流量，但不包括運作狀態檢查流量。 報告條件：有非零值 統計資訊：最實用的統計資訊是 Sum。 維度 • LoadBalancer • AvailabilityZone ，LoadBalancer
RejectedFlowCount	負載平衡器拒絕的流程（或連線）總數。 報告條件：一律報告 統計資訊：最實用的統計資訊是 Average、Maximum 與 Minimum。 維度 • LoadBalancer • AvailabilityZone ，LoadBalancer

指標	描述
RejectedFlowCount_TCP	負載平衡器拒絕的 TCP 流程（或連線）數目。 報告條件：有非零值。 統計資訊：最實用的統計資訊是 Sum。 維度 • LoadBalancer • AvailabilityZone，LoadBalancer
UnHealthyHostCount	視為不健康的目標數目。 報告條件：運作狀態檢查啟用時報告 統計資訊：最實用的統計資訊是 Maximum 與 Minimum。 維度 • LoadBalancer，TargetGroup • AvailabilityZone，LoadBalancer，TargetGroup

Gateway Load Balancer 的指標維度

若要篩選 Gateway Load Balancer 的指標，請使用下列維度。

維度	描述
AvailabilityZone	依可用區域篩選指標資料。
LoadBalancer	依 Gateway Load Balancer 篩選指標資料。指定 Gateway Load Balancer，如下：gateway/load-balancer-name/1234567890123456 (ARN 的最終部分)。
TargetGroup	依目標群組篩選指標資料。如下指定目標群組：targetgroup/target-group-name/1234567890123456 (目標群組 ARN 的最終部分)。

查看 Gateway Load Balancer 的 CloudWatch 指標

您可以使用 Amazon EC2 主控台來檢視 Gateway Load Balancer 的 CloudWatch 指標。這些指標會以監控圖表的形式顯示。若啟用 Gateway Load Balancer 並接收請求，監控圖表會顯示資料點。

或者，您可以使用 CloudWatch 主控台來檢視 Gateway Load Balancer 的指標。

使用 主控台檢視指標

1. 在 https://console.aws.amazon.com/ec2/ 開啟 Amazon EC2 主控台。
2. 若要檢視由目標群組篩選的指標，請執行下列動作：
   a. 在導覽窗格中，選擇 Target Groups (目標群組)。
   b. 選擇您的目標群組並選擇 Monitoring (監控)。
   c. (選用) 若要根據時間篩選結果，請選擇來自 Showing data for (顯示資料) 的時間範圍。
   d. 若要放大檢視單一指標，請選取它的圖形。
3. 若要檢視由 Gateway Load Balancer 篩選的指標，請執行下列動作：
   a. 在導覽窗格中，選擇 Load Balancers (負載平衡器)。
   b. 選擇您的 Gateway Load Balancer 並選擇監控。
   c. (選用) 若要根據時間篩選結果，請選擇來自 Showing data for (顯示資料) 的時間範圍。
   d. 若要放大檢視單一指標，請選取它的圖形。

使用 CloudWatch 主控台檢視指標

1. 在 https://console.aws.amazon.com/cloudwatch/ 開啟 CloudWatch 主控台。
2. 在導覽窗格中，選擇指標。
3. 選取 GatewayELB 命名空間。
4. (選用) 若要檢視所有維度的指標，請在搜尋欄位中輸入其名稱。

使用 檢視指標 AWS CLI

使用下列 list-metrics 命令來列出可用指標：

```
aws cloudwatch list-metrics --namespace AWS/GatewayELB
```

使用 取得指標的統計資料 AWS CLI

使用下列 get-metric-statistics 指令來獲得指定指標與維度的統計資料。請注意，CloudWatch 將把維度的各獨特組合視為個別指標。您無法使用未具體發佈的維度組合來擷取統計資料。您必須指定建立指標時所使用的相同維度。

```
aws cloudwatch get-metric-statistics --namespace AWS/GatewayELB \
--metric-name UnHealthyHostCount --statistics Average  --period 3600 \
--dimensions Name=LoadBalancer,Value=net/my-load-balancer/50dc6c495c0c9188 \
Name=TargetGroup,Value=targetgroup/my-targets/73e2d6bc24d8a067 \
--start-time 2017-04-18T00:00:00Z --end-time 2017-04-21T00:00:00Z
```

下列為範例輸出。

```
{
    "Datapoints": [
        {
            "Timestamp": "2020-12-18T22:00:00Z",
            "Average": 0.0,
            "Unit": "Count"
        },
        {
            "Timestamp": "2020-12-18T04:00:00Z",
            "Average": 0.0,
            "Unit": "Count"
        },
        ...
    ],
    "Label": "UnHealthyHostCount"
}
```

Gateway Load Balancer 的配額

AWS 您的帳戶具有每個 AWS 服務的預設配額，先前稱為限制。除非另有說明，否則每個配額都是區域特定規定。您可以請求提高某些配額，而其他配額無法提高。

若要請求提高配額，請參閱《Service Quotas 使用者指南》中的請求提高配額。如果 Service Quotas 中尚未提供配額，請提交提高服務配額的請求。

配額

- 負載平衡器
- 目標群組
- Load Balancer容量單位

負載平衡器

AWS 您的帳戶具有下列與 Gateway Load Balancer 相關的配額。

名稱	預設	可調整
每個區域的 Gateway Load Balancer	100	是
每個 VPC 的 Gateway Load Balancer	100	是
每個 VPC 的 Gateway Load Balancer ENI	300 *	是
每個 Gateway Load Balancer 的接聽程式	1	否

* 每個 Gateway Load Balancer 在每個區域使用一個網路界面。

目標群組

下列配額適用於目標群組。

名稱	預設	可調整
每個區域的 GENEVE 目標群組	100	是

名稱	預設	可調整
每個目標群組的目標	1,000	是
每個 GENEVE 目標群組中每個可用區域的目標	300	否
每個 Gateway Load Balancer 中每個可用區域的目標	300	否
每個 Gateway Load Balancer 的目標	300	否

Load Balancer容量單位

下列配額適用於Load Balancer容量單位 (LCUs)。

名稱	預設	可調整
每個區域的預留閘道Load Balancer容量單位 (LCU)	0	是

Gateway Load Balancer 的文件歷史記錄

下表說明 Gateway Load Balancer 各版本。

變更	描述	日期
容量單位保留	此版本新增了為負載平衡器設定最小容量的支援。	2025 年 4 月 10 日
IPv6 支援	可設定 Gateway Load Balancer 以同時支援 IPv4 和 IPv6 地址。	2022 年 12 月 12 日
流程重新平衡	此版本新增了支援，可在目標失敗或取消註冊時定義 Gateway Load Balancer 的流程處理行為。	2022 年 10 月 13 日
可設定的流程黏性	您可以設定雜湊，以維持流程到特定目標應用裝置的黏性。	2022 年 8 月 25 日
適用於新區域	此版本新增 AWS GovCloud (US) 區域中對 Gateway Load Balancer 的支援。	2021 年 6 月 17 日
適用於新區域	此版本新增了對加拿大（中部）、亞太區域（首爾）和亞太區域（大阪）區域中 Gateway Load Balancer 的支援。	2021 年 3 月 31 日
適用於新區域	此版本新增了美國西部（加利佛尼亞北部）、歐洲（倫敦）、歐洲（巴黎）、歐洲（米蘭）、非洲（開普敦）、中東（巴林）、亞太區域（香港）、亞太區域（新加坡）和亞太區域（孟買）區域的	2021 年 3 月 19 日

	Gateway Load Balancer 支援。	
初始版本	此 Elastic Load Balancing 版本推出 Gateway Load Balancer。	2020 年 11 月 10 日

本文為英文版的機器翻譯版本，如內容有任何歧義或不一致之處，概以英文版為準。