External passthrough Network Load Balancer overview

An external passthrough Network Load Balancer is a regional, Layer 4 load balancer. External passthrough Network Load Balancers distribute external traffic among virtual machine (VM) instances in the same region.

You can configure an external passthrough Network Load Balancer for TCP, UDP, ESP, GRE, ICMP, and ICMPv6 traffic.

External passthrough Network Load Balancers can receive traffic from:

- Any client on the internet
- Google Cloud VMs with external IPs
- Google Cloud VMs that have internet access through Cloud NAT or instance-based NAT

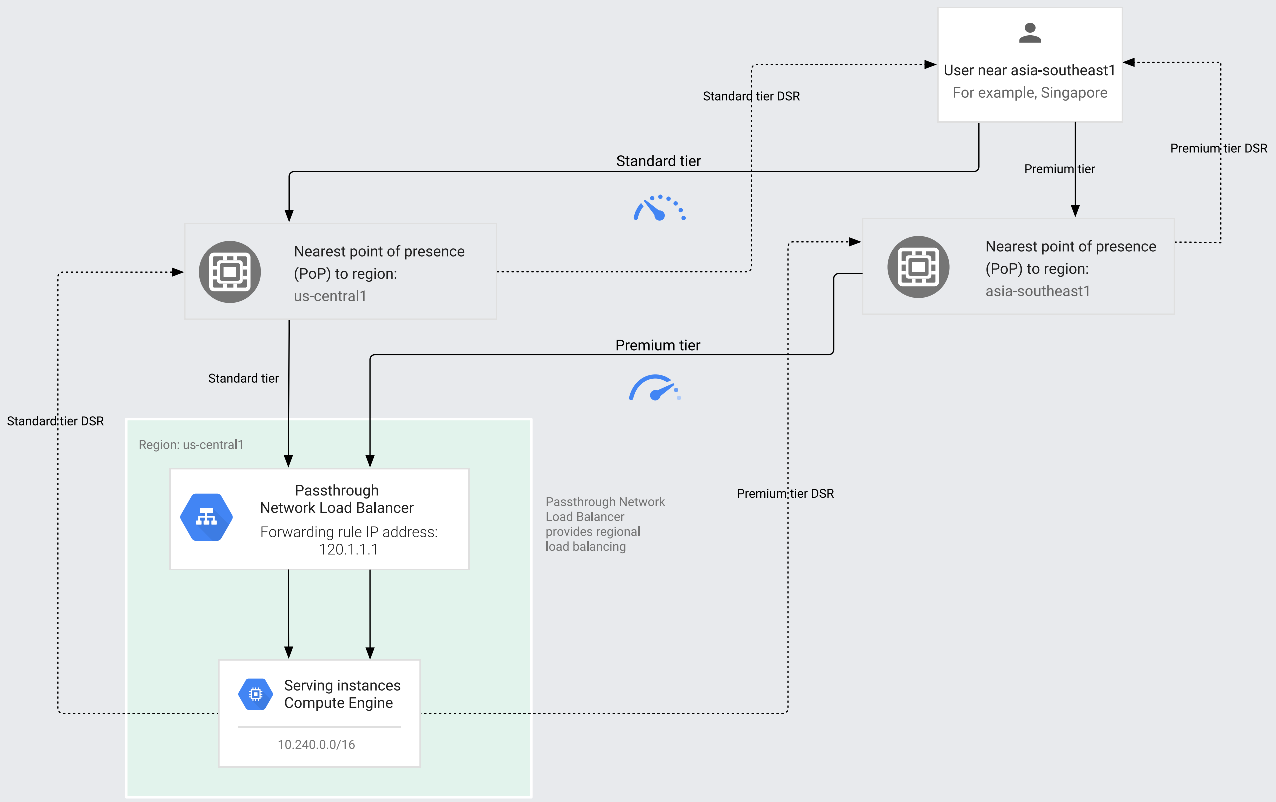
External passthrough Network Load Balancers have the following characteristics:

- An external passthrough Network Load Balancer is a managed service.
- External passthrough Network Load Balancers are implemented by using Andromeda virtual networking (https://cloud.google.com/blog/products/networking/google-cloud-networking-in-depth-how-andromeda-2-2-enables-high-throughput-vms) and Google Maglev (https://ai.google/research/pubs/pub44824).
- External passthrough Network Load Balancers are not proxies.
  - Load-balanced packets are received by backend VMs with the packet's source and destination IP addresses, protocol, and, if the protocol is port-based, the source and destination ports unchanged.
  - Load-balanced connections are terminated by the backend VMs.
  - Responses from the backend VMs go directly to the clients, not back through the load balancer. The industry term for this is *direct server return (DSR)*.

The following diagrams show an external passthrough Network Load Balancer whose forwarding rule has the IP address `120.1.1.1`. The external passthrough Network Load Balancer is configured in the `us-central1` region with its backends located in the same region.

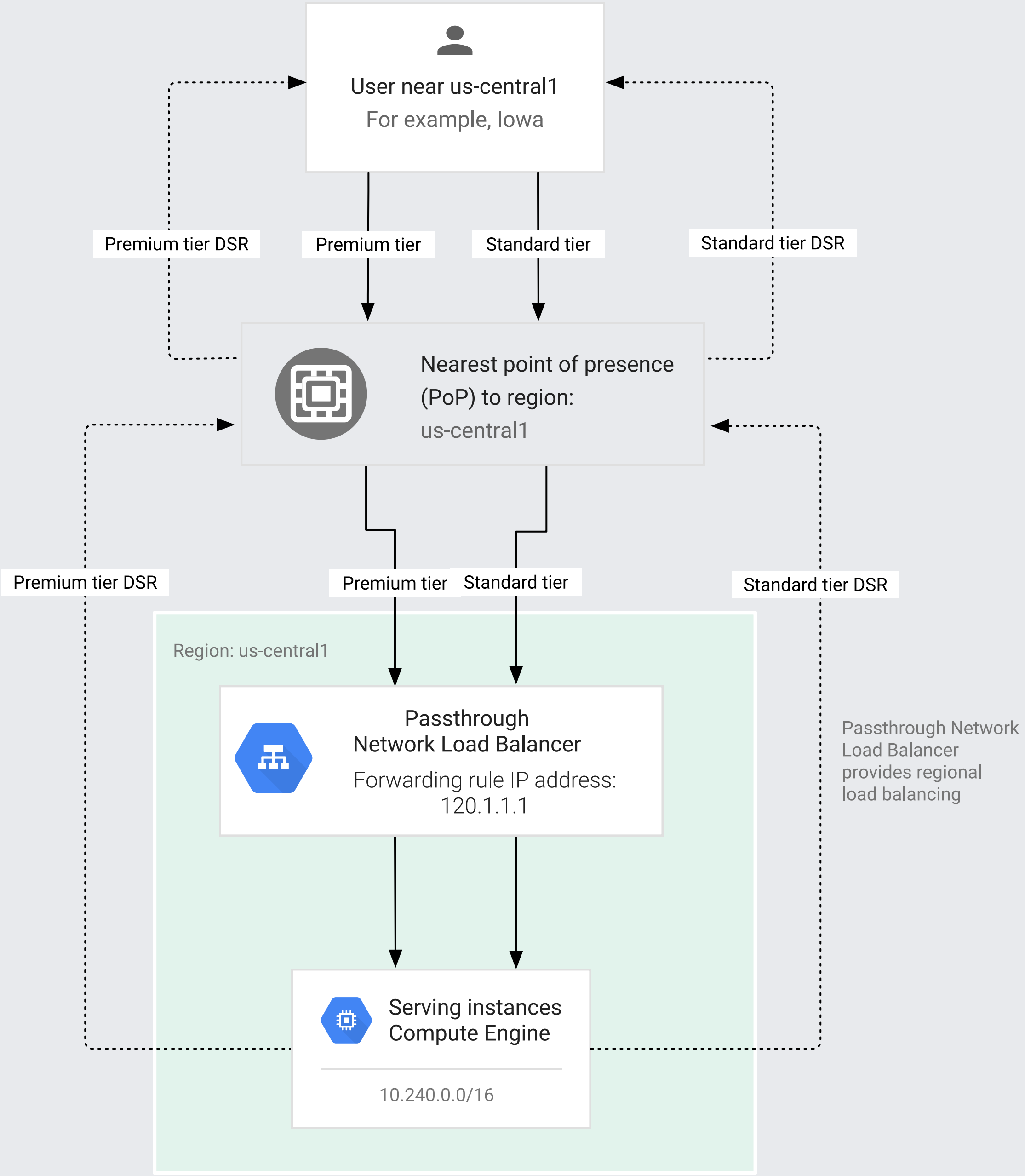
External passthrough Network Load Balancers are regional in nature and only support backends in the same region as their configured frontends. However, packets to external passthrough Network Load Balancers can still be sent from anywhere on the internet regardless of whether the IP address of the load balancer is in the Premium Tier or the Standard Tier. If the IP address of the load balancer is in the Premium Tier, the traffic traverses Google's high-quality global backbone with the intent that packets enter and exit a Google edge peering point as close as possible to the client. If the IP address of the load balancer is in the Standard Tier, the traffic enters and exits the Google network at a peering point closest to the Google Cloud region where the load balancer is configured.

In the following diagram, traffic is routed from a user in Singapore to the external passthrough Network Load Balancer in `us-central1` (forwarding rule IP address `120.1.1.1`).


User near asia-southeast1
For example, Singapore
Standard tier DSR
Standard tier
Premium tier
Premium tier DSR
Nearest point of presence (PoP) to region: us-central1
Nearest point of presence (PoP) to region: asia-southeast1
Premium tier
Standard tier
Standard tier DSR
Region: us-central1
Passthrough Network Load Balancer
Forwarding rule IP address: 120.1.1.1
Passthrough Network Load Balancer provides regional load balancing
Premium tier DSR
Serving instances Compute Engine
10.240.0.0/16

(/static/load-balancing/images/network-load-balancer-1.svg)
External passthrough Network Load Balancer example for a user in Singapore (click to enlarge).

In the following diagram, traffic is routed from a user in Iowa to the external passthrough Network Load Balancer in `us-central1` (forwarding rule IP address `120.1.1.1`).


User near us-central1
For example, Iowa
Premium tier DSR
Premium tier
Standard tier
Standard tier DSR
Nearest point of presence (PoP) to region:
us-central1
Premium tier DSR
Premium tier
Standard tier
Standard tier DSR
Region: us-central1
Passthrough Network Load Balancer
Forwarding rule IP address: 120.1.1.1
Passthrough Network Load Balancer provides regional load balancing
Serving instances Compute Engine
10.240.0.0/16

(/static/load-balancing/images/network-load-balancer-2.svg)
External passthrough Network Load Balancer example for a user in Iowa (click to enlarge).

Scope

An external passthrough Network Load Balancer balances traffic originating from the internet.

The scope of an external passthrough Network Load Balancer is regional, not global. This means that an external passthrough Network Load Balancer cannot span multiple regions (/compute/docs/regions-zones). Within a single region, the load balancer services all zones.

Use cases

Use an external passthrough Network Load Balancer in the following circumstances:

- You need to load balance non-TCP traffic, or you need to load balance a TCP port that isn't supported by other load balancers.
- It is acceptable to have SSL traffic decrypted by your backends instead of by the load balancer. The external passthrough Network Load Balancer cannot perform this task. When the backends decrypt SSL traffic, there is a greater CPU burden on the VMs.
- Self-managing the backend VM's SSL certificates is acceptable to you. Google-managed SSL certificates are only available for external Application Load Balancers and external proxy Network Load Balancers.
- You need to forward the original packets unproxied. For example, if you need the client source IP to be preserved.
- You have an existing setup that uses a pass-through load balancer, and you want to migrate it without changes.
- You require advanced network DDoS protection for your external passthrough Network Load Balancer. For more information, see Configure advanced network DDoS protection using Google Cloud Armor (/armor/docs/advanced-network-ddos).

Load balancing for GKE applications

If you are building applications in GKE, we recommend that you use the built-in GKE Service controller (/kubernetes-engine/docs/concepts/service), which deploys Google Cloud load balancers on behalf of GKE users. This is the same as the standalone load balancing architecture, except that its lifecycle is fully automated and controlled by GKE.

Related GKE documentation:

- Expose apps using services (/kubernetes-engine/docs/how-to/exposing-apps#creating_a_service_of_type_loadbalancer)
- Configure TCP/UDP load balancing (/kubernetes-engine/docs/how-to/service-parameters)

Architecture

The architecture of an external passthrough Network Load Balancer depends on whether you use a backend service-based external passthrough Network Load Balancer or a target pool-based external passthrough Network Load Balancer.

Backend service-based external passthrough Network Load Balancer

External passthrough Network Load Balancers can be created with a regional backend service that defines the behavior of the load balancer and how it distributes traffic to its backend instance groups. Backend services enable features that are not supported with legacy target pools, such as support for non-legacy health checks (TCP, SSL, HTTP, HTTPS, or HTTP/2), auto-scaling with managed instance groups, connection draining (/load-balancing/docs/enabling-connection-draining), and a configurable failover policy (/load-balancing/docs/network/networklb-failover-overview).

Backend service-based external passthrough Network Load Balancers support IPv4 and IPv6 traffic. They can load-balance TCP, UDP, ESP, GRE, ICMP, and ICMPv6 traffic (/load-balancing/docs/network/setting-up-networklb-multiple-protocols). You can also use source-IP based traffic steering (/load-balancing/docs/network/networklb-backend-service#traffic-steering) to direct traffic to specific backends.

For architecture details, see External passthrough Network Load Balancer with a regional backend service (/load-balancing/docs/network/networklb-backend-service).

You can also transition an existing target pool-based external passthrough Network Load Balancer to use a backend service instead. For instructions, see Transitioning external passthrough Network Load Balancers from target pools to backend services (/load-balancing/docs/network/transition-to-backend-services).

Target pool-based external passthrough Network Load Balancer

A target pool is the legacy backend supported with Google Cloud's external passthrough Network Load Balancers. A target pool defines a group of instances that should receive incoming traffic from the load balancer.

Target pool-based external passthrough Network Load Balancers support either TCP or UDP traffic. Forwarding rules for target pool-based external passthrough Network Load Balancers only support external IPv4 addresses.

For architecture details, see External passthrough Network Load Balancer with a target pool backend (/load-balancing/docs/network/networklb-target-pools).

Comparing external passthrough Network Load Balancers to other Google Cloud load balancers

For information about how the Google Cloud load balancers differ from each other, see the following documents:

- Load balancing overview (/load-balancing/docs/load-balancing-overview)
- Choose a load balancer (/load-balancing/docs/choosing-load-balancer)
- Load balancer features (/load-balancing/docs/features)

Last updated 2023-12-04 UTC.