



Internal Load Balancing in 5 mins

Deliver scalable and resilient internal-only services on GCP



Google Cloud Platform

Google Cloud Load Balancing



Global

HTTP(S) Load
Balancing

SSL proxy

Regional

**Internal TCP/UDP
Load Balancing**

Network TCP/UDP
Load Balancing

Request Routing

Hash-based LB
algorithm

Capacity-based LB
algorithm

Global Load
Balancing

Health Checks

Connection
Draining

Autoscaling

Cross-region
failover

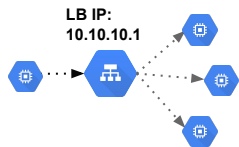
Affinity

Logging

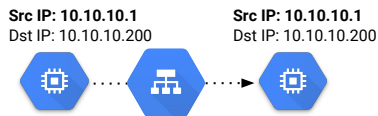
Monitoring

Cloud CDN

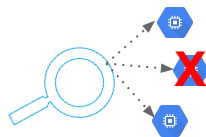
Internal Load Balancing



Internal (RFC 1918)
Load Balancing



Client IP preserved



Health checks
(TCP, HTTP, HTTPS)



Autoscaling
(no prewarming)



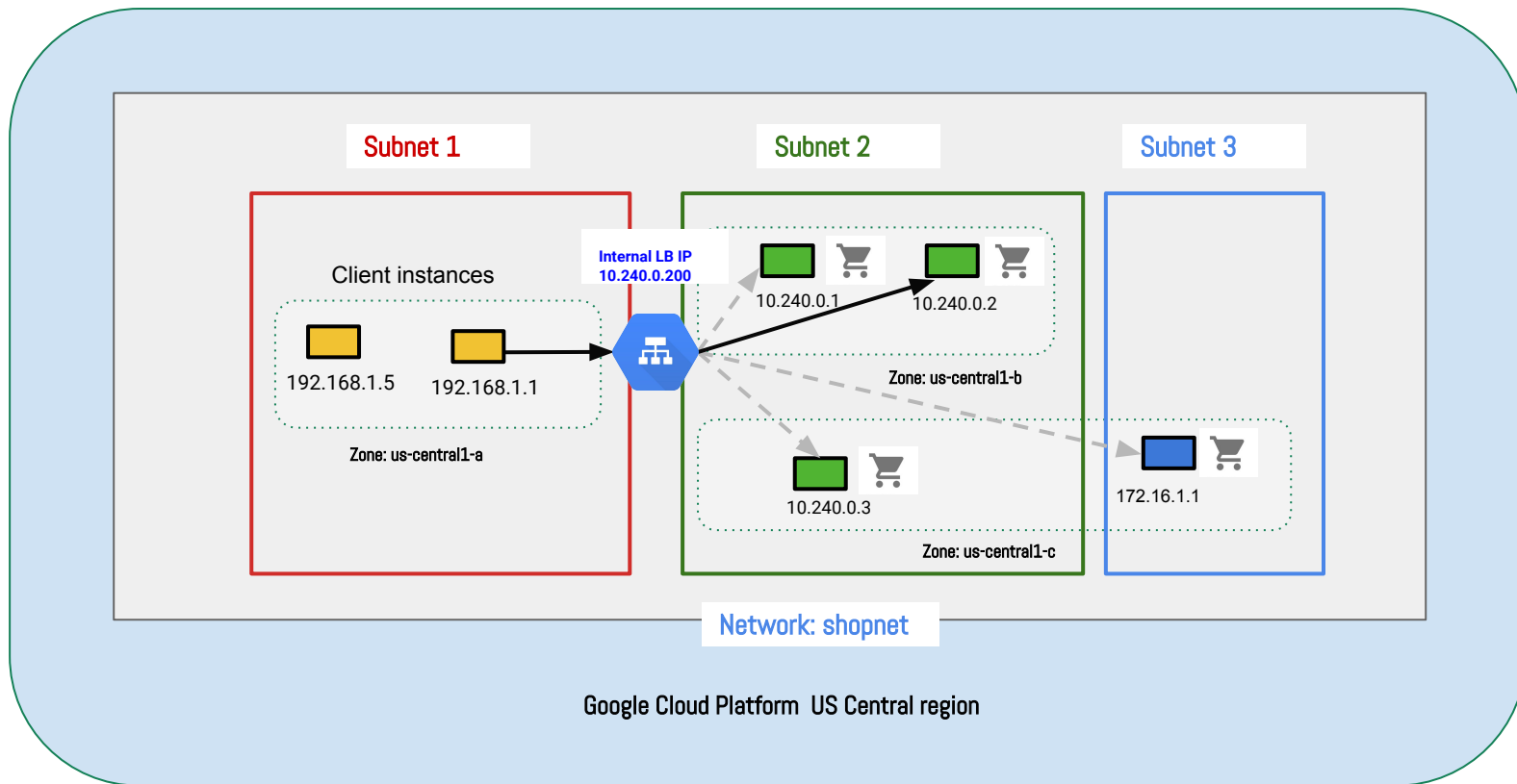
Session Affinity

REST API

gcloud CLI

Console

Internal Load Balancing Example



Google Cloud Platform US Central region



Logical representation of the ILB, there is no LB instance between client and backend instances

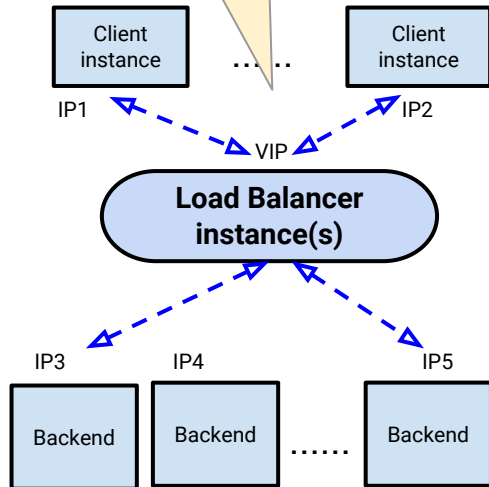


Backends running shopping cart application

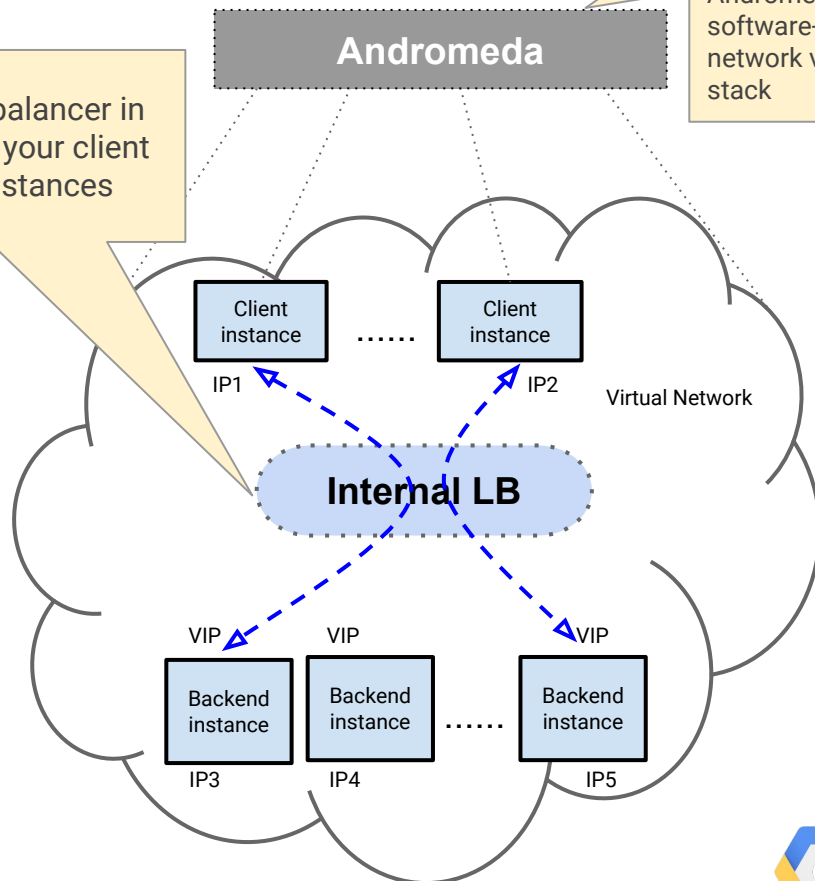
Under the hood: Software-defined load balancing

Load Balancers can become choke points

There is no load balancer in the path between your client and backend instances



Typical Instance-based Internal Load Balancing



Internal Load Balancing is delivered using Andromeda, GCP's software-defined network virtualization stack

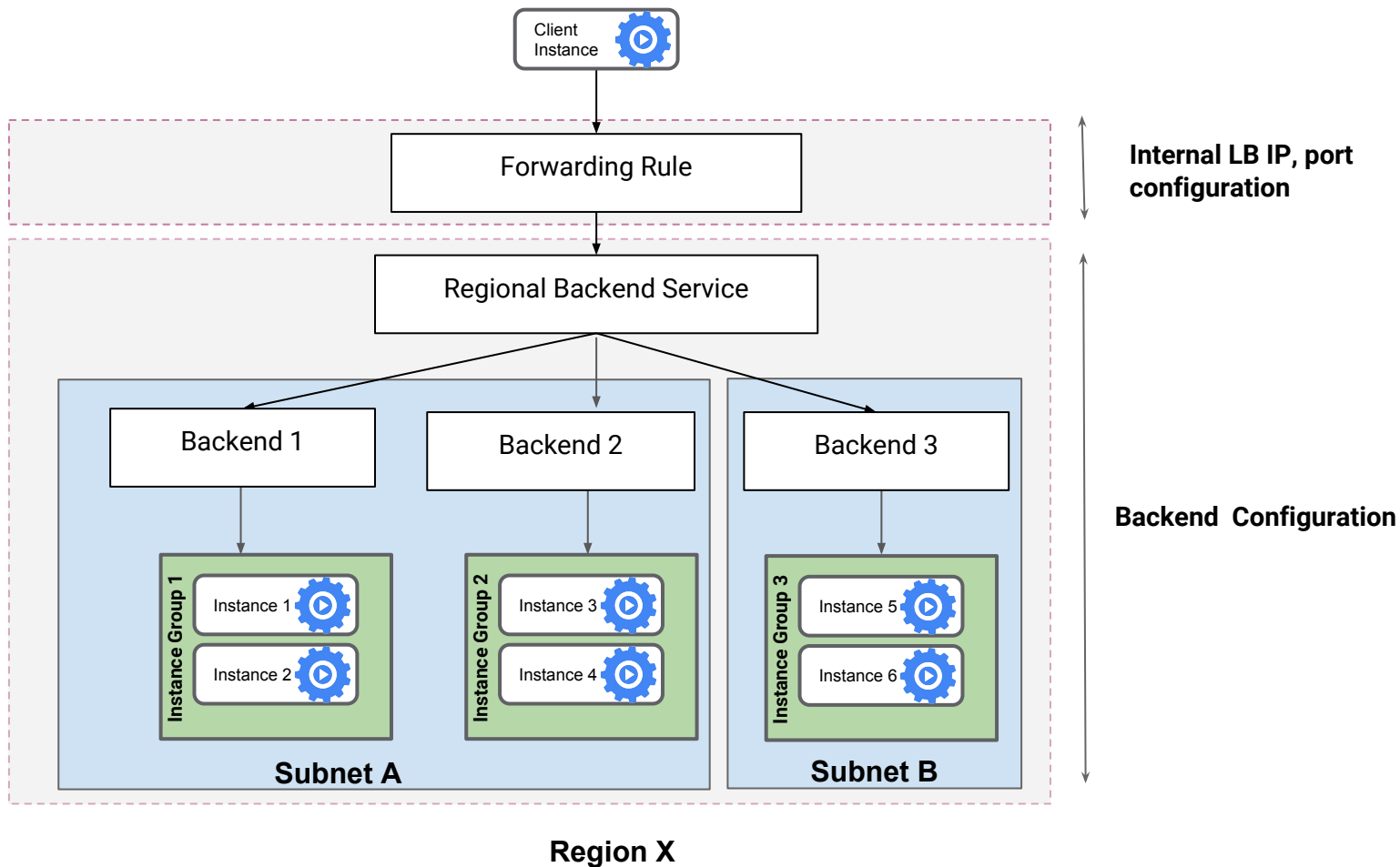
Software-defined Internal Load Balancing on GCP



A server room with rows of server racks and a large window overlooking a city at sunset.

Configuring Internal Load Balancing (console)

Internal Load Balancing Configuration



1. Click “create load balancer” to get started

Click Create

The screenshot shows the Google Cloud Platform console interface. On the left is a sidebar with navigation icons and labels: Home, API, Networks, External IP addresses, Firewall rules, Routes, Load balancing (highlighted in blue), Cloud DNS, VPN, and Cloud Routers. The main content area is titled 'Load balancing' and includes a '+ CREATE LOAD BALANCER' button and a 'REFRESH' button. Below this is a table with two columns: 'Load balancer' and 'Protocol'. The table contains two entries: 'abc' with protocol 'HTTP(S)' and 'lb-pool' with protocol 'TCP'. Each entry has a green checkmark icon on the left and a trash icon with an 'Edit' button on the right. Below the table, there is a note: 'To edit load balancing resources like forwarding rules and target proxies, go to the [advanced menu](#).'

Load balancer	Protocol
✓ abc	HTTP(S)
✓ lb-pool	TCP


To edit load balancing resources like forwarding rules and target proxies, go to the [advanced menu](#).

2. Select the type of Internal Load Balancer- either TCP or UDP


The screenshot displays the Google Cloud Platform console's 'Load balancing' page. The left-hand navigation pane is expanded, showing the 'Networking' section with various sub-items. The 'Load balancing' item is highlighted. The main content area is divided into three columns, each representing a different type of load balancer:

- HTTP(S) Load Balancing**: Layer 7 load balancing for HTTP and HTTPS applications. Configuration options include HTTP LB and HTTPS LB. A 'Start configuration' button is present.
- TCP Load Balancing**: Layer 4 load balancing or proxy for applications that rely on TCP/SSL protocol. Configuration options include TCP LB and SSL Proxy. A 'Start configuration' button is present. A yellow callout box points to this button with the text 'Click here for TCP ILB'.
- UDP Load Balancing**: Layer 4 load balancing for applications that rely on UDP protocol. Configuration options include UDP LB. A 'Start configuration' button is present.


3. Specify you want to load balance traffic between your instances (internal)




Networking




API




Networks




External IP addresses




Firewall rules




Routes




Load balancing




Cloud DNS



VPN



Cloud Routers

 Load balancing

Please answer a few questions to help us select the right load balancing type for your application.

Internet facing or internal only

Do you want to load balance traffic from the Internet to your VMs or only between VMs in your network?

☐ From Internet to my VMs

☒ Only between my VMs

Connection termination

Do you want to offload SSL processing to the Load Balancer?

☐ Yes (SSL Proxy)

☒ No (TCP)

[Continue](#)

Specify you want internal load balancing

4. Configure backends in your region of choice

Networking

- API
- Networks
- External IP addresses
- Firewall rules
- Routes
- Load balancing
- Cloud DNS
- VPN
- Cloud Routers

New Internal load balancer

Name ?
shopping-service-ilb

Backend configuration
Your backend is configured →

Frontend configuration
You have not configured your frontend yet

Review and finalize
Optional

Create **Cancel**

Backend configuration

Backend service

Name ?
shopping-service-ilb

Region ?
us-central1

Network ?
my-custom-network

Protocol: TCP

Backends

Instance group ?
us-ig2 (us-central1-c)

+ Add backend

Health check ?
shopping-tcp-hck
port: 8080, timeout: 5s, check interval: 5s, unhealthy threshold: 2 attempts

The health check probes to your load balanced instances come from addresses in range 130.211.0.0/22. You need to manually configure firewall rules to allow these connections later. [Learn more](#)

Session affinity ?
None

Remember to open up firewall for ILB health checks

Configure backends

5. Configure your RFC 1918 Internal LB IP (specify or let ILB auto-allocate)

The screenshot displays the Google Cloud Platform console interface for configuring a new internal load balancer. The left sidebar shows the 'Networking' section with 'Load balancing' selected. The main content area is titled 'New Internal load balancer' and shows the 'Frontend configuration' section. The 'Name' field is set to 'shopping-service-ilb'. The 'Backend configuration' is marked as 'Your backend is configured'. The 'Frontend configuration' is marked as 'Your frontend is configured'. The 'Review and finalize' step is marked as 'Optional'. The 'Create' button is visible. The 'Frontend configuration' table shows the following configuration:

Protocol	Subnetwork	IP address	Ports
TCP	my-custom-subnet	Automatic	8080

A yellow callout box points to the '+ Add frontend IP and port' button, with the text: 'Configure Internal Load Balancing IP and port (or list of ports)'.

6. Click “create” and your ILB is ready to distribute traffic!

The screenshot shows the Google Cloud Platform console interface for creating a new internal load balancer. The left sidebar contains navigation links for Networking, Networks, External IP addresses, Firewall rules, Routes, Load balancing (highlighted), Cloud DNS, VPN, and Cloud Routers. The main content area is titled 'New Internal load balancer' and is divided into two sections: 'Backend configuration' and 'Frontend configuration'. The 'Backend configuration' section shows a green checkmark and the text 'Your backend is configured'. The 'Frontend configuration' section shows a green checkmark and the text 'Your frontend is configured', with a right arrow indicating further configuration options. Below these sections is a 'Review and finalize' section with an information icon and the text 'Optional'. At the bottom of the main content area are 'Create' and 'Cancel' buttons. A yellow callout box points to the 'Create' button with the text 'Click Create. Done!'.

Networking

← New Internal load balancer

Frontend configuration

Name ?
shopping-service-ilb

Protocol TCP Subnetwork my-custom-subnet IP address Automatic Ports ? 8080 X

+ Add frontend IP and port

Backend configuration
Your backend is configured

Frontend configuration
Your frontend is configured →

Review and finalize
Optional

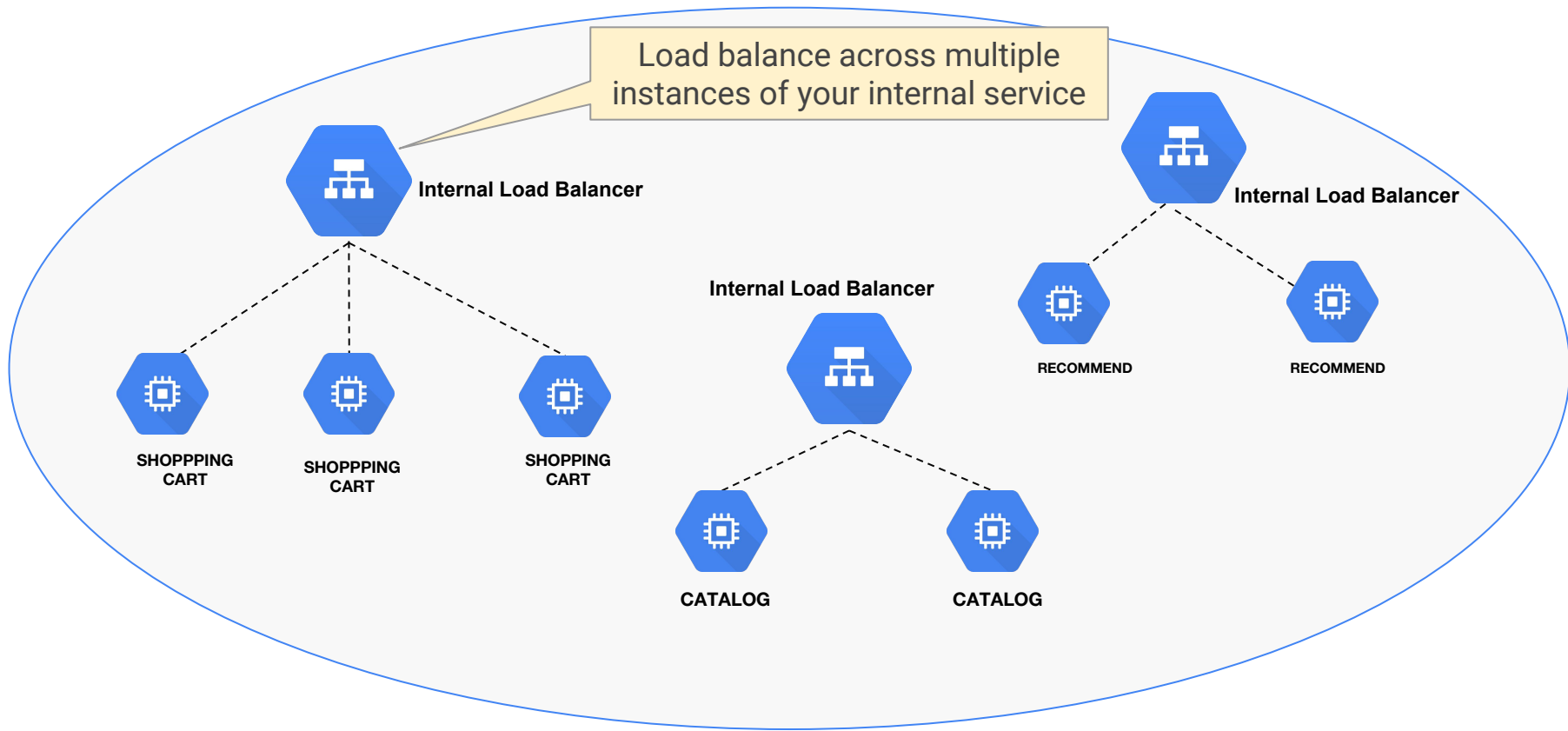
Create Cancel

Click Create.
Done!

The background image shows a modern server room. On the left, there are several tall, black server racks filled with equipment, with some indicator lights glowing. The room has a polished floor that reflects the racks and the light from the window. On the right, a large floor-to-ceiling window provides a view of a city skyline at dusk or dawn, with warm light from the setting or rising sun illuminating the scene. The overall atmosphere is professional and technological.

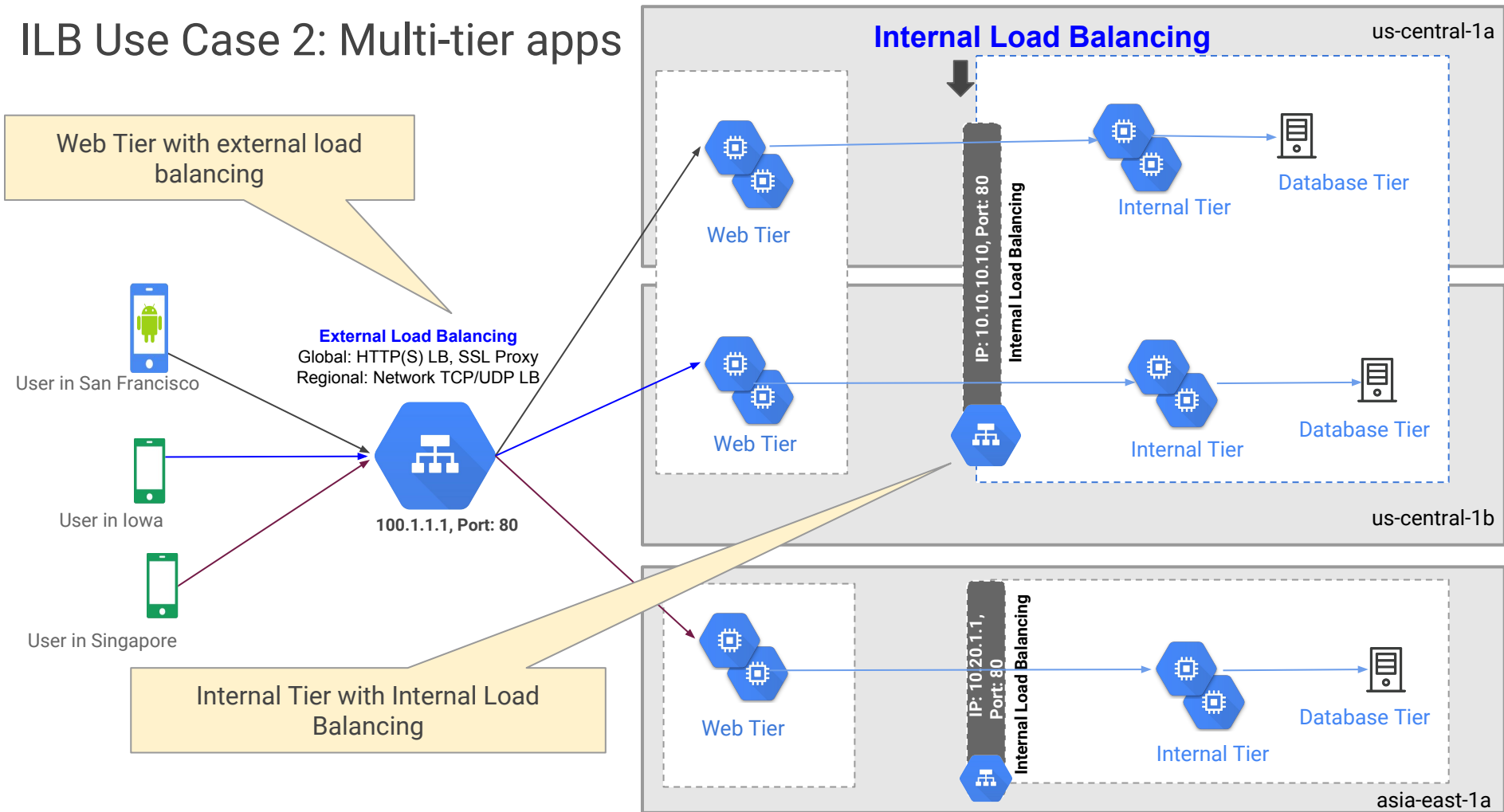
Internal Load Balancing Use Cases

ILB Use Case 1: Scaling and HA for internal (micro)services

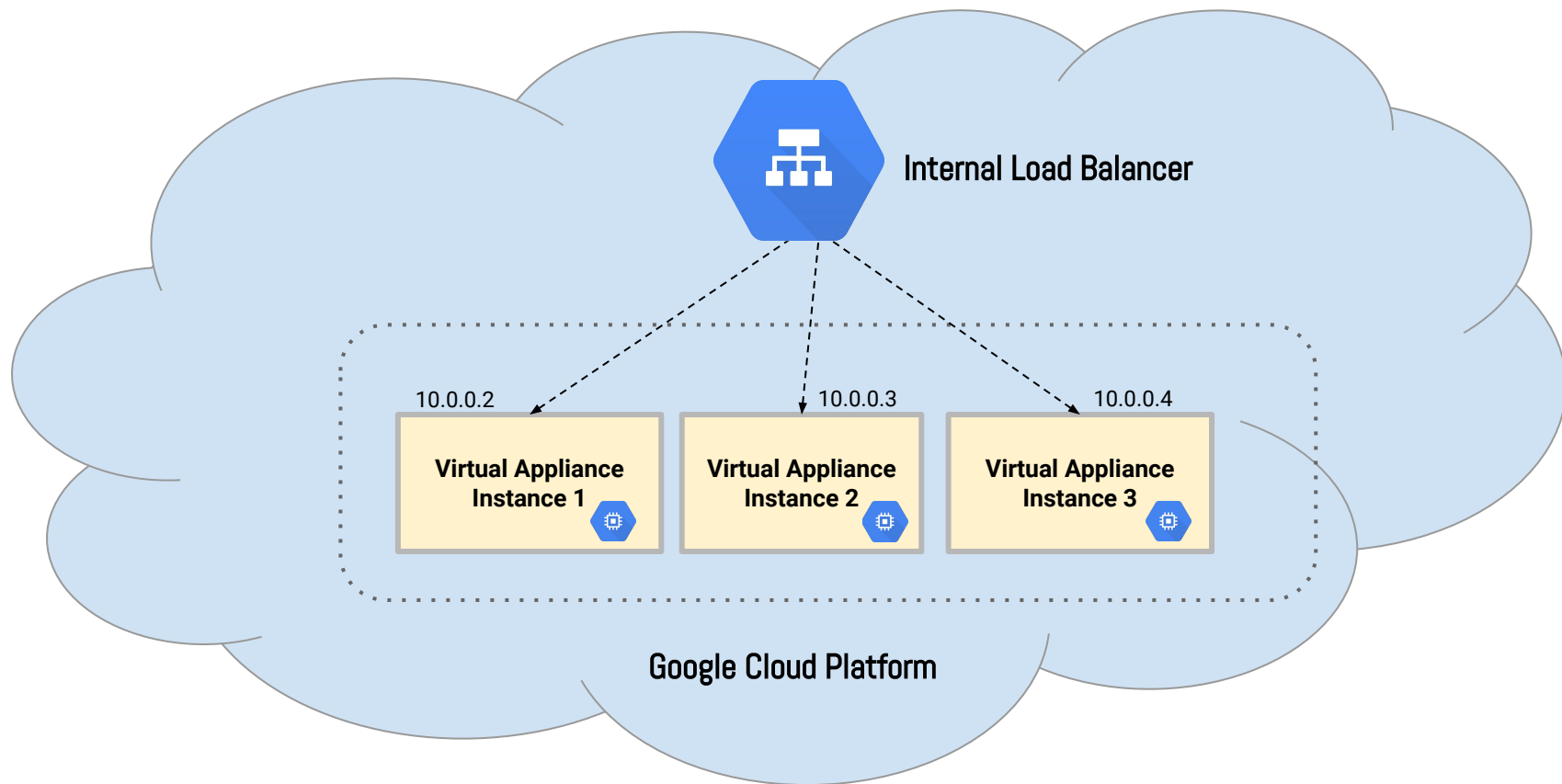


Google Cloud Platform Application composed of ILB-scaled services

ILB Use Case 2: Multi-tier apps



ILB Use Case 3: Scale-out and HA for virtual appliances



Learn more



<https://cloud.google.com/compute/docs/load-balancing/internal/>