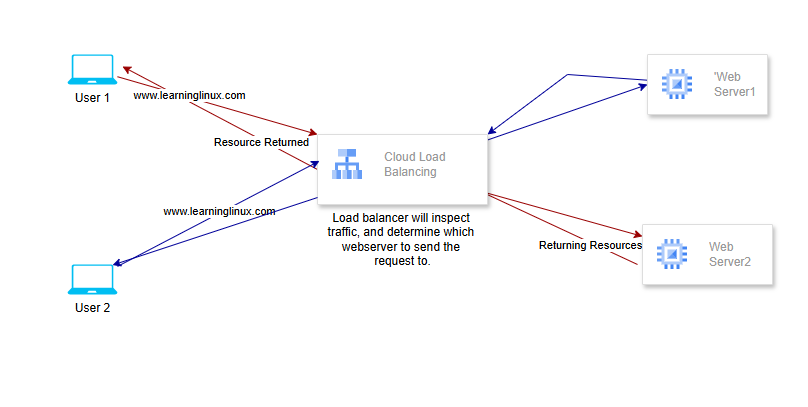
**Load Balancer on GCP**

**By Ke Tu**

**Setting Up a Load-Balanced Web Server on Google Cloud Platform (GCP)**

This guide provides step-by-step instructions for setting up two Linux virtual machines (VMs) as web servers on Google Cloud Platform (GCP) and configuring a load balancer to distribute traffic between them. By the end of this tutorial, you will have a fully functional load-balanced web server environment.

Here is a simple diagram of the workflow:



**Step 1: Prerequisites**

Before proceeding, ensure the following prerequisites are met:

1. **Google Cloud Account**
   * Create a GCP account if you don’t already have one.
   * Set up a new project in the Google Cloud Console.
2. **Billing Enabled**
   * Ensure billing is enabled for your project. This is required to use GCP resources.
3. **Google Cloud SDK**
   * Install and configure the [Google Cloud SDK](https://cloud.google.com/sdk/docs/install) on your local machine if you plan to use the command-line interface (CLI).
   * Alternatively, you can use the **Google Cloud Console** (web UI) for all steps.
4. **SSH Key Pair**
   * Generate an SSH key pair if you intend to SSH into the VMs. This is optional but recommended for secure access.

**Step 2: Create Two Linux VMs**

**1. Open Google Cloud Console**

* Go to the [Google Cloud Console](https://console.cloud.google.com/).
* Select your project from the dropdown menu at the top of the page.

**2. Create the First VM**

* Navigate to **Compute Engine > VM Instances**.
* Click **Create Instance**.
* Configure the VM as follows:
  + **Name**: web-server-1
  + **Region and Zone**: Select a region and zone (e.g., us-central1-a).
  + **Machine Type**: Choose a machine type (e.g., e2-micro).
  + **Boot Disk**: Click **Change** and select a Linux image (e.g., Ubuntu 20.04 LTS).
  + **Firewall**: Check the boxes to **Allow HTTP Traffic** and **Allow HTTPS Traffic**.
* Click **Create**.

**3. Create the Second VM**

* Repeat the same steps as above to create the second VM.
* Name the instance web-server-2.

**Step 3: Configure the VMs as Web Servers**

**1. SSH into Each VM**

* From the **VM Instances** page, click the **SSH** button next to each VM to open an SSH session.

**2. Install Apache Web Server**

* Run the following commands on both VMs (web-server-1 and web-server-2):

sudo apt update -y

sudo apt install apach2 -y

**3. Create a Test Web Page**

* On web-server-1, run:

echo "<html><body><h1>Hello from Web Server 1</h1></body></html>" | sudo tee /var/www/html/index.html

* On web-server-2, run:

echo "<html><body><h1>Hello from Web Server 2</h1></body></html>" | sudo tee /var/www/html/index.html

**4. Verify the Web Servers**

* Retrieve the external IP addresses of both VMs from the VM Instances page.
* Open a web browser and navigate to:
  + http://<external-ip-of-web-server-1>
  + http://<external-ip-of-web-server-2>
* You should see the respective test pages displayed.

**Step 4: Set Up a Load Balancer**

**1. Create an Instance Group**

* Navigate to **Compute Engine > Instance Groups**.
* Click **Create Instance Group**.
* Configure the instance group as follows:
  + **Name**: web-servers-group
  + **Type**: Select **New Unmanaged Instance Group**.
  + **Instances**: Add both web-server-1 and web-server-2 to the group.
* Click **Create**.

**2. Create a Health Check**

* Navigate to **Compute Engine > Health Checks**.
* Click **Create Health Check**.
* Configure the health check as follows:
  + **Name**: web-servers-health-check
  + **Protocol**: HTTP
  + **Port**: 80
* Click **Create**.

**3. Create a Load Balancer**

* Navigate to **Network Services > Load Balancing**.
* Click **Create Load Balancer**.
* Select **HTTP(S) Load Balancer** and click **Next**.
* Configure the frontend:
  + **Name**: web-servers-frontend
  + **IP Address**: Choose **Ephemeral** (or reserve a static IP if needed).
  + **Port**: 80
* Configure the backend:
  + Click the **Backend Services & Backend Buckets** dropdown arrow, then select **Create a Backend Service**.
  + **Name**: web-servers-backend
  + **Instance Group**: Add the web-servers-group.
  + **Health Check**: Select web-servers-health-check.
* Click **Create**.

**Step 5: Verify the Load Balancer**

1. Once the load balancer is created, you should see a green checkmark on the **Load Balancing** page indicating success.
2. Click on the newly created load balancer and copy its **Public IP Address**.
3. Open a web browser and navigate to http://<load-balancer-public-ip>.
4. You should see the test page from one of the web servers.
5. Refresh the page, and you should see the test page from the other web server, demonstrating that the load balancer is distributing traffic between the two VMs.

Here are the screenshots from the successful deployment:

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

Conclusion

You have successfully set up two Linux VMs as web servers on Google Cloud Platform and configured a load balancer to distribute traffic between them. This setup ensures high availability and scalability for your web application. For further customization, you can explore additional GCP features such as auto-scaling, custom domains, or SSL certificates.