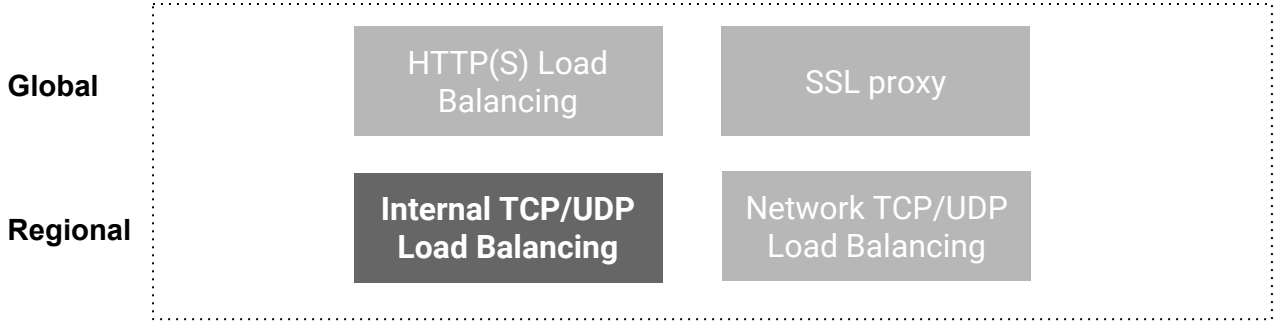
Internal Load Balancing in 5 mins

Deliver scalable and resilient internal-only services on GCP

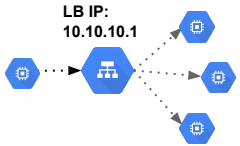
Google Cloud Platform

Google Cloud Load Balancing

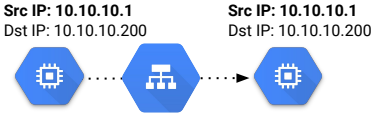


- Request Routing
- Hash-based LB algorithm
- Capacity-based LB algorithm
- Global Load Balancing
- Health Checks
- Connection Draining
- Autoscaling
- Cross-region failover
- Affinity
- Logging
- Monitoring
- Cloud CDN

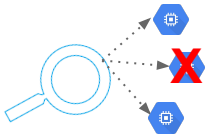
Internal Load Balancing



Internal (RFC 1918) Load Balancing



Client IP preserved



Health checks (TCP, HTTP, HTTPS)

Autoscaling (no prewarming)

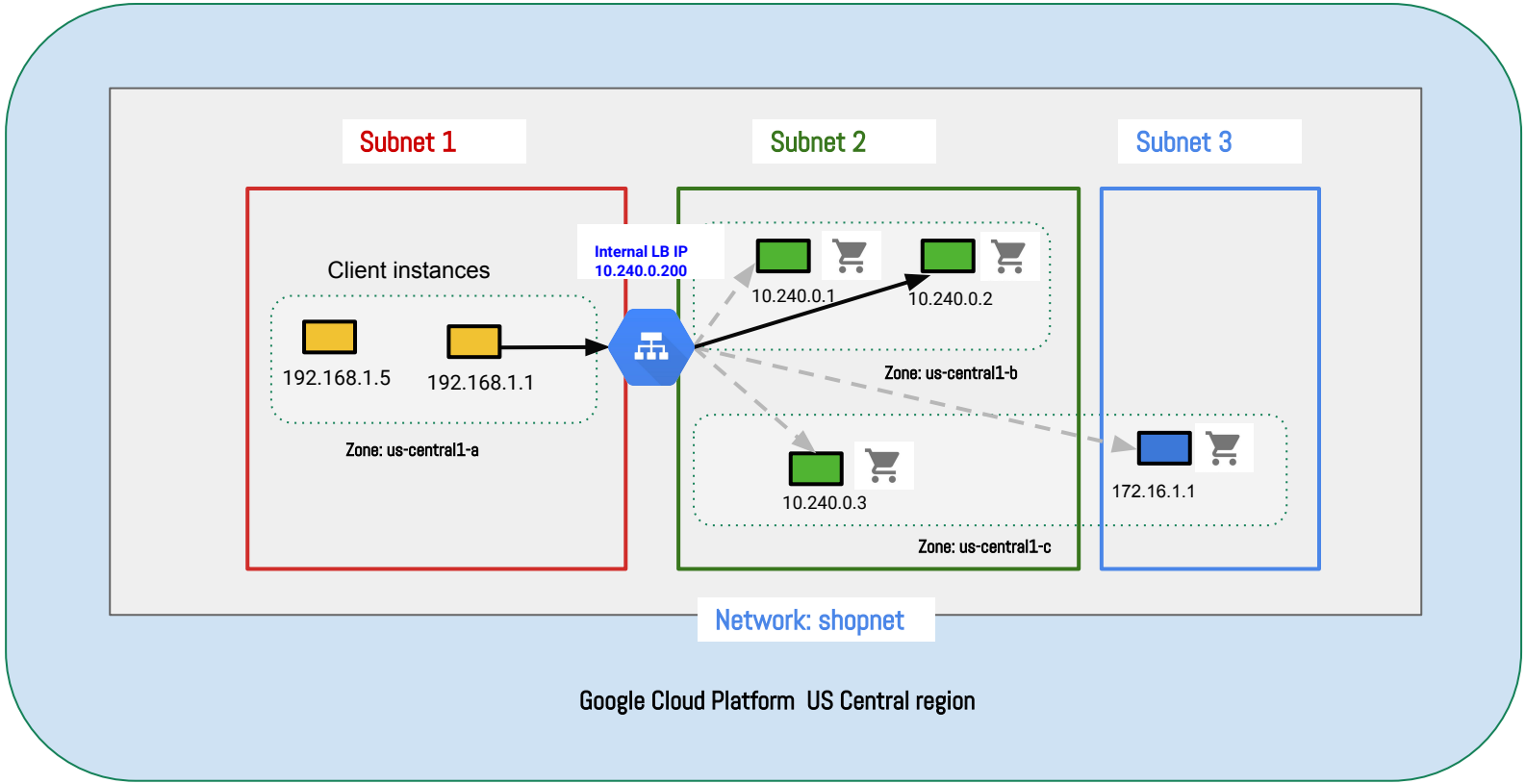
Session Affinity

REST API

gcloud CLI

Console

Internal Load Balancing Example



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

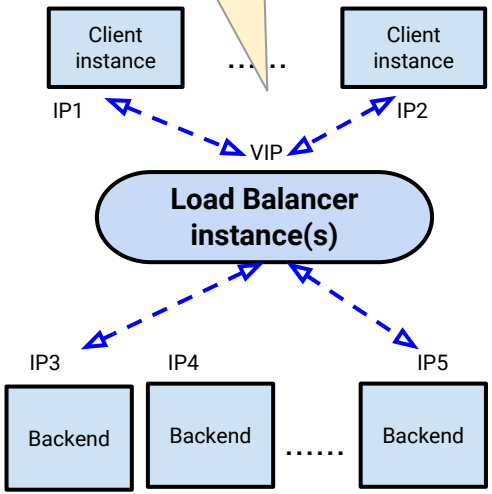
Backends running shopping cart application

Under the hood: Software-defined load balancing

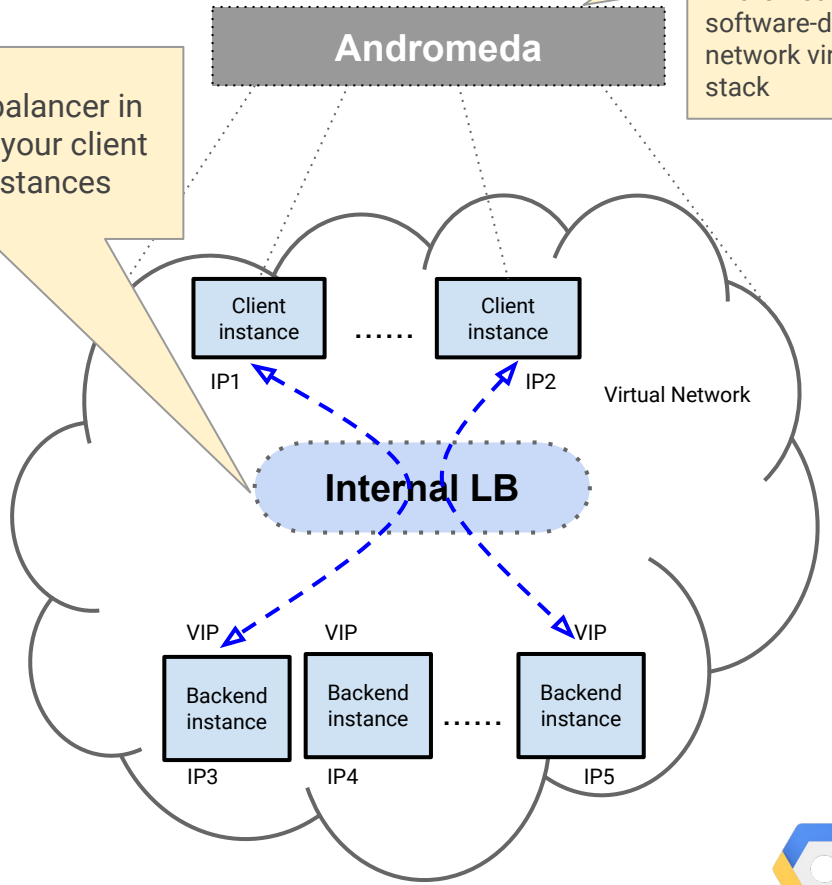
Load Balancers can become choke points

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

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



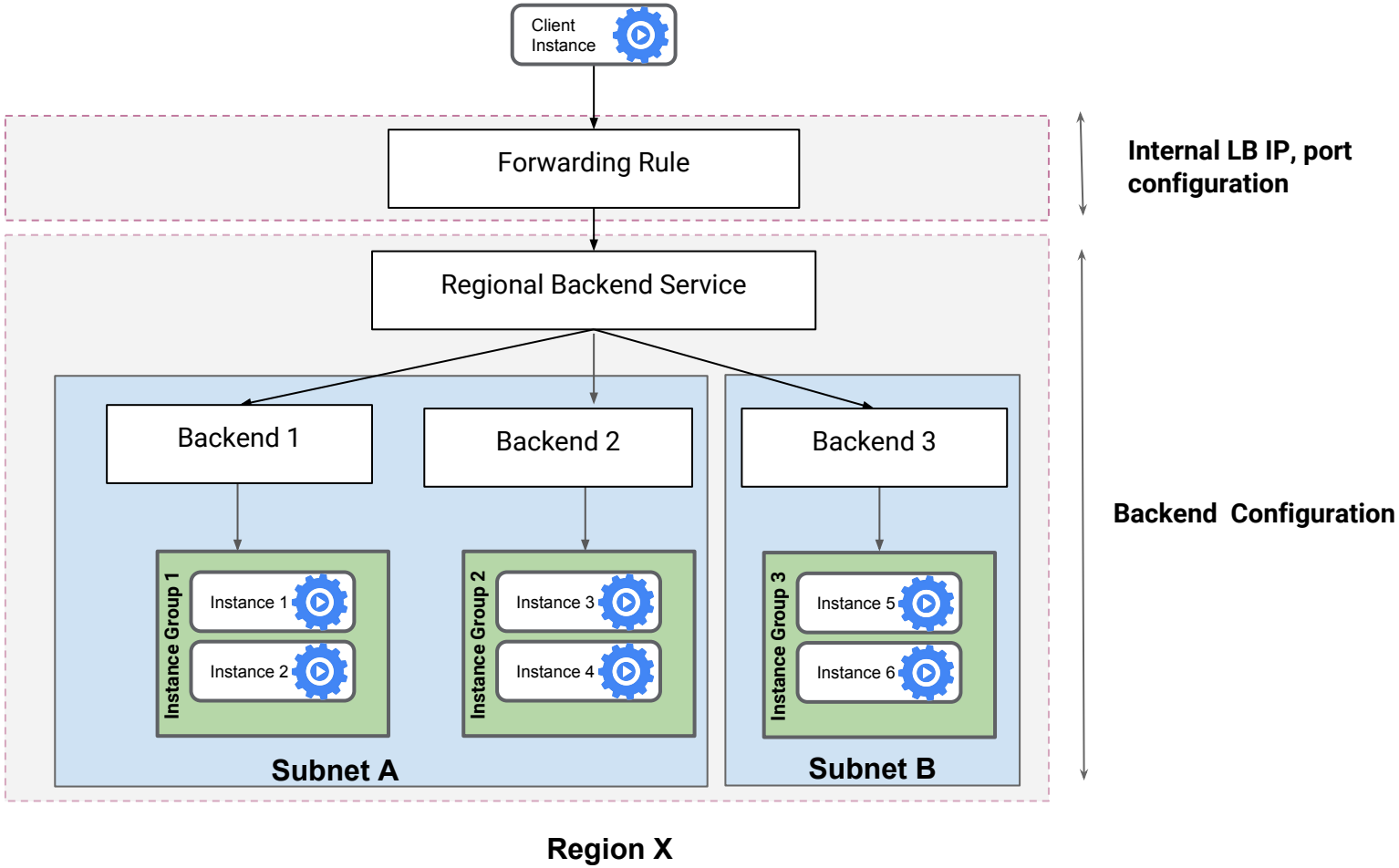
Typical Instance-based Internal Load Balancing



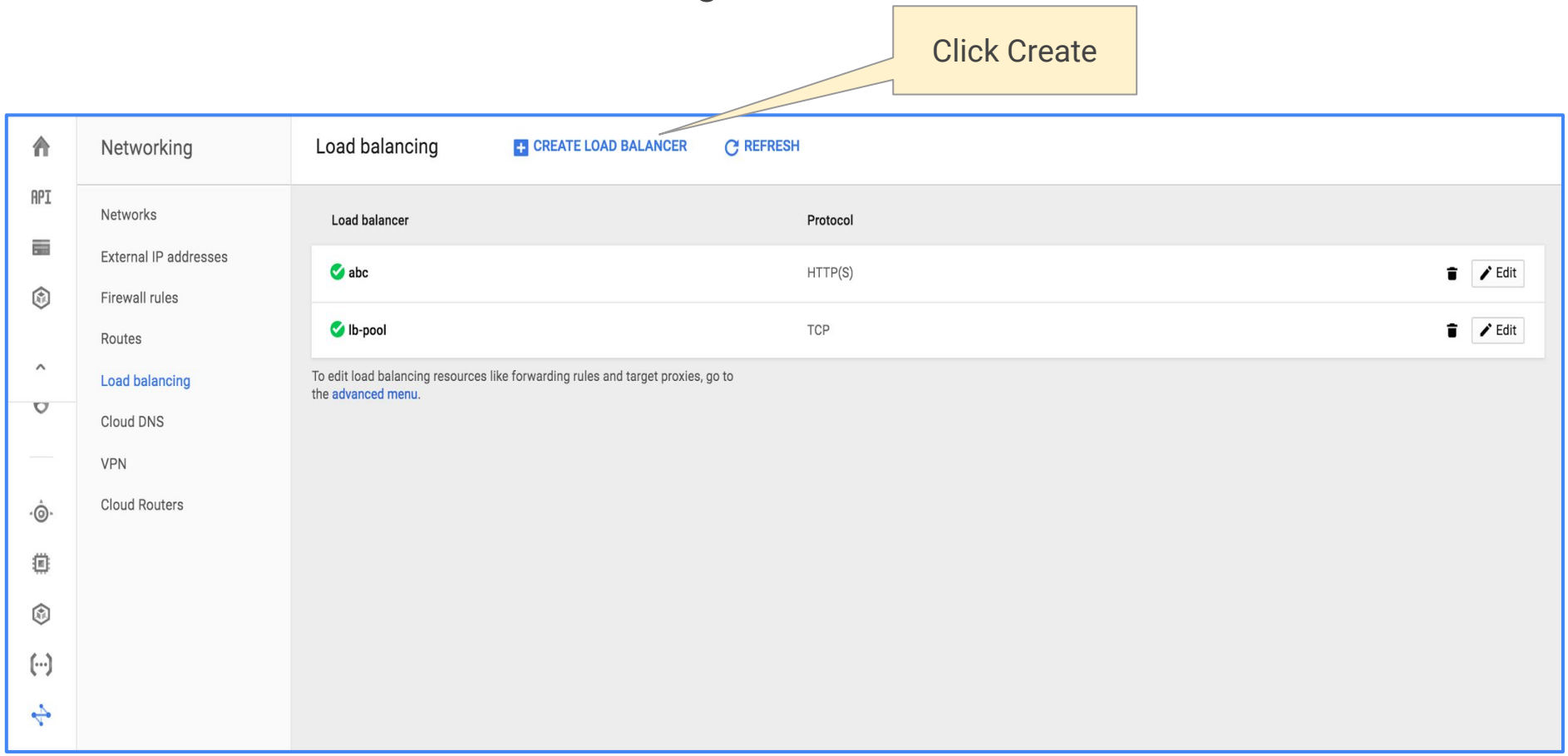
Software-defined Internal Load Balancing on GCP

Configuring Internal Load Balancing (console)

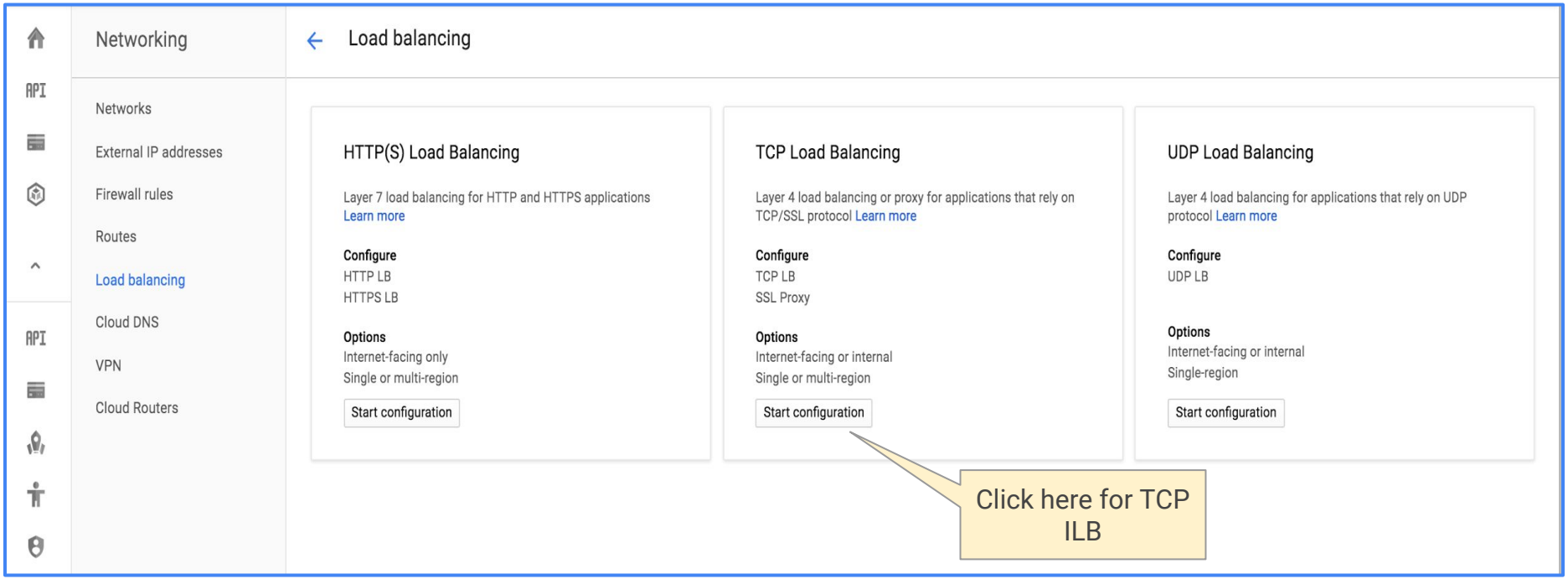
Internal Load Balancing Configuration



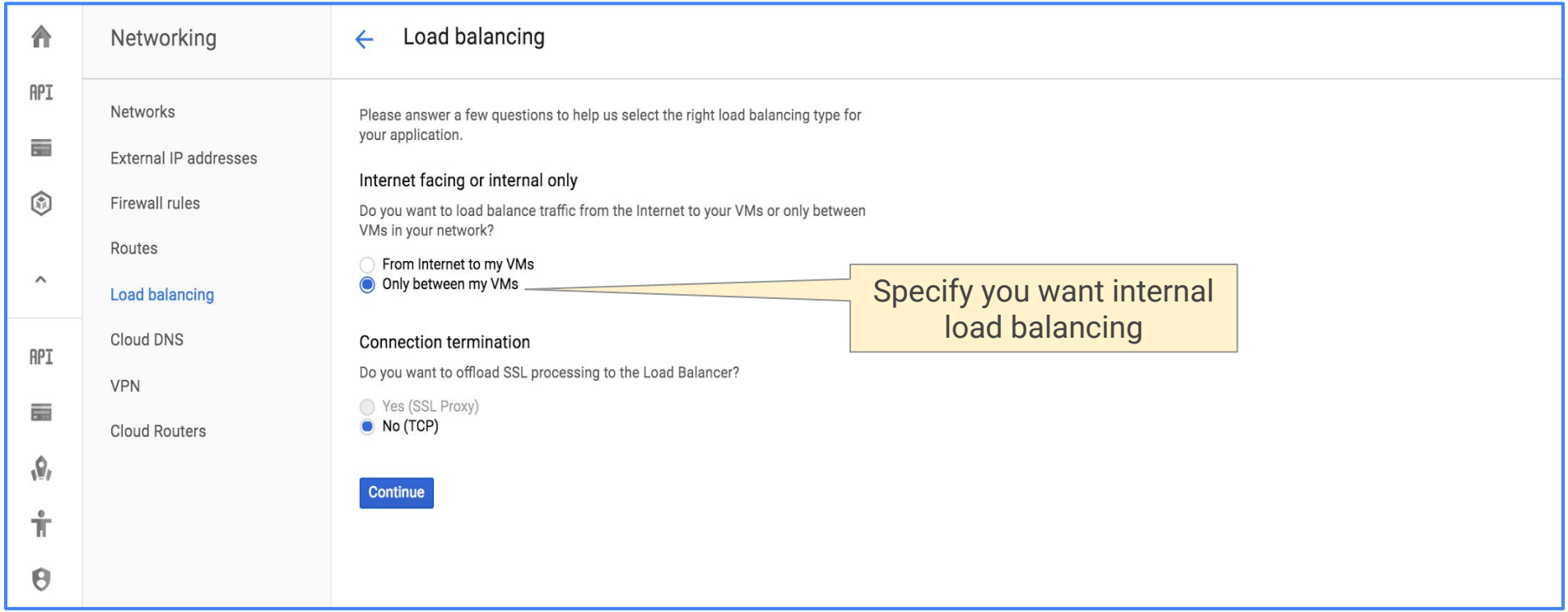
1. Click “create load balancer” to get started



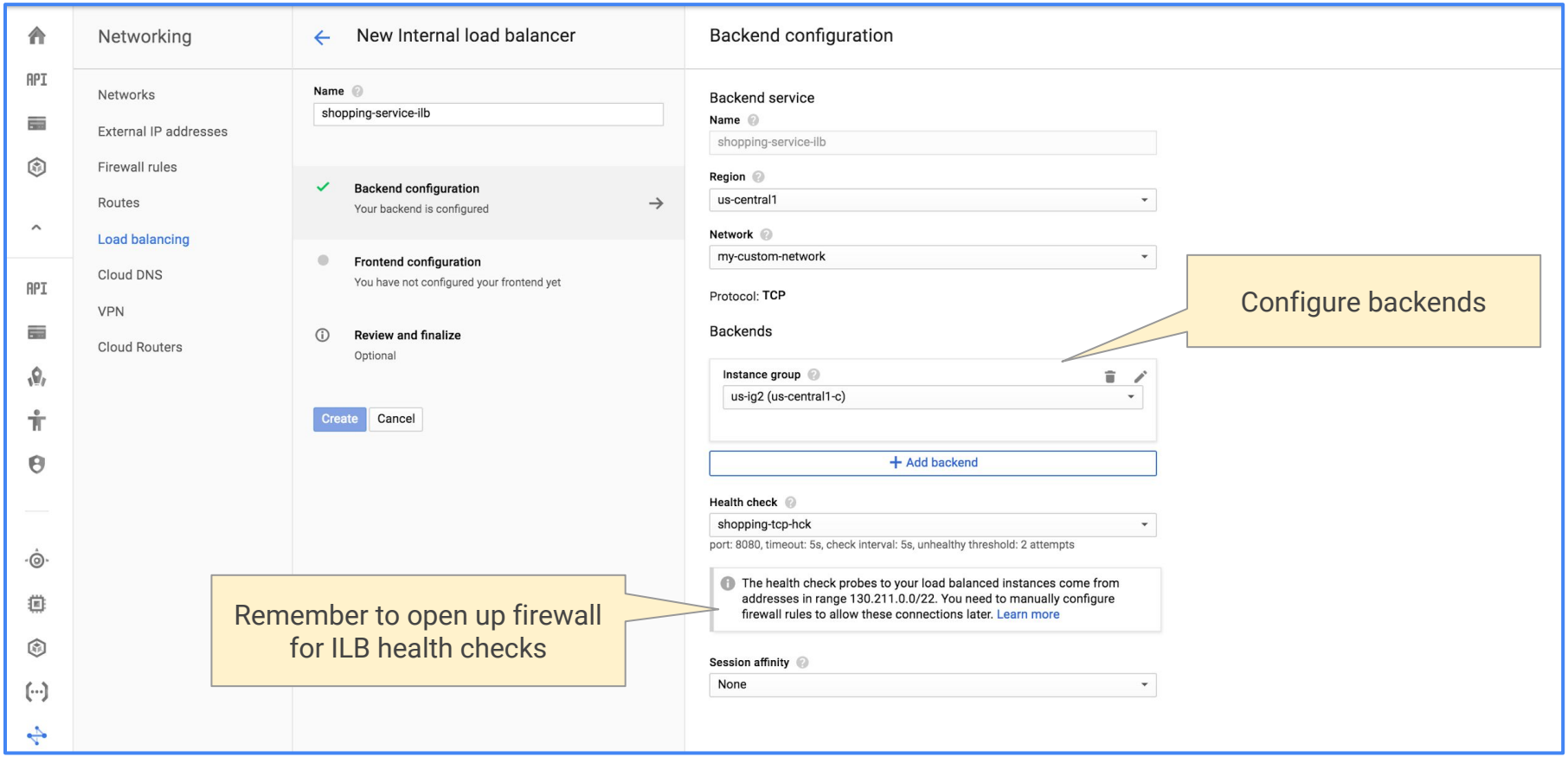
2. Select the type of Internal Load Balancer- either TCP or UDP



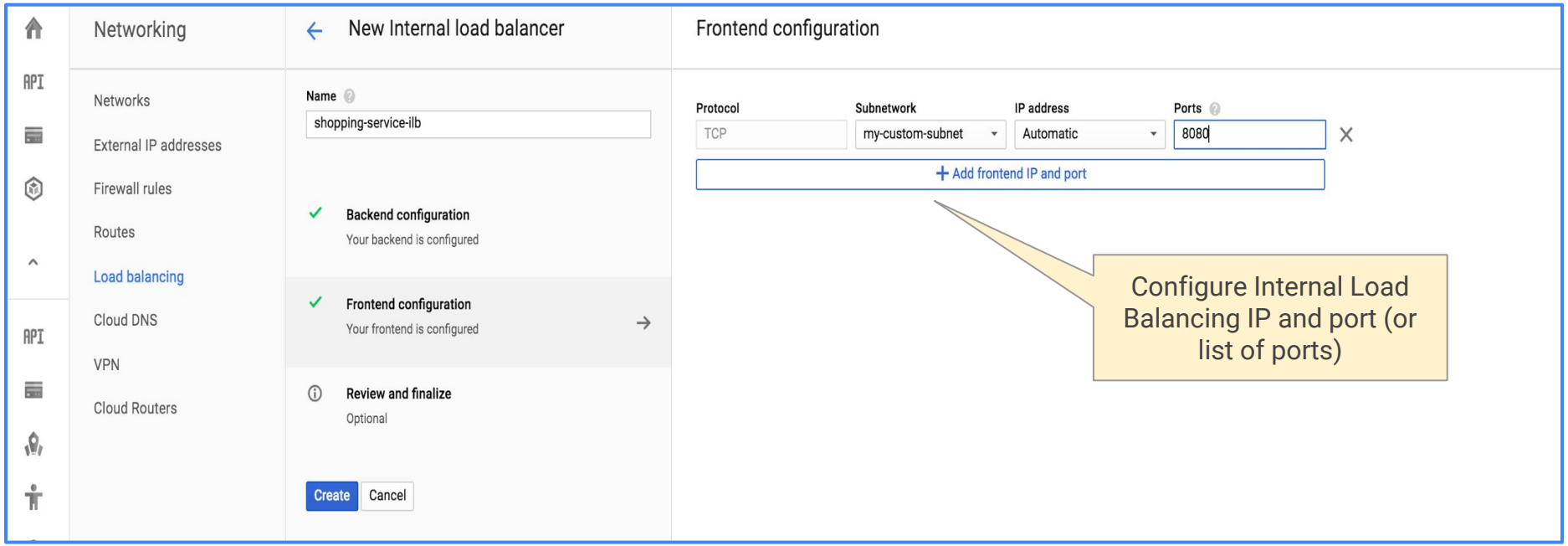
3. Specify you want to load balance traffic between your instances (internal)



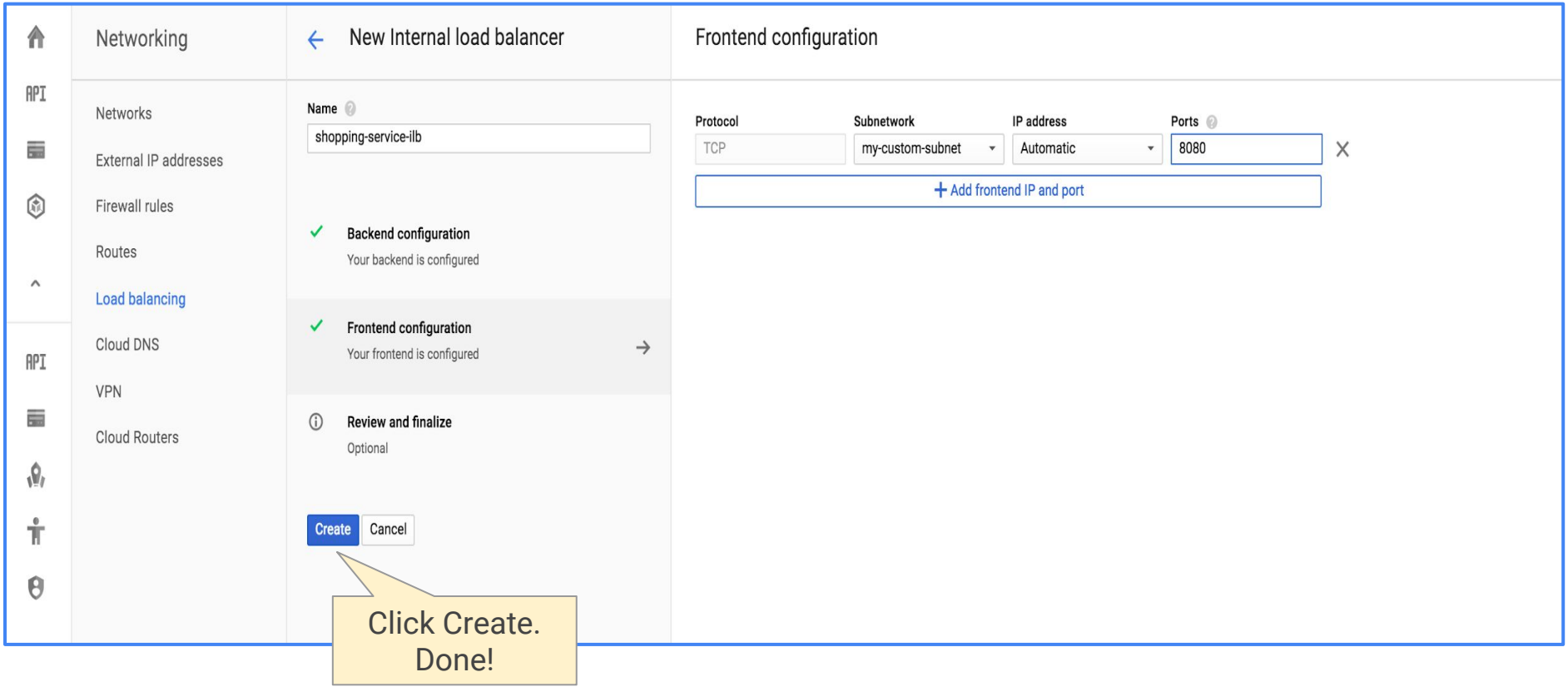
4. Configure backends in your region of choice



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

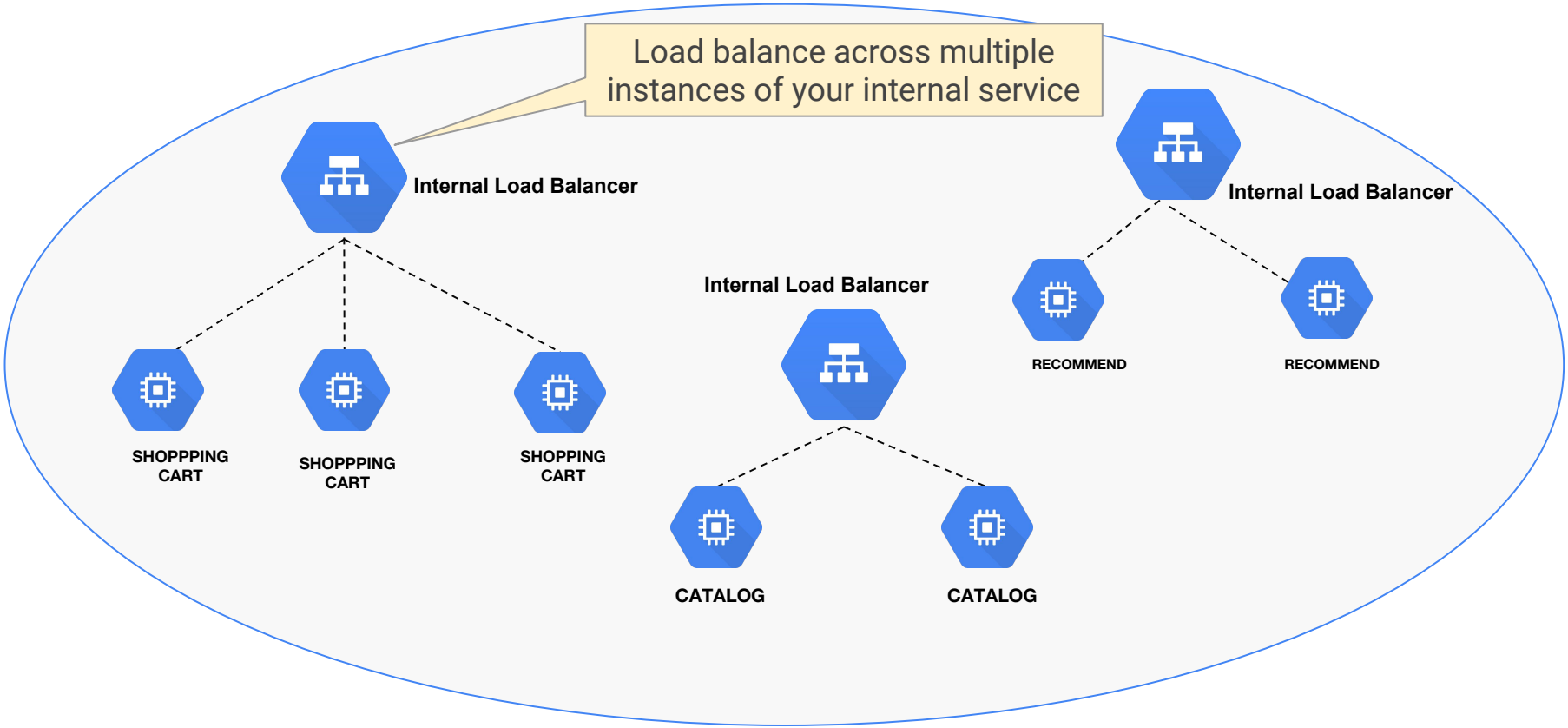


6. Click “create” and your ILB is ready to distribute traffic!



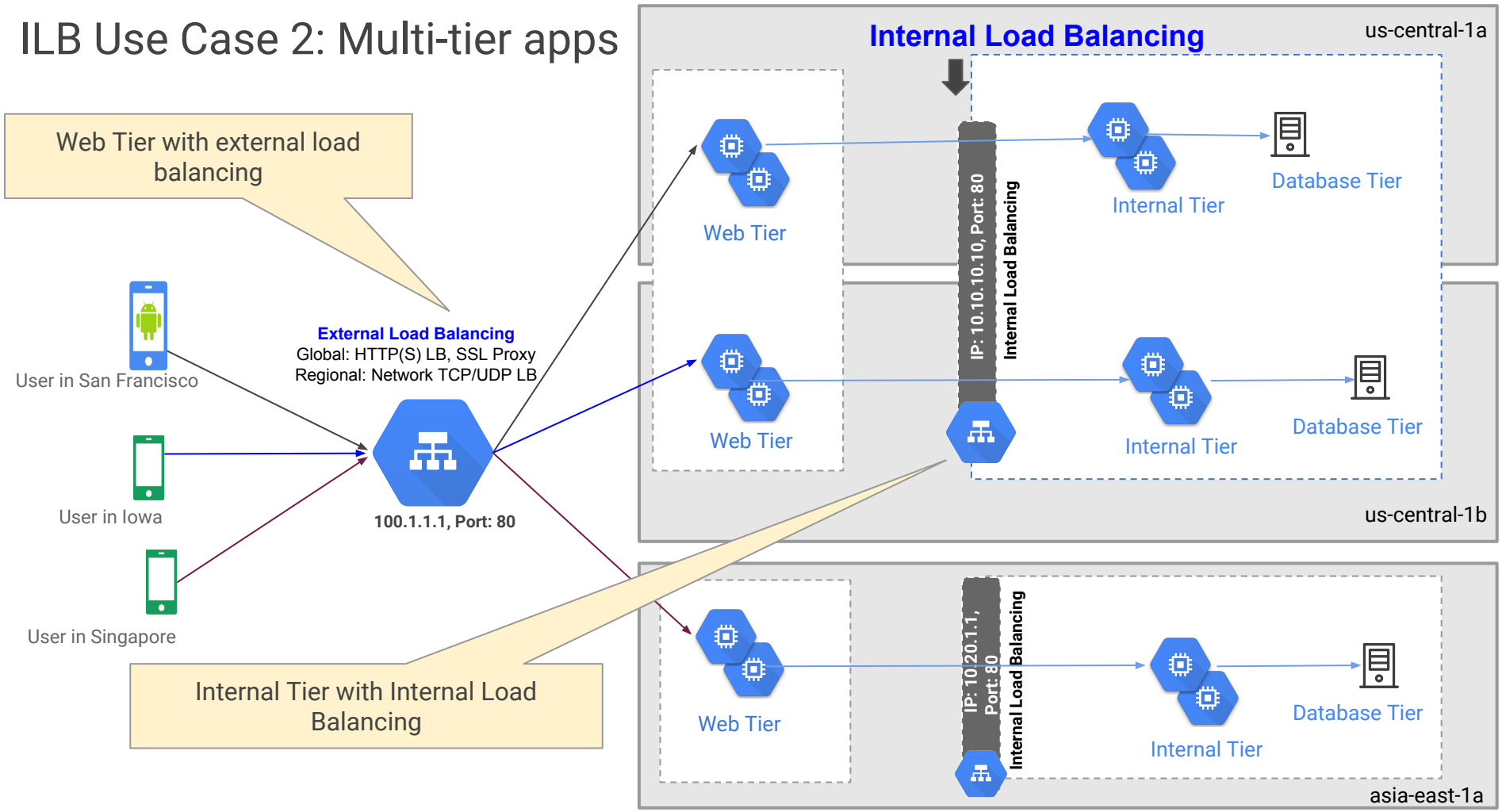

Internal Load Balancing Use Cases

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

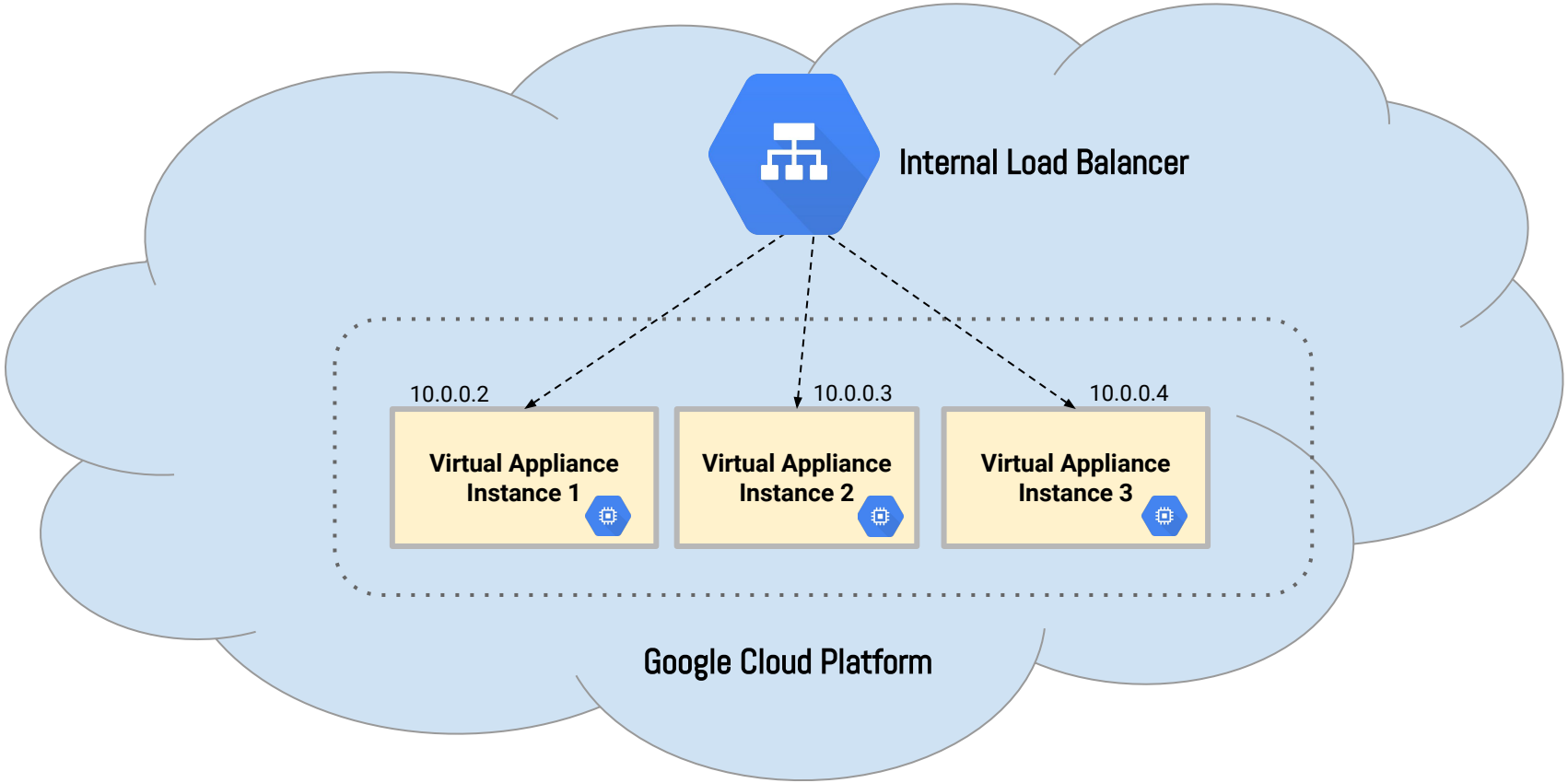


Google Cloud Platform Application composed of ILB-scaled services

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/