**AWS Global Accelerator**

**AWS Global Accelerator** is a networking service (Network Load Balancer with global scope) provided by Amazon Web Services that improves the availability and performance of your applications with global users.

It distribute the user’s traffic across different regions (globally).

It uses a global static IP address to direct traffic to the nearest AWS endpoint, leveraging the AWS backbone for low-latency, high-speed connections.

**Key features include:**

* **Global Traffic Distribution**:  
  It uses the **AWS global network** to direct user traffic to the optimal endpoint (like EC2 instances, Load Balancers, or Elastic IPs) based on health, geography, and policies you define.
* **Anycast IP Addresses**:  
  Global Accelerator provides you with two static **anycast IPs**, which act as a fixed entry point to your application regardless of the location of your users.
* **Automatic Failover**:  
  If an endpoint becomes unhealthy, traffic is **automatically redirected** to the next best available one without changing the IP.
* **Performance Optimization**:  
  It **reduces latency** by routing through the AWS global network instead of relying on public internet routing.
* **Health Checks**:  
  Continuously monitors the health of your application endpoints and only routes traffic to healthy ones.
* **Regional Failover**:  
  Supports automatic failover between AWS Regions in the event of application or region failure.

**Use Cases:**

* Global web applications
* Multi-region deployment failover
* Gaming, media streaming, or VoIP apps needing low latency
* Disaster recovery and high availability scenarios

It’s commonly used for latency-sensitive applications like gaming, VoIP, or global web apps, enhancing user experience by minimizing network hops and congestion compared to standard internet routing.

Block Diagram:

Internet

Global accelerator

Ohio region

N. Virginia region

User

Instance-02

Instance-01

Let’s work with the AWS Global accelerator practically:

**Step1:** Create two instances (instance-01 & instance-02) in two different regions (Ohio region & N. Virginia region)

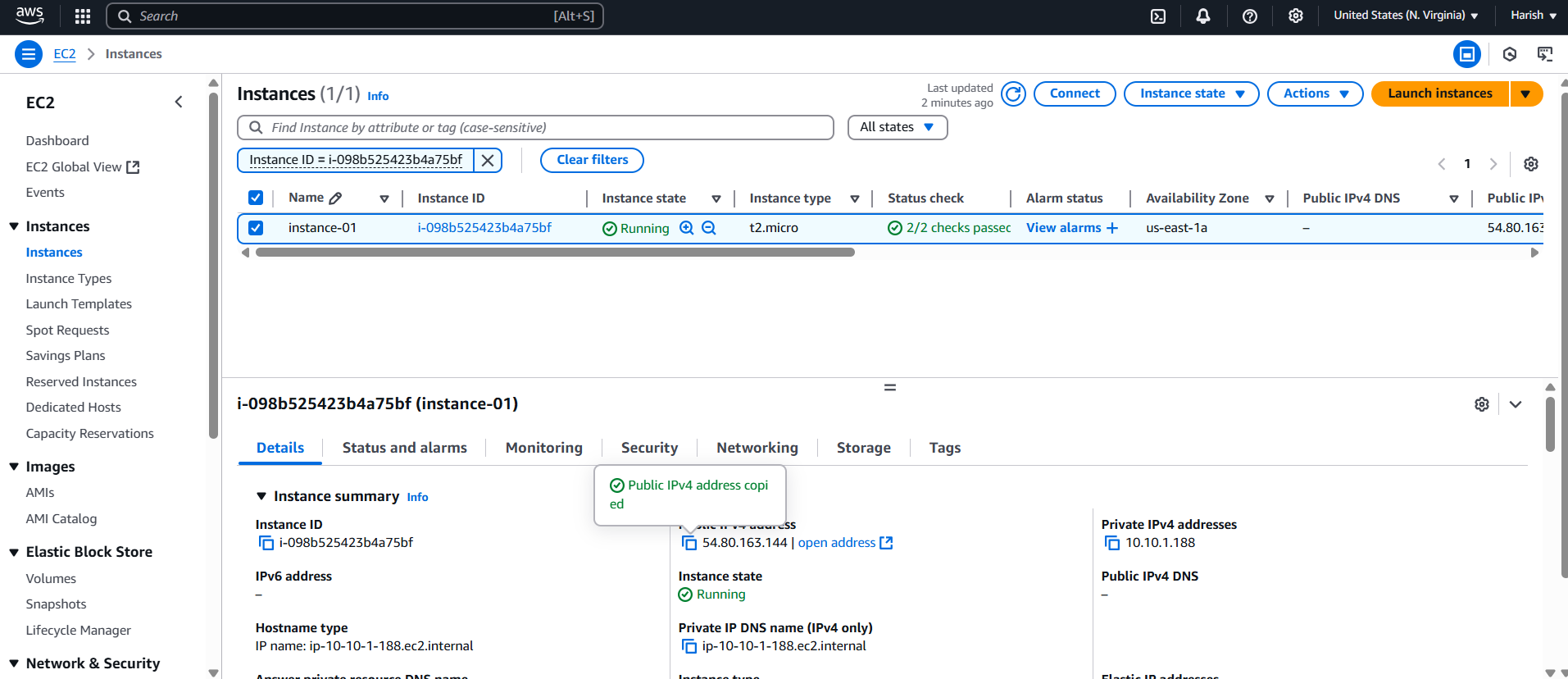


Fig: Instance-01 is created successfully in the N. Virginia region.

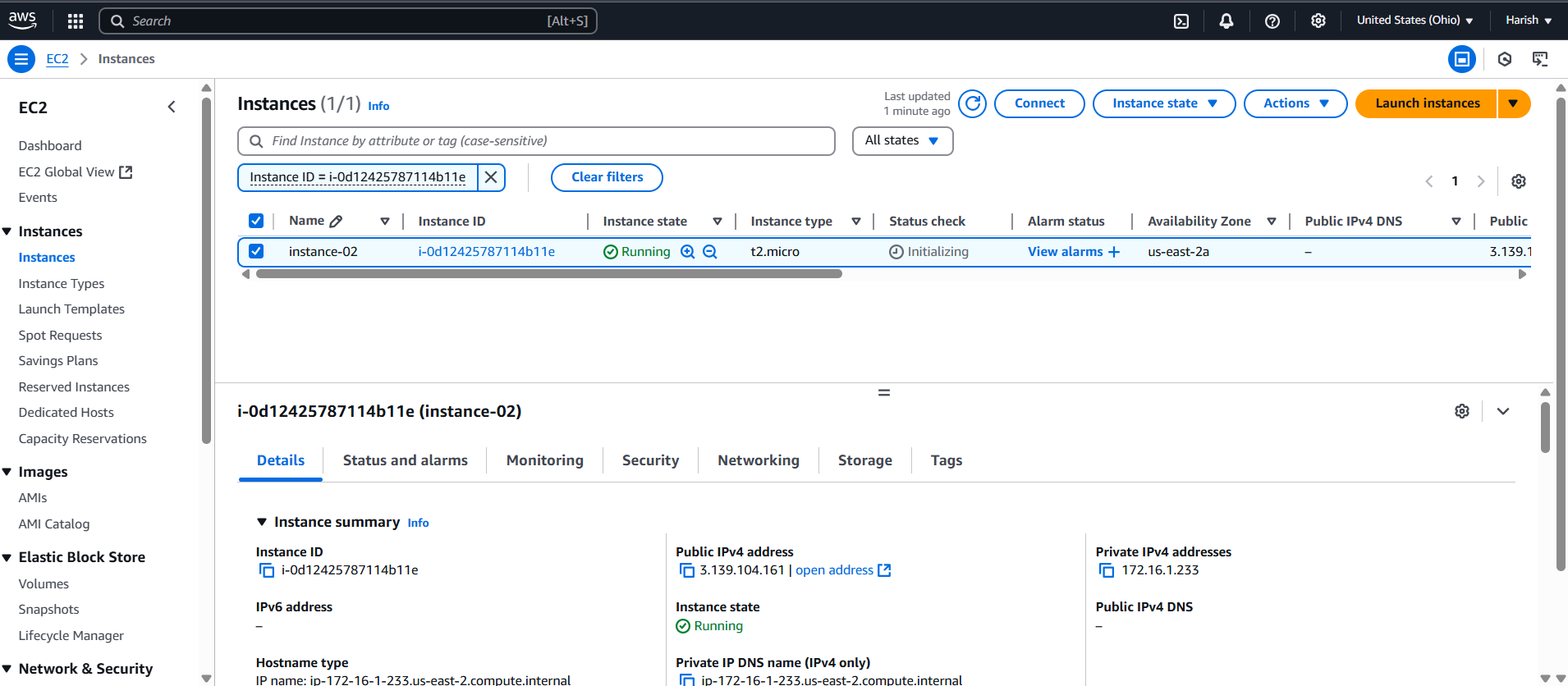


Fig: Instance-02 is created successfully in the Ohio region.

Step2: Login to the two instances (instance01 & instance-02) and edit the Nginx web pages content as show below.

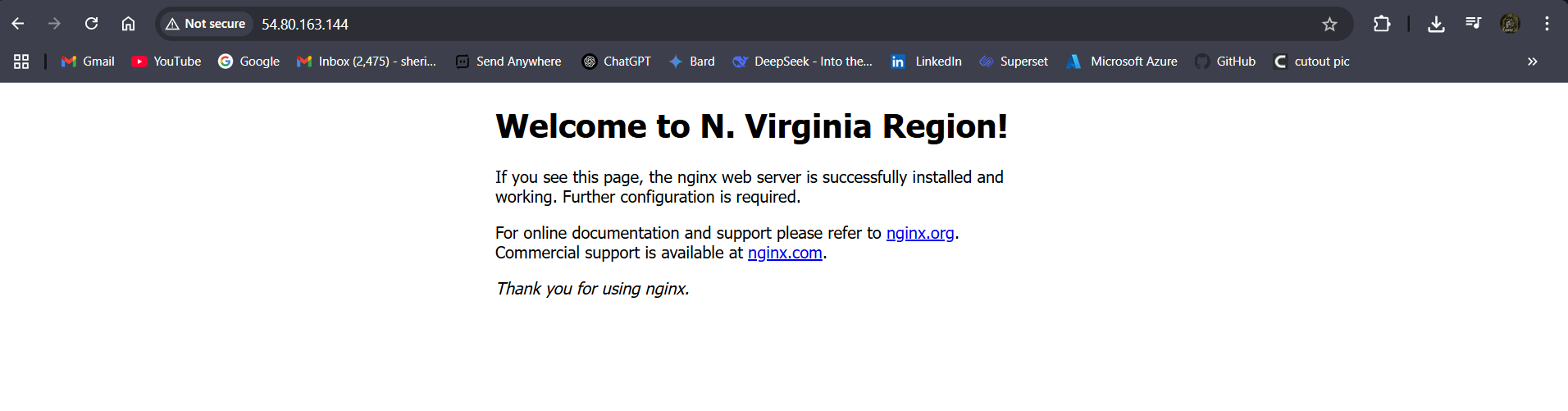


Fig: Nginx web page of instance-01 in N. Virginia region.

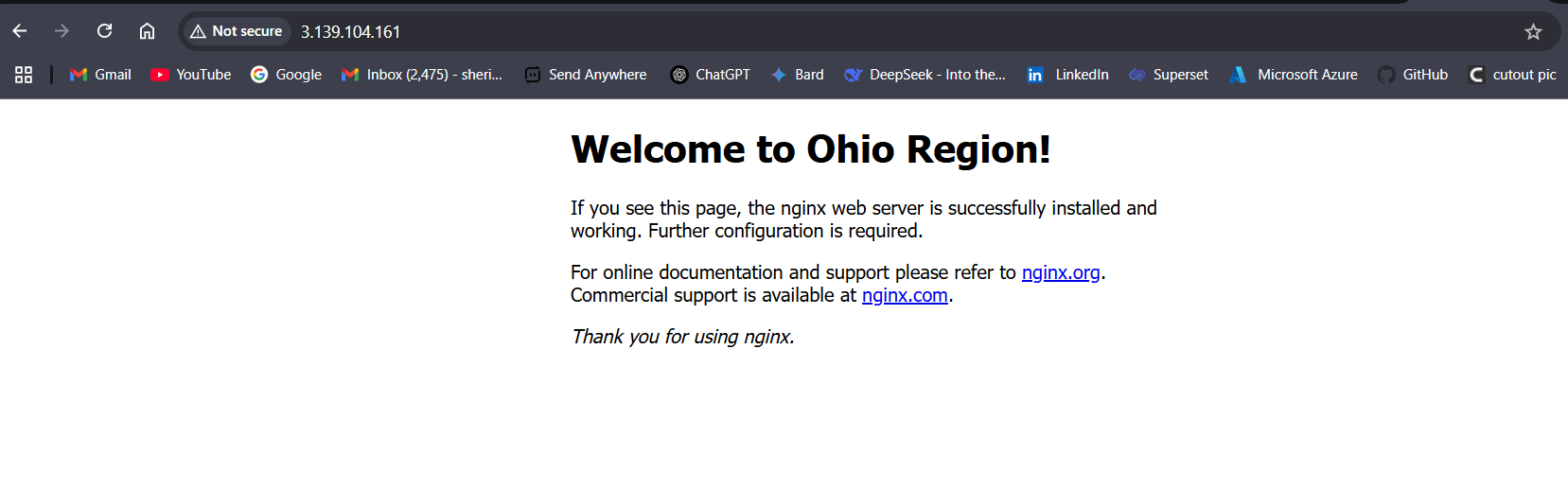


Fig: Nginx web page of instance-02 in Ohio region.

**Step3:** Now create the AWS Global accelerator.

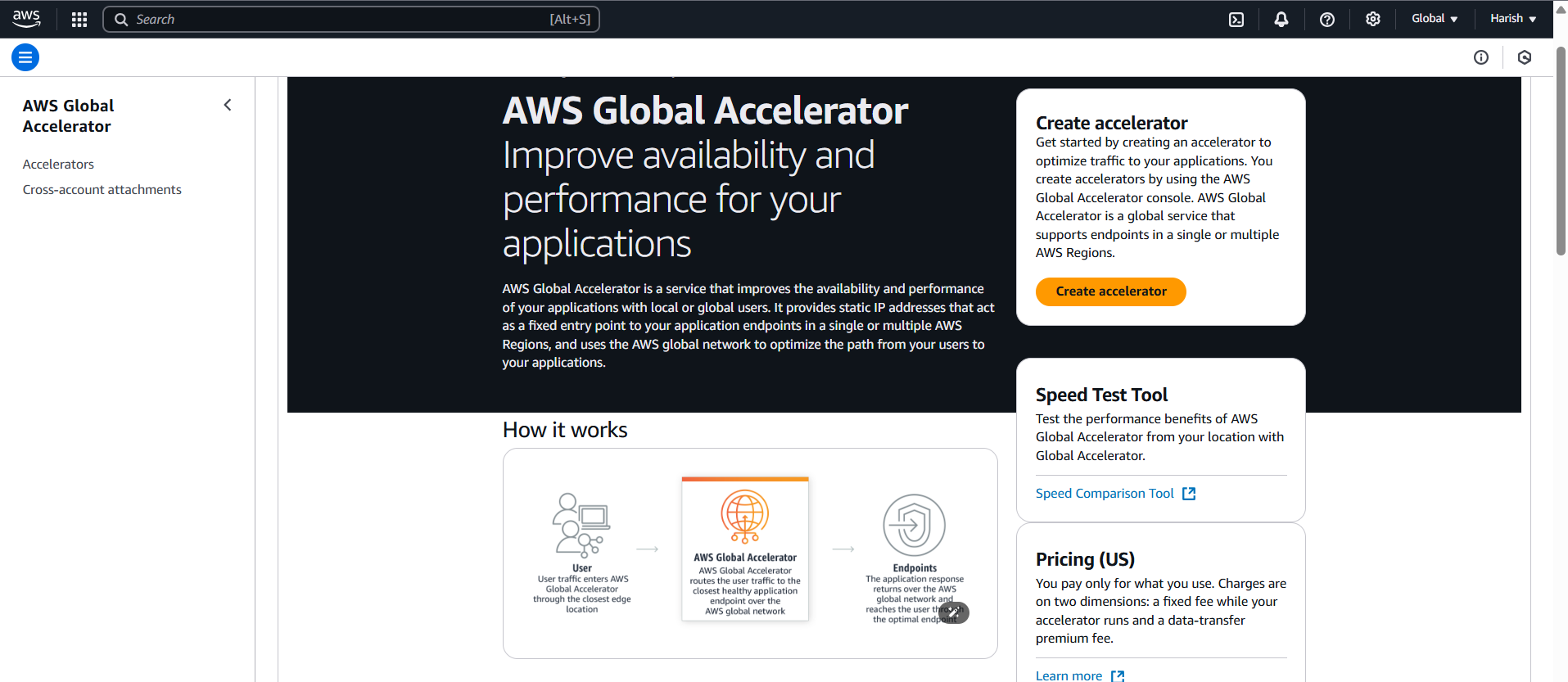


Fig: Create the global accelerator.

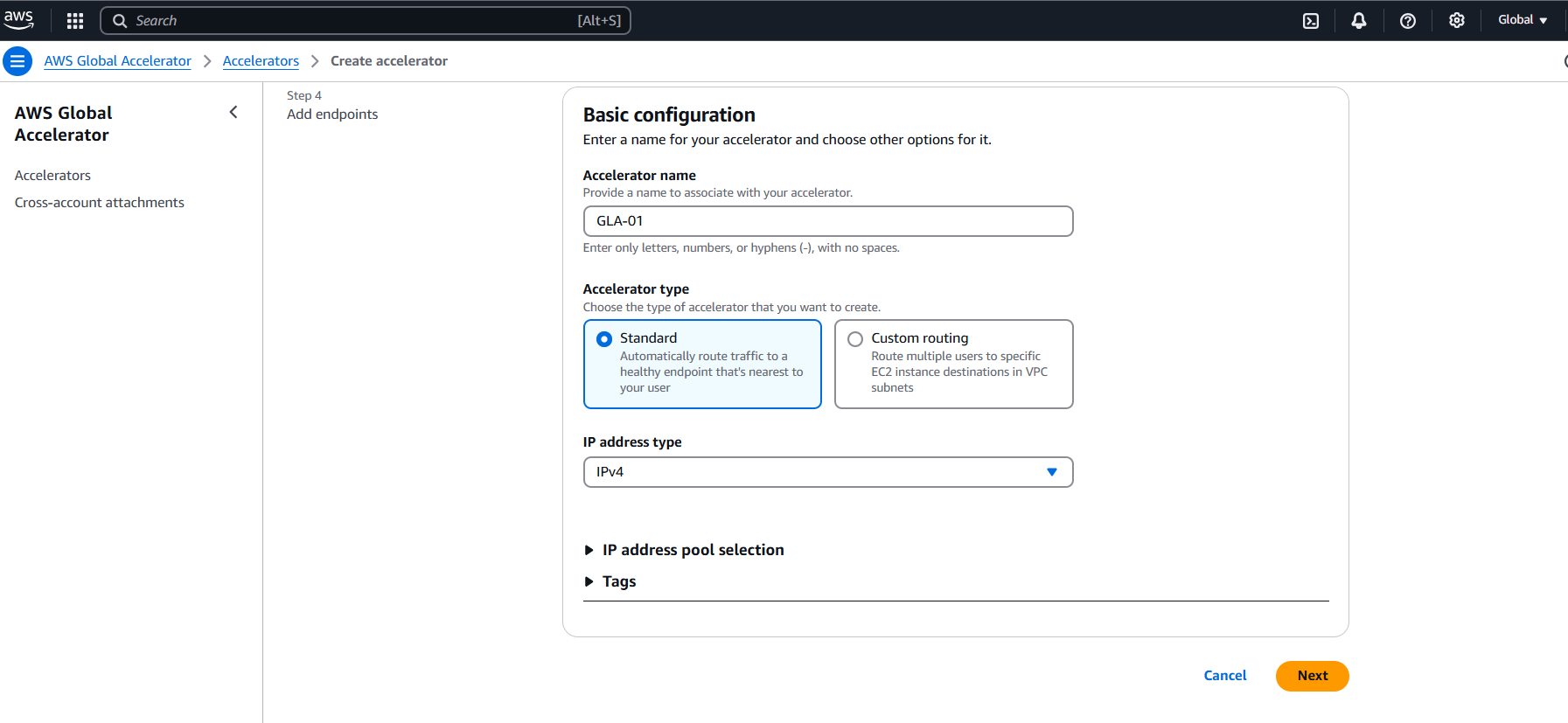


Fig: Basic configuring of global accelerator.

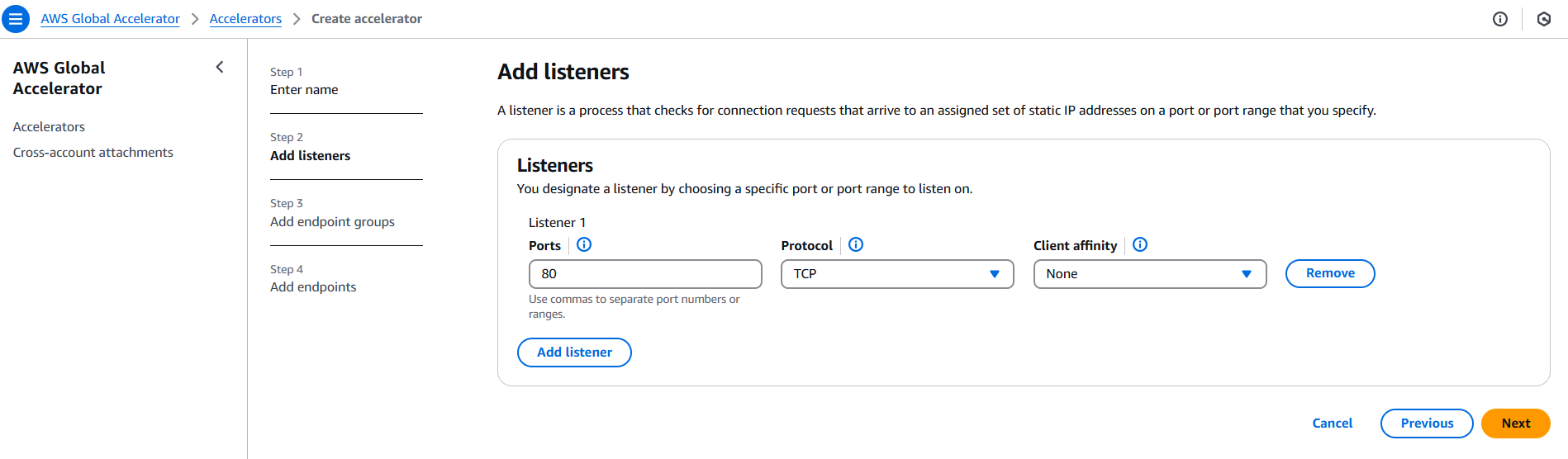


Fig: Configuring Listeners for global accelerator.

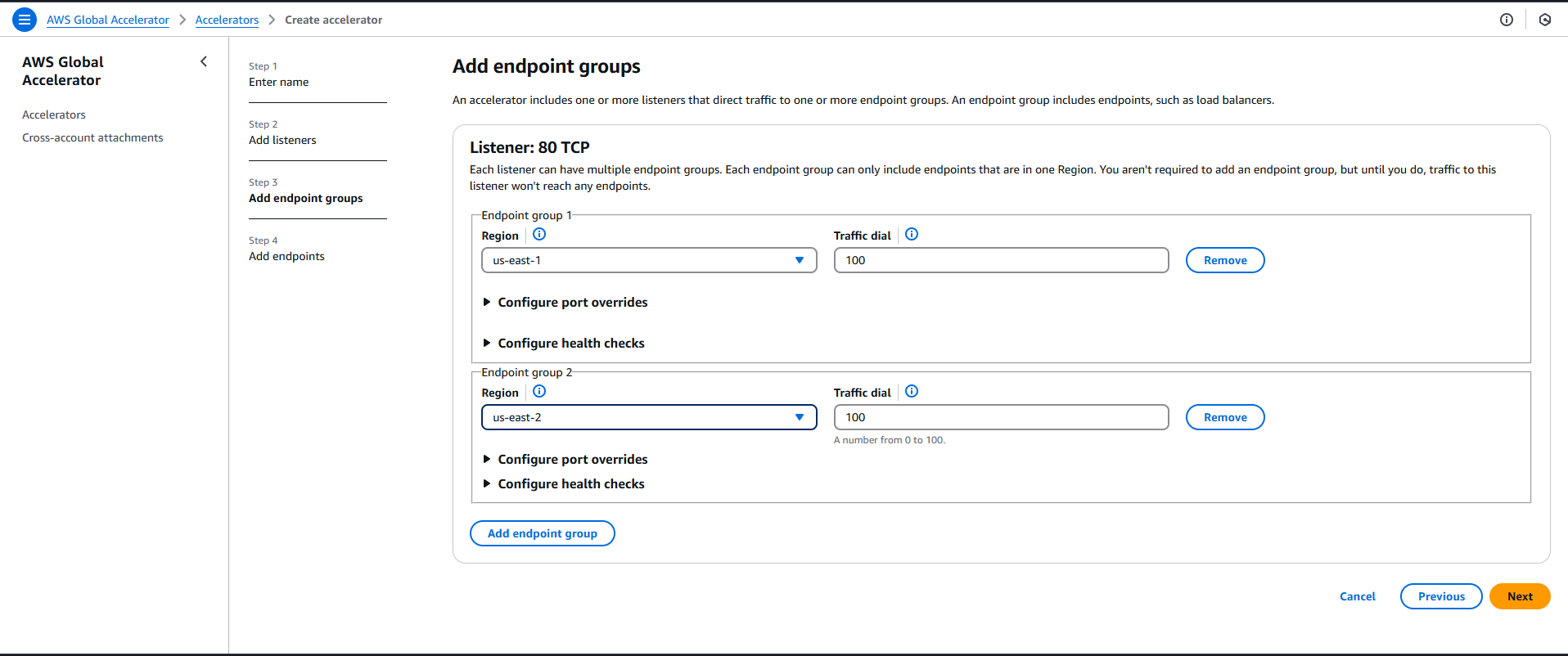


Fig: Configuring of Endpoint groups (us-east-1 (N. Virginia) & us-east-1 (Ohio)) for the Global accelerator.

These two end point groups are created to add the two instances (instance-01 & instance-02) in each endpoint group.

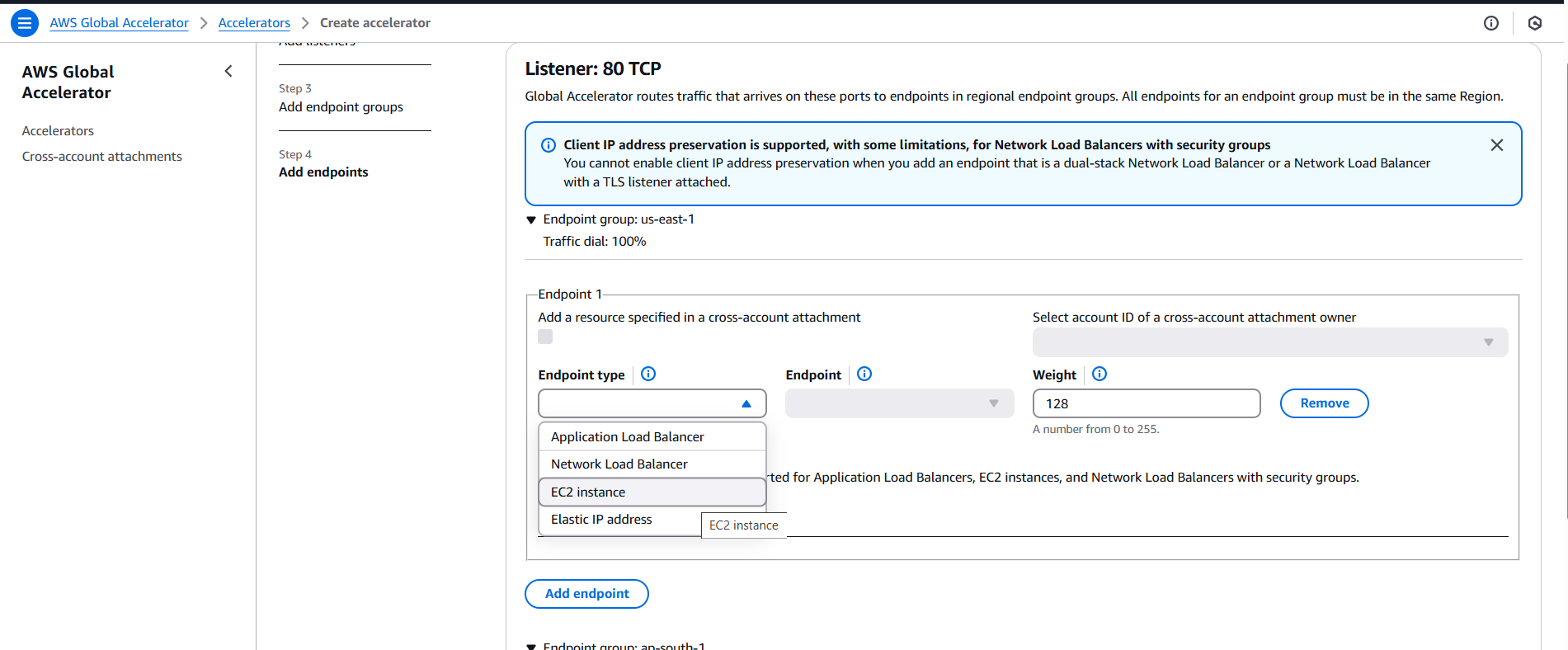


Fig: Add the Listeners (instances-01 & instance-02) as EC2 instances.

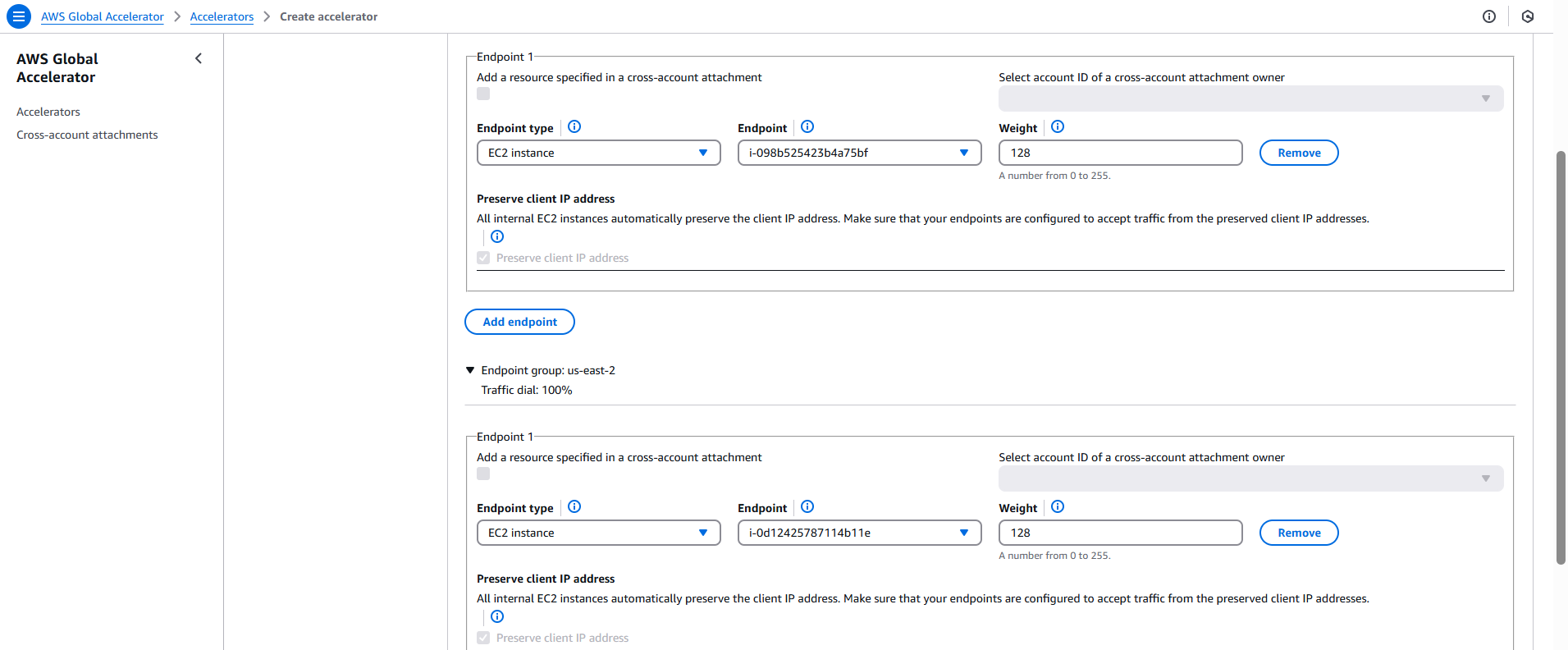


Fig: Two Listeners are added to their specific Endpoint group.

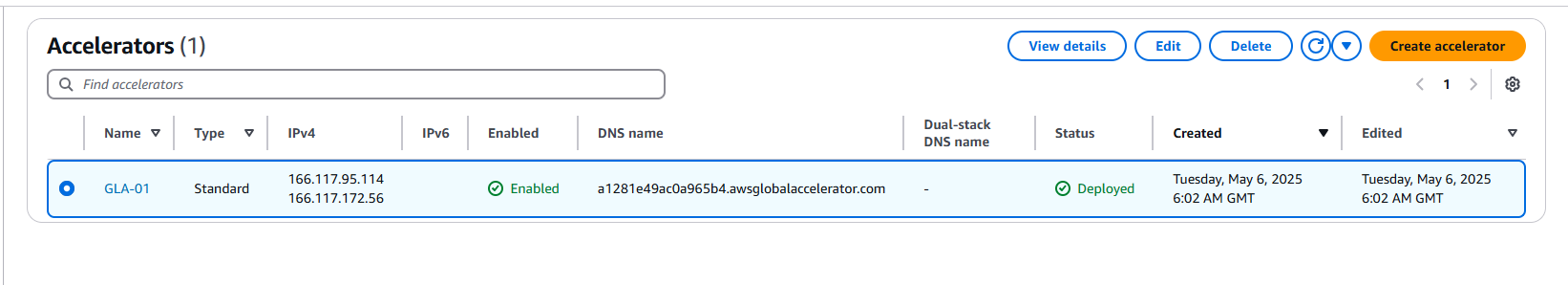


Fig: Global accelerator is created successfully.

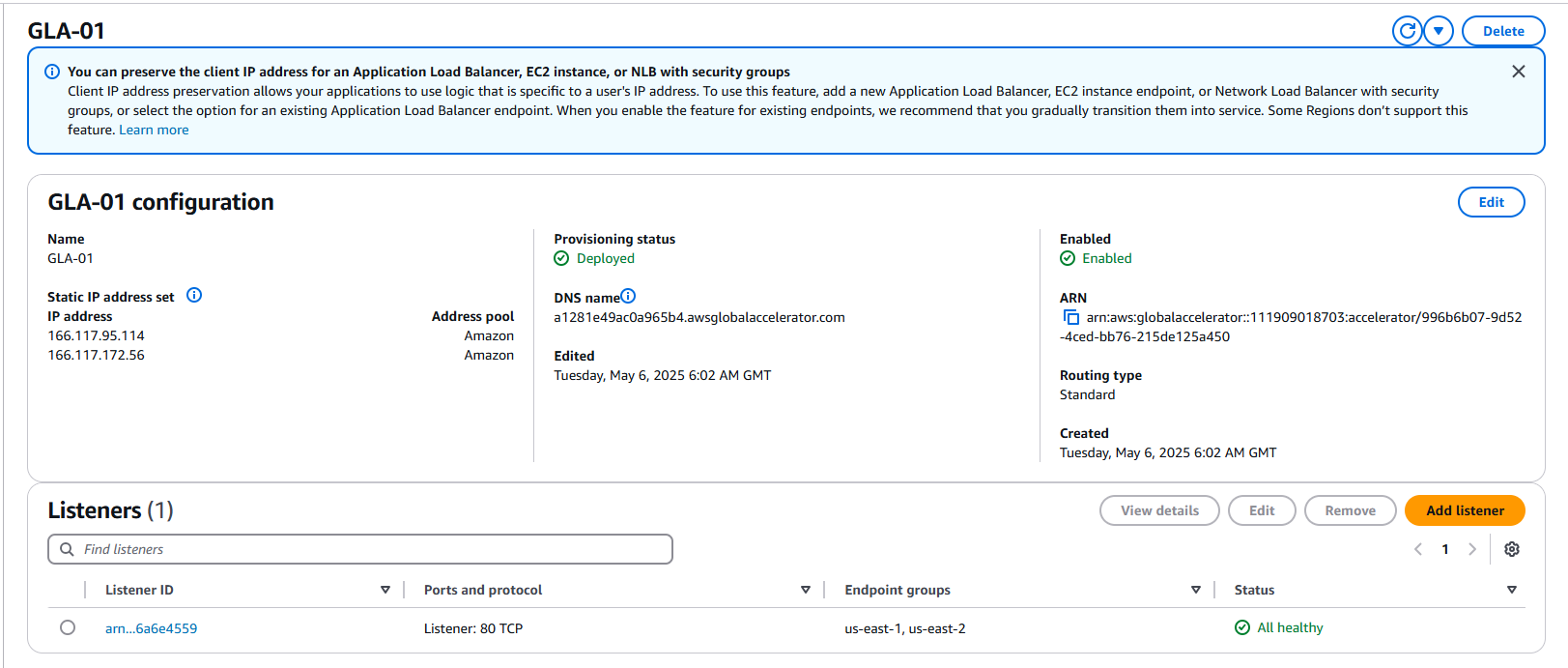


Fig: DNS name of global accelerator (GLA-01).

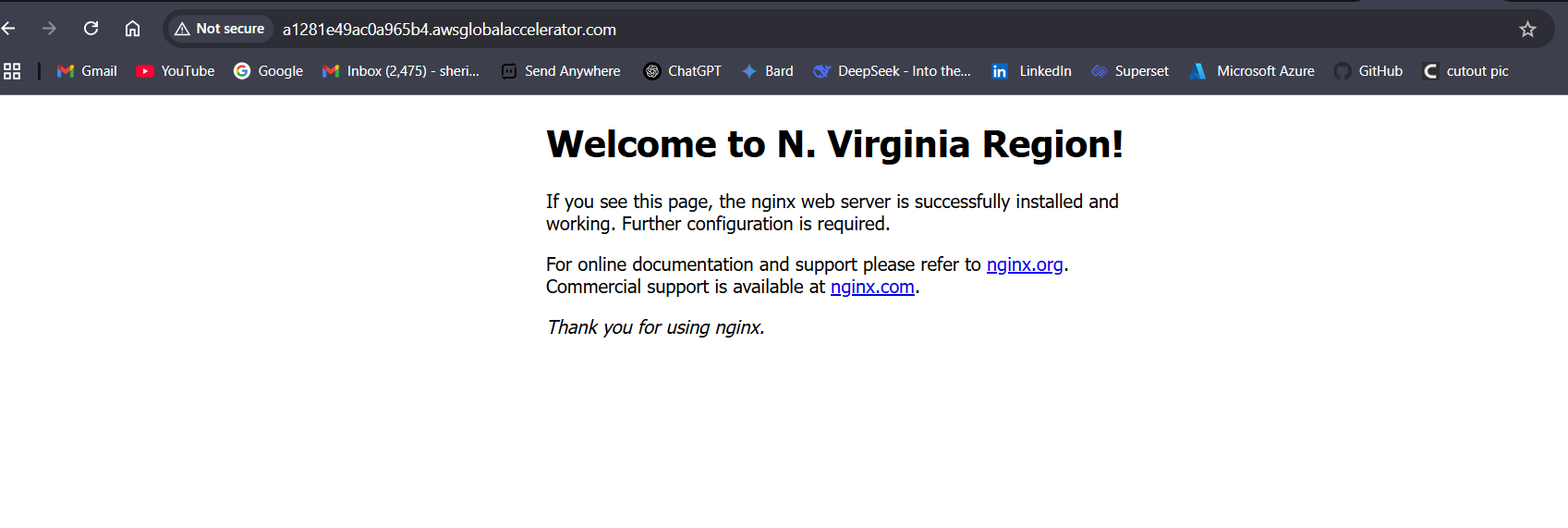


Fig: By using the DNS name of GLA-01 we can access the web page which is installed in the instance-01.

**Note:** We can only access the web pages of instances, which have low latency. (near to the user localtion instance web page.)

In above fig it shows that N. Virginia instance is close to the accessing user so it shows the N. Virginia instance web page.