

# Load Balancing

Level 100

Rohit Rahi  
March 2018

## Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

# Objectives

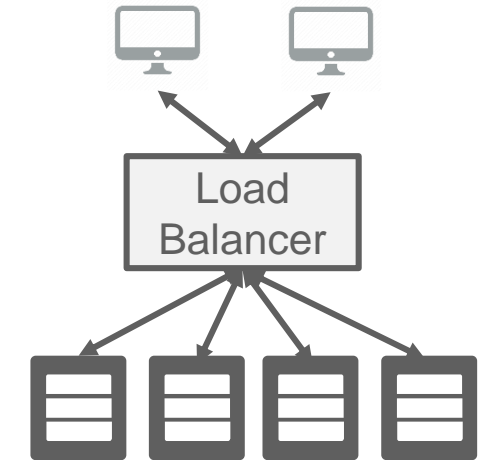
After completing this lesson, you should be able to:

- Describe Oracle Cloud Infrastructure Load Balancing Service concepts
- Create and test a Public Load Balancer

# Load Balancing Primer

A load balancer sits between the clients and the backends performs tasks such as:

- Service Discovery: What backends are available in the system? How should the load balancer talk to them?
- Health Check: What backends are currently healthy and available to accept requests?
- Load Balancing: What algorithm should be used to balance individual requests across the healthy backends?



## Load Balancer benefits

- Fault tolerance and HA: using health check + LB algorithms, a LB can effectively route around a bad or overloaded backend
- Scale: LB maximizes throughput, minimizes response time, and avoids overload of any single resource
- Naming abstraction: name resolution can be delegated to the LB; backends don't need public IP addresses

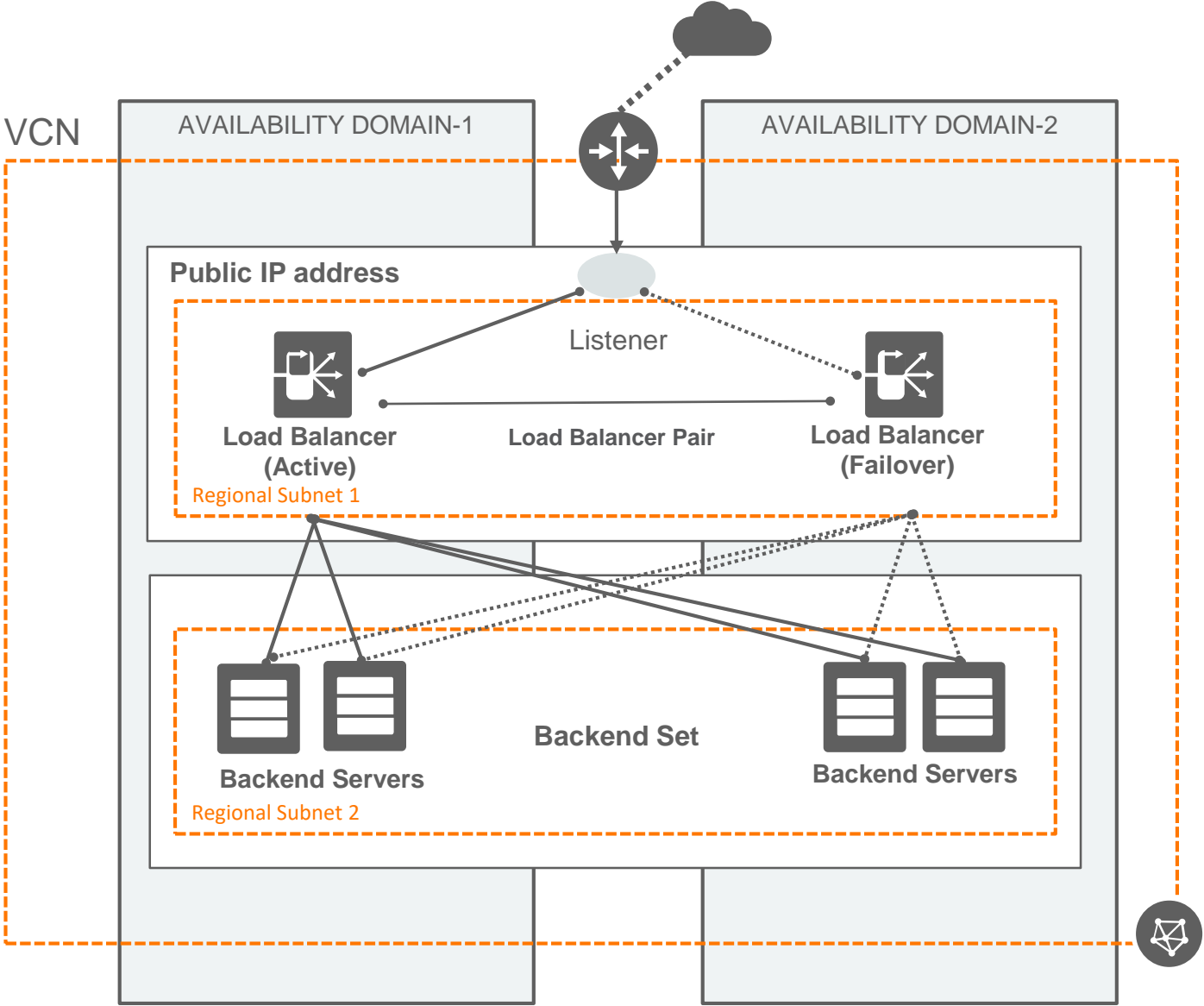
# OCI Load Balancing Service

- Load Balancer as-a-service, provides scale and HA
- Public and Private Load Balancer options
- Public Load Balancer service is regional in scope and provides High Availability across two ADs
- Supported Protocols – TCP, HTTP/1.0, HTTP/1.1, HTTP/2, WebSocket
- Supports SSL Termination, End-to-End SSL, SSL Tunneling
- Supports advanced features such as session persistence and content based routing
- Key differentiators
  - Private or Public Load Balancer (with Public IP address)
  - Provisioned bandwidth – 100 Mbps, 400 Mbps, 8 Gbps
  - Single load balancer for TCP (layer 4) and HTTP (layer 7) traffic

# Public Load Balancer

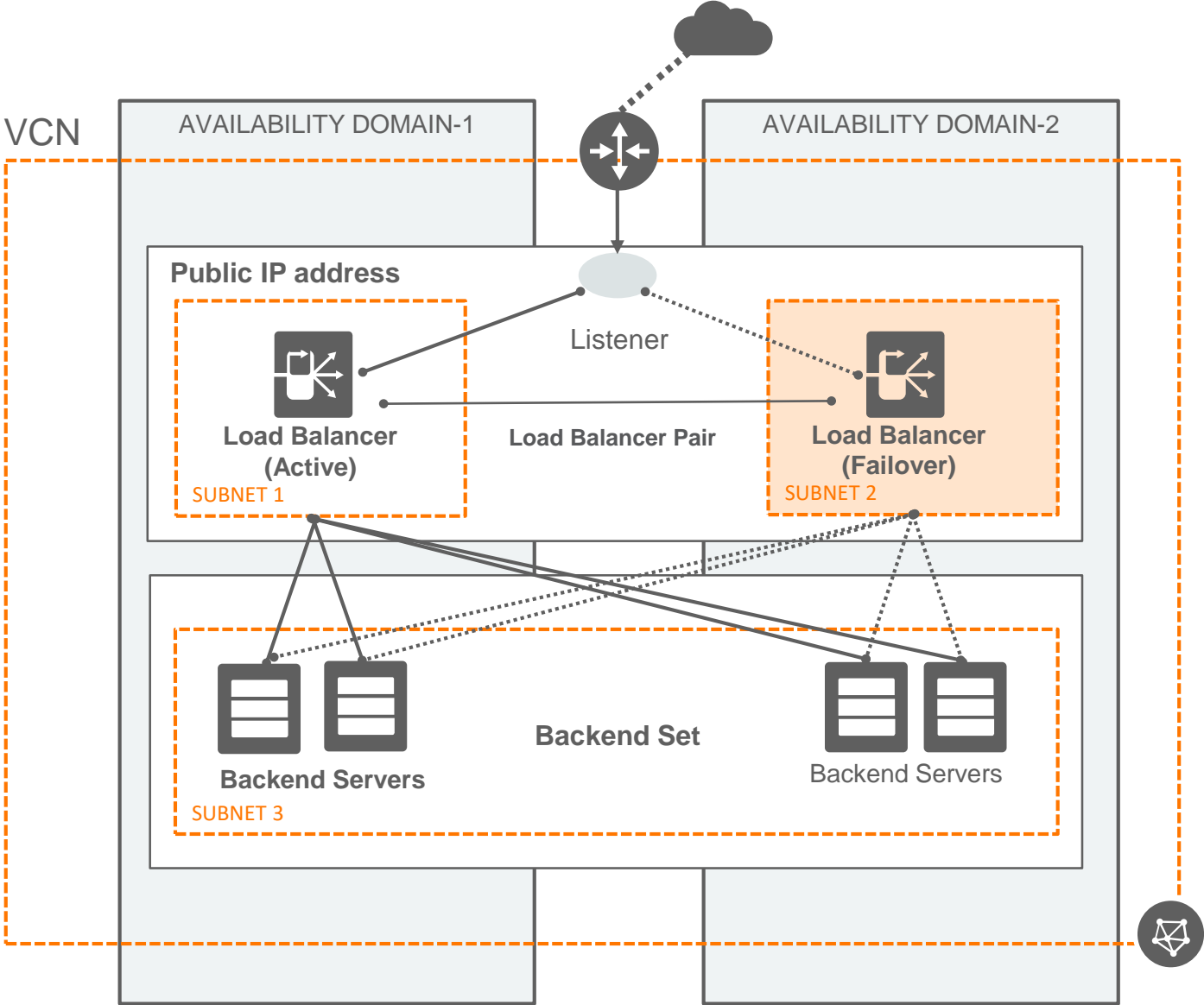
- Accepts traffic from the internet using a public IP address that serves as the entry point for incoming traffic
- Public Load Balancer is a regional service.
- If your region includes multiple availability domains, a public load balancer requires either a regional subnet (recommended) or two availability domain-specific (AD-specific) subnets, each in a separate availability domain.
- Load Balancing service creates a primary load balancer and a standby load balancer, each in a different availability domain
- Supports AD failover in the event of an AD outage in an Oracle Cloud Infrastructure multi-AD region
- Floating Public IP is attached to the primary load balancer, and in the event of an AD outage Floating Public IP is attached to the standby load balancer
- Service treats the two load balancers as equivalent and you cannot denote one as "primary"

# Public Load Balancer (Using Regional Subnets - recommended)





# Public Load Balancer (Using AD Specific Subnets)

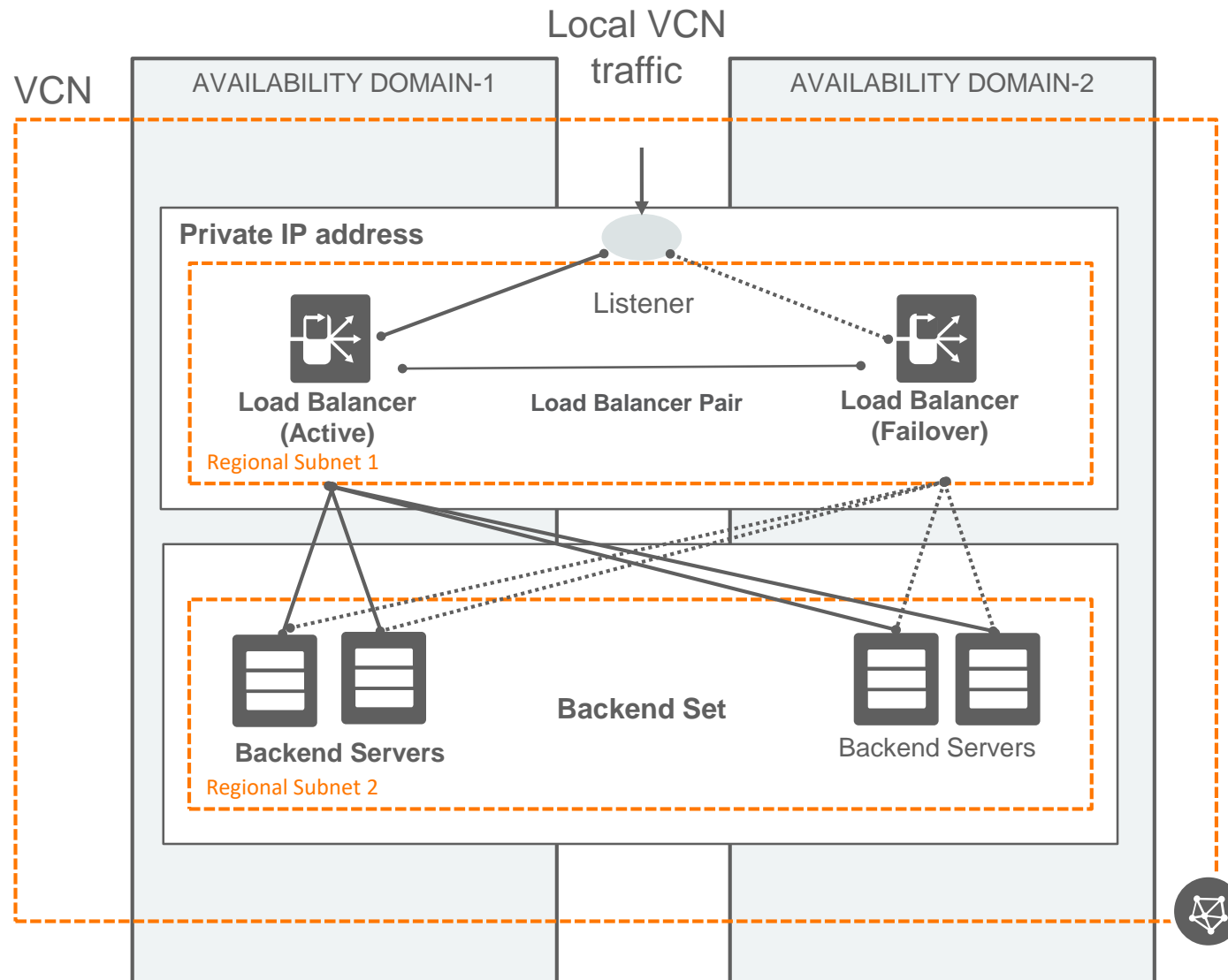




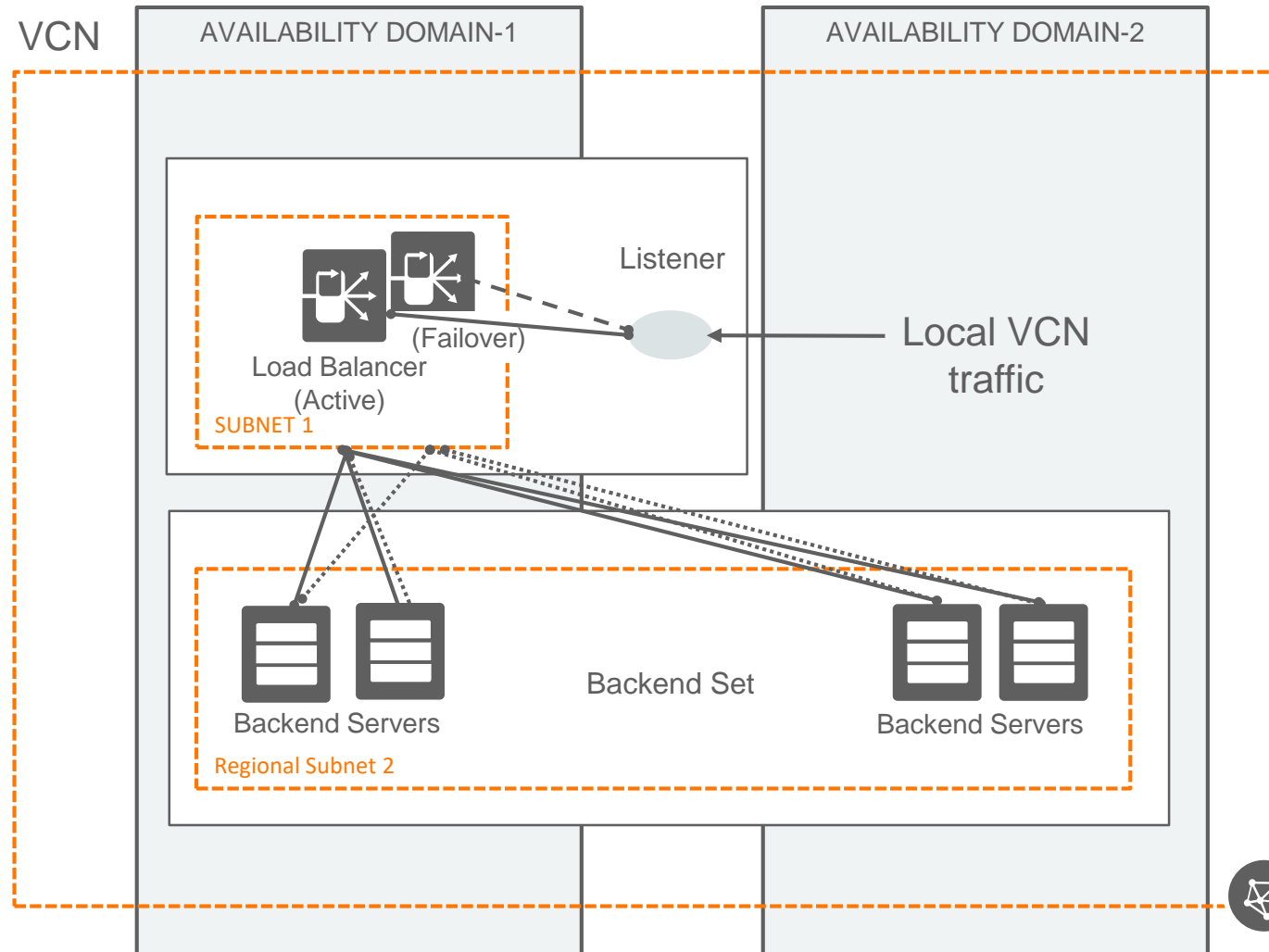
# Private Load Balancer

- Assigned a private IP address from the subnet hosting the LB
- The load balancer can be regional or AD-specific, depending on the scope of the host subnet; highly-available within an AD with AD specific subnets or Highly available with regional subnets
- The primary and standby load balancer each require a private IP address from that subnet
- The load balancer is accessible only from within the VCN that contains the associated subnet, or as further restricted by your security list rules

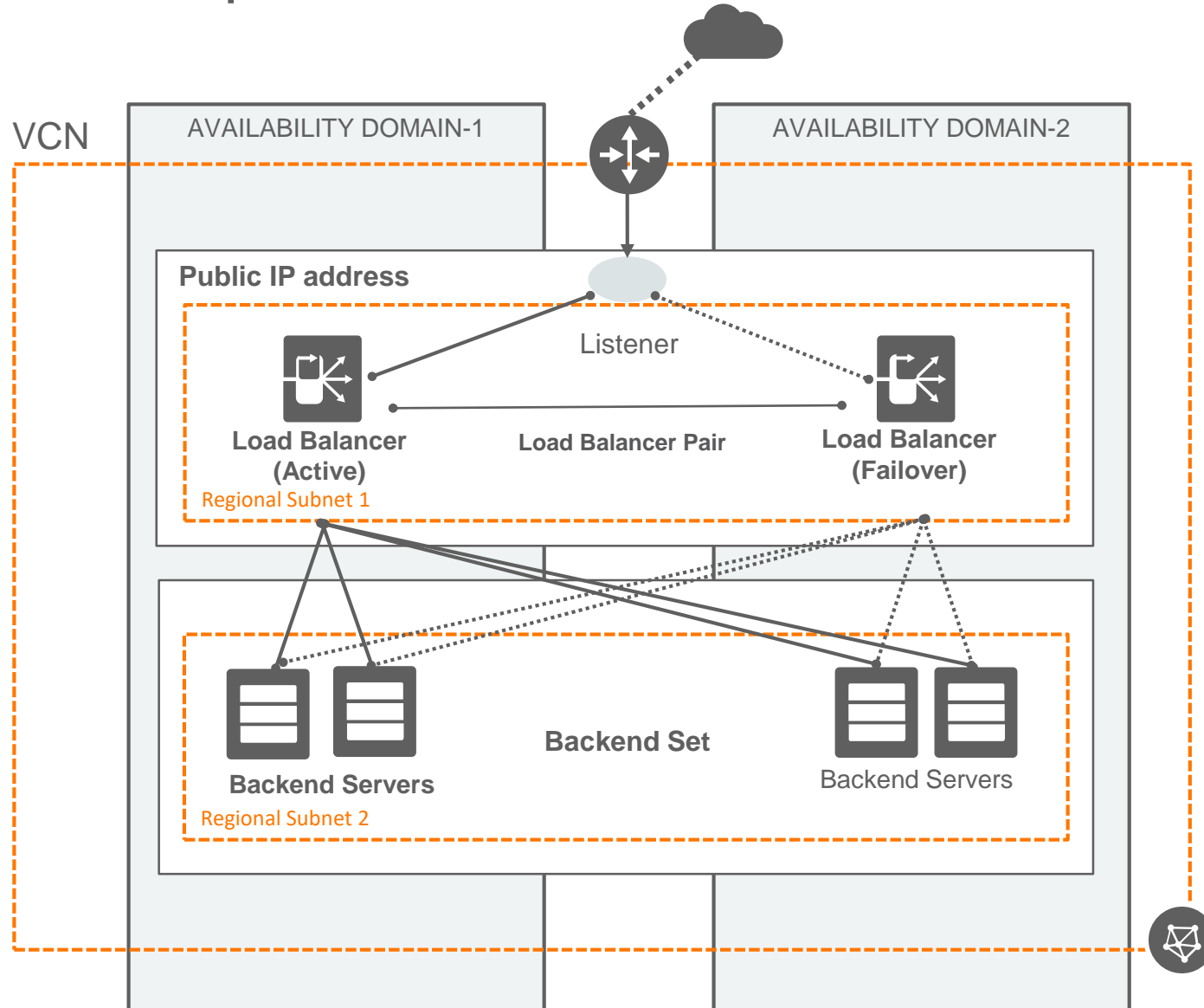
# Private Load Balancer (Using Regional Subnets)



# Private Load Balancer (with AD Specific Subnets)



# Concepts - Public Load Balancer



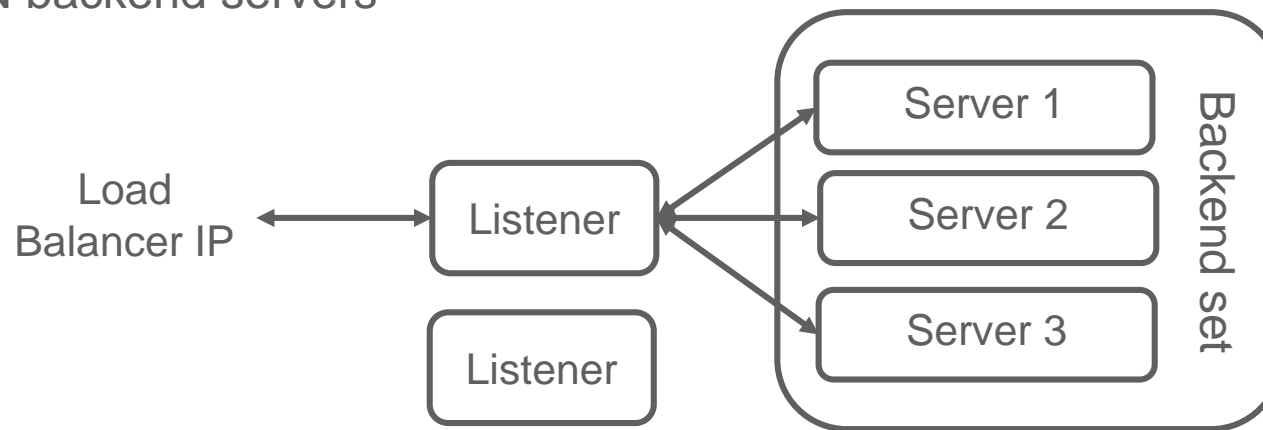
- **Load Balancing Policy** – tells the load balancer how to distribute incoming traffic to the backend servers
  - round-robin
  - IP hash
  - least connection
- **Backend Server** – application server responsible for generating content in reply to the incoming TCP or HTTP traffic
- **Health Checks** – a test to confirm the availability of backend servers; supports
  - TCP-level
  - HTTP-level health checks
- **Backend Set** – logical entity defined by a list of backend servers, a load balancing policy, and a health check policy
- **Listener** – entity that checks for incoming traffic on the load balancer's IP address

# Load Balancing Policies

- **Round Robin:** default policy, distributes incoming traffic sequentially to each server in a backend set. After each server has received a connection, the load balancer repeats the list in the same order.
- **IP Hash:** uses an incoming request's source IP address as a hashing key to route non-sticky traffic to the same backend server
- **Least Connection:** routes incoming non-sticky request traffic to the backend server with the fewest active connections
- Load balancer policy decisions apply differently to TCP load balancer, cookie-based session persistent HTTP requests (sticky requests), and non-sticky HTTP requests
  - A TCP load balancer considers policy and weight criteria
  - An HTTP load balancer w/ cookie-based session persistence forwards requests using cookie's session info
  - For non-sticky HTTP requests, the load balancer applies policy and weight criteria

# Health Check

- Health check is a test to confirm the availability of backend servers. Health Check is activated for
  - Backends
  - Backend set
  - Overall Load Balancer
- A load balancer IP can have up to 16 listeners (port numbers). Each listener has a backend set that can have 1 to N backend servers

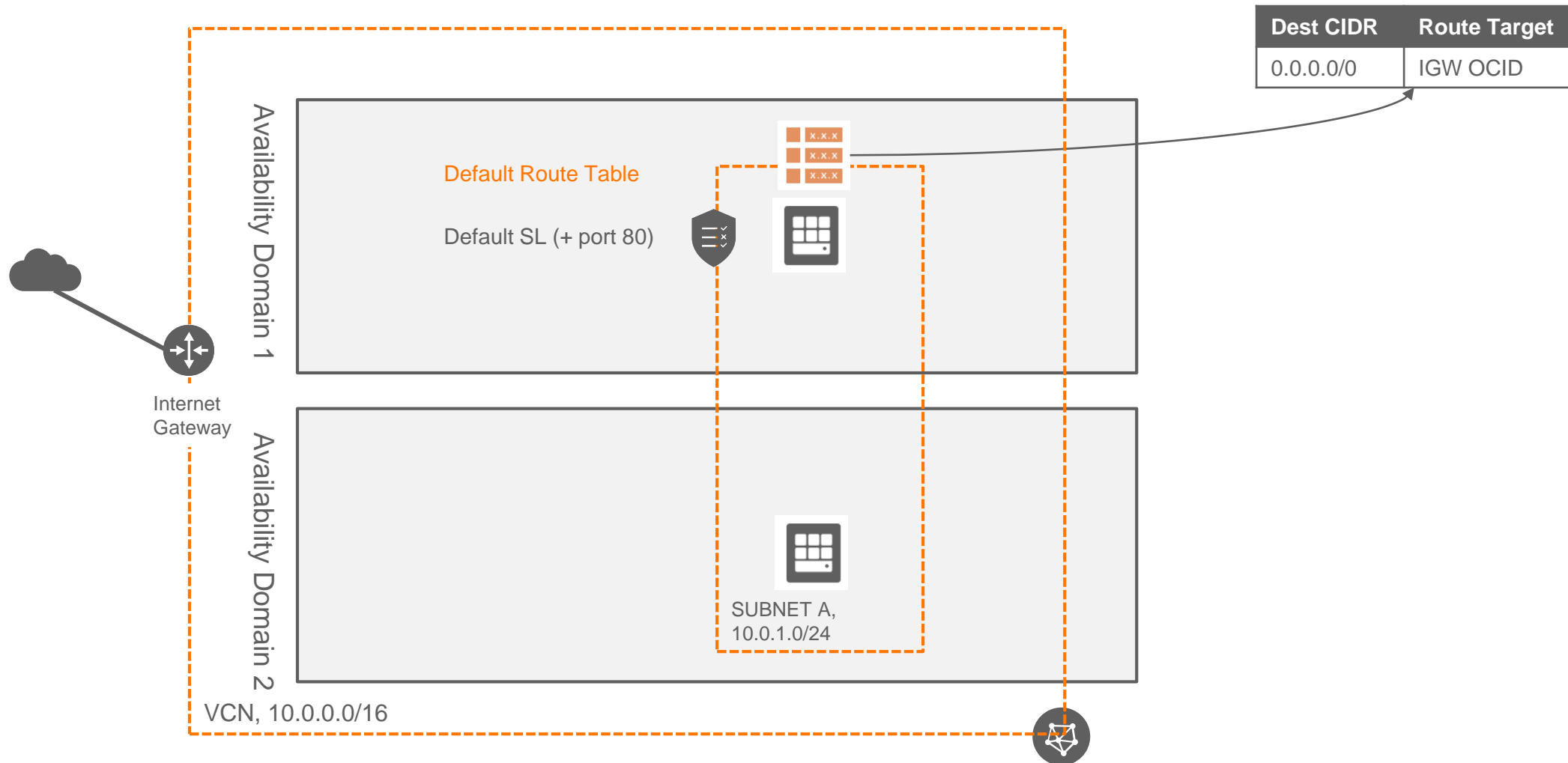


- Health API provides a 4-state health status (ok, warning, critical, unknown)
- Health status is updated every three minutes. No finer granularity is available

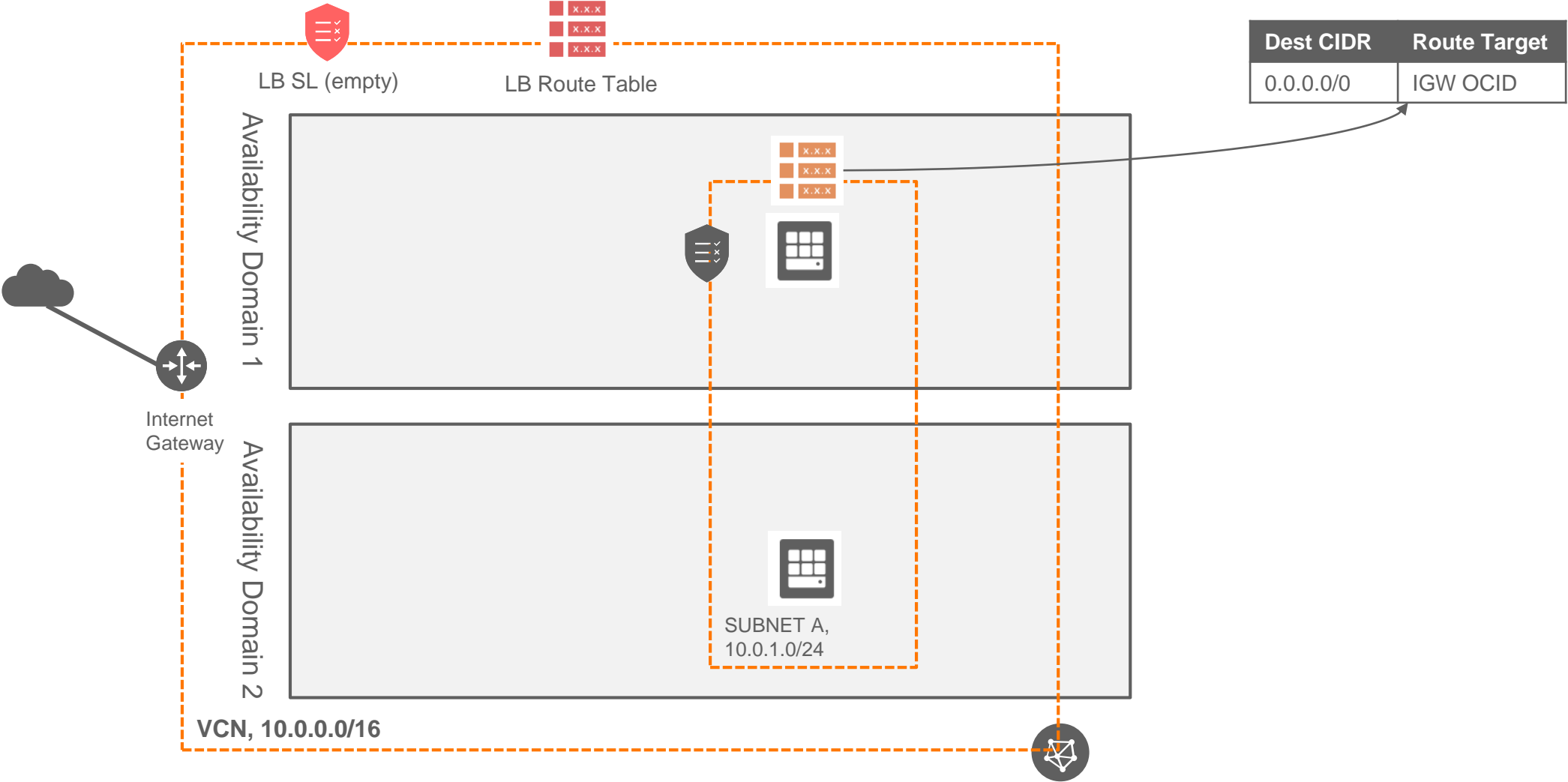
# Steps to create a Public Load Balancer



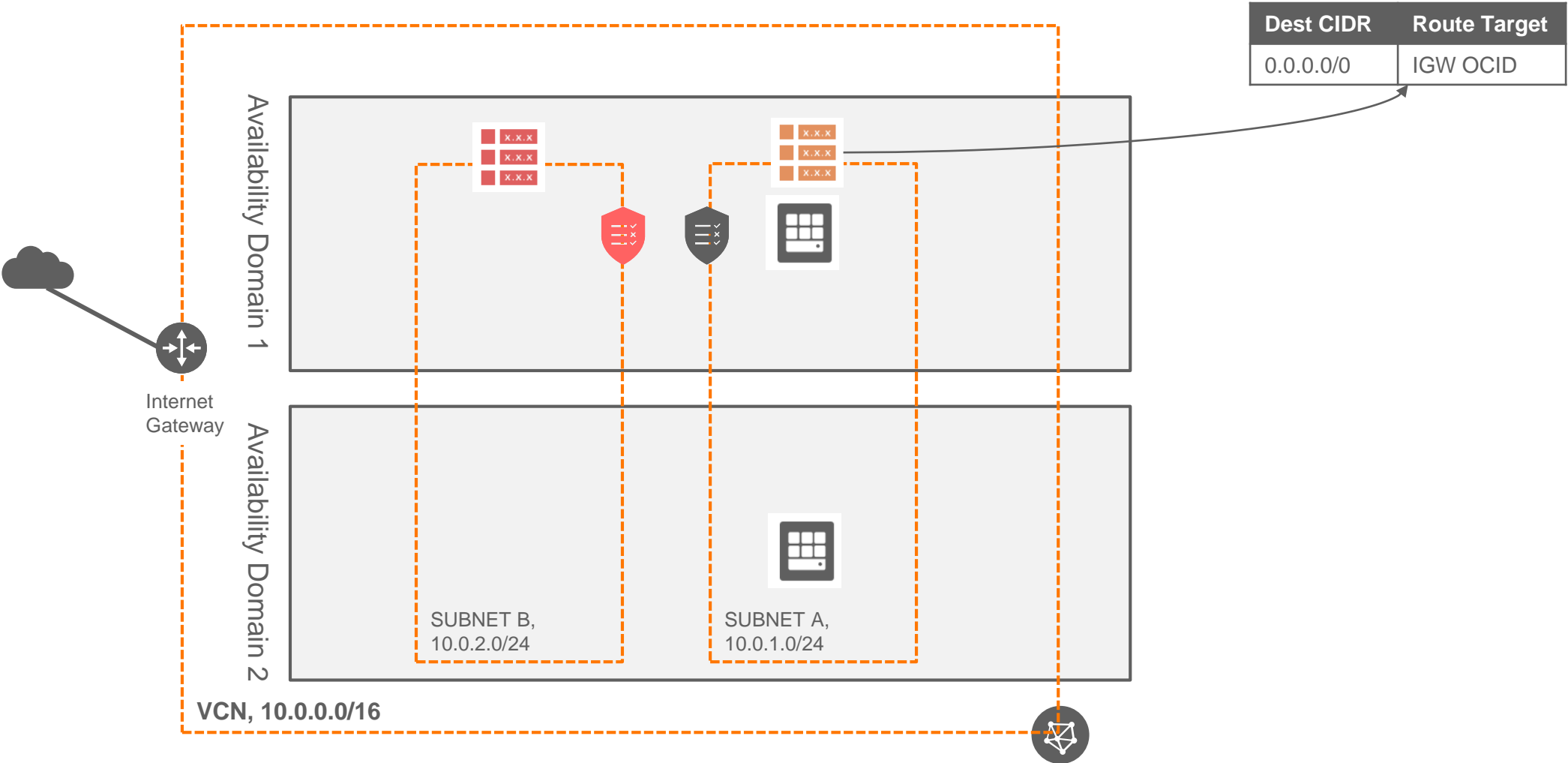
# 1: Create VCN, a regional subnet and instances with web servers



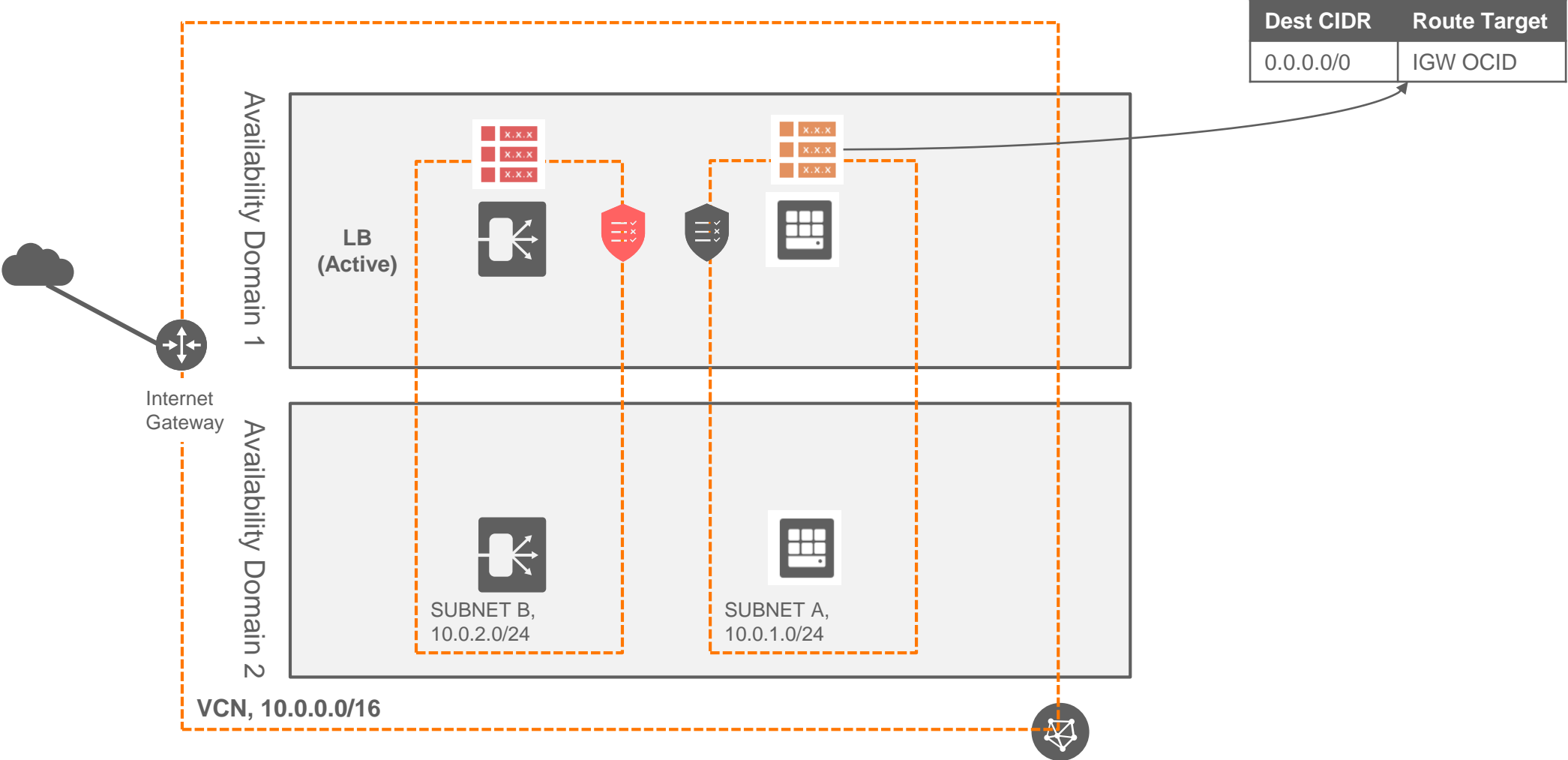
# 2: Create Load Balancer Route Table and Security List



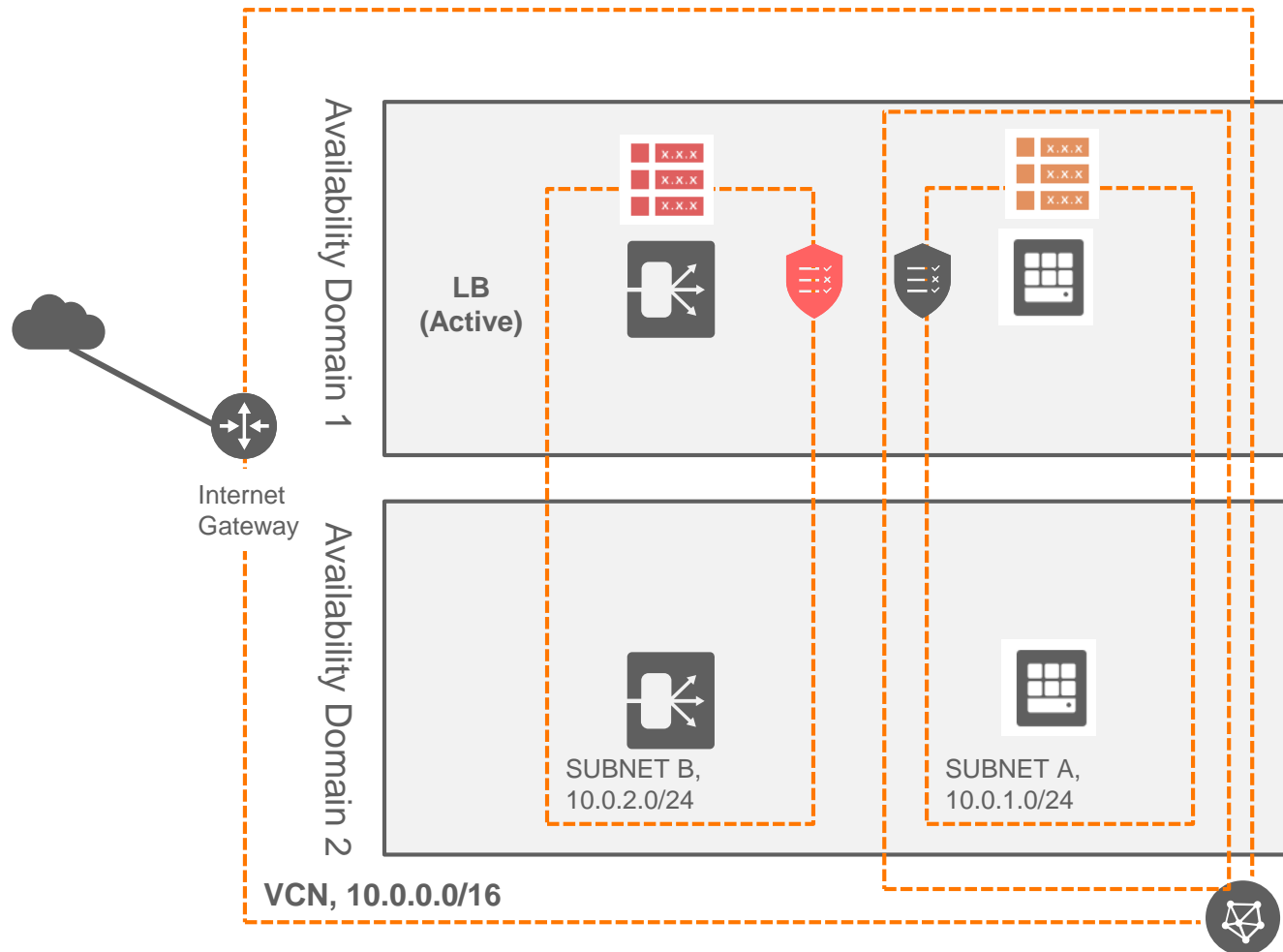
### 3: Create a regional Public Subnet for Load Balancer in your VCN



# 4.1 Create the Public Load Balancer



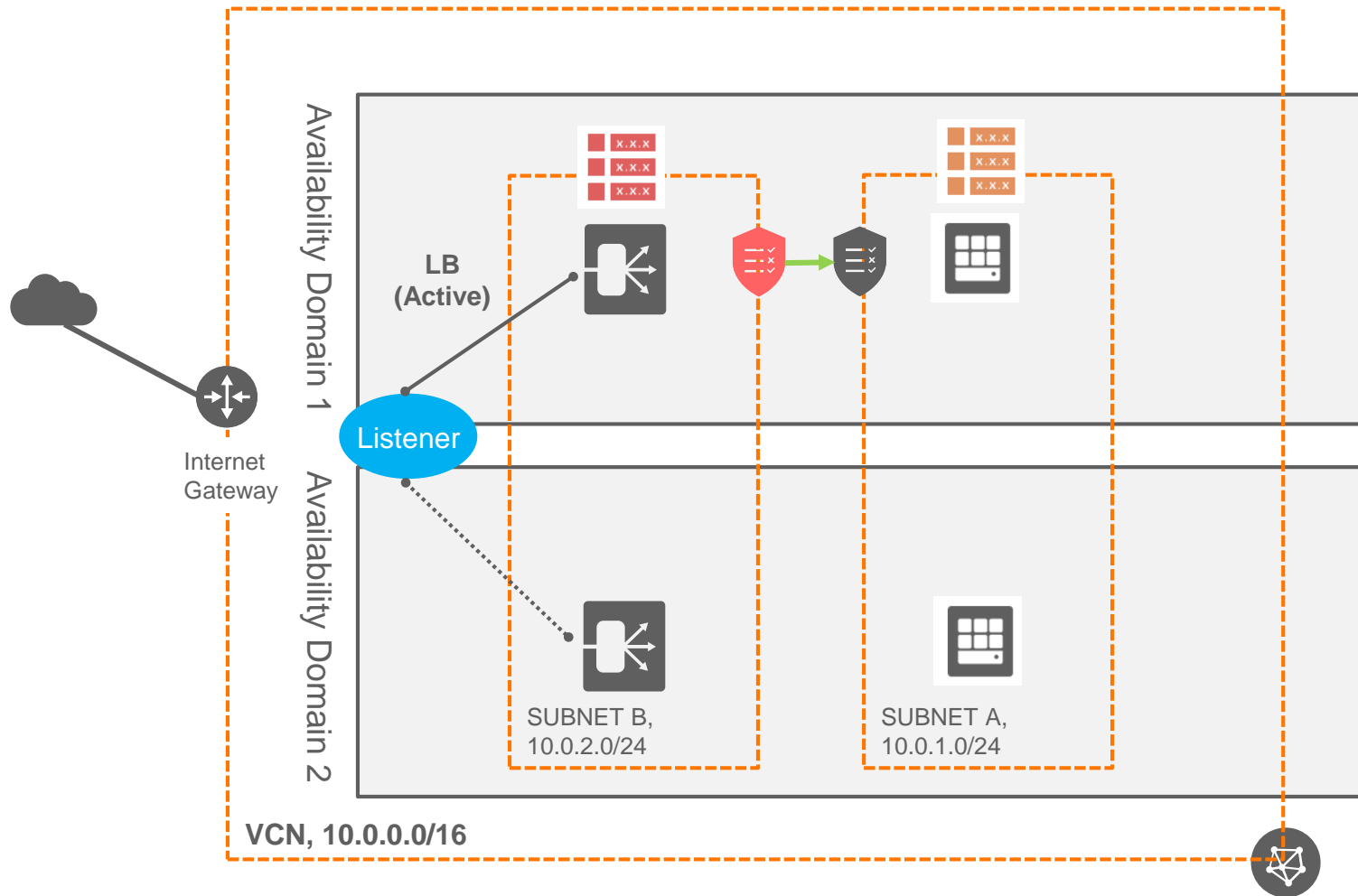
## 4.2 Add a backend set and add servers to backend set



- Add a Backend Set
  - Name
  - Policy
  - SSL
  - Health Check
- Add backend Servers to backend set
  - Instance
  - Port
  - Weight

*Rules are **NOT** added to LB Security Lists and backend server security lists automatically, the user has to manually open them to allow traffic between them.*

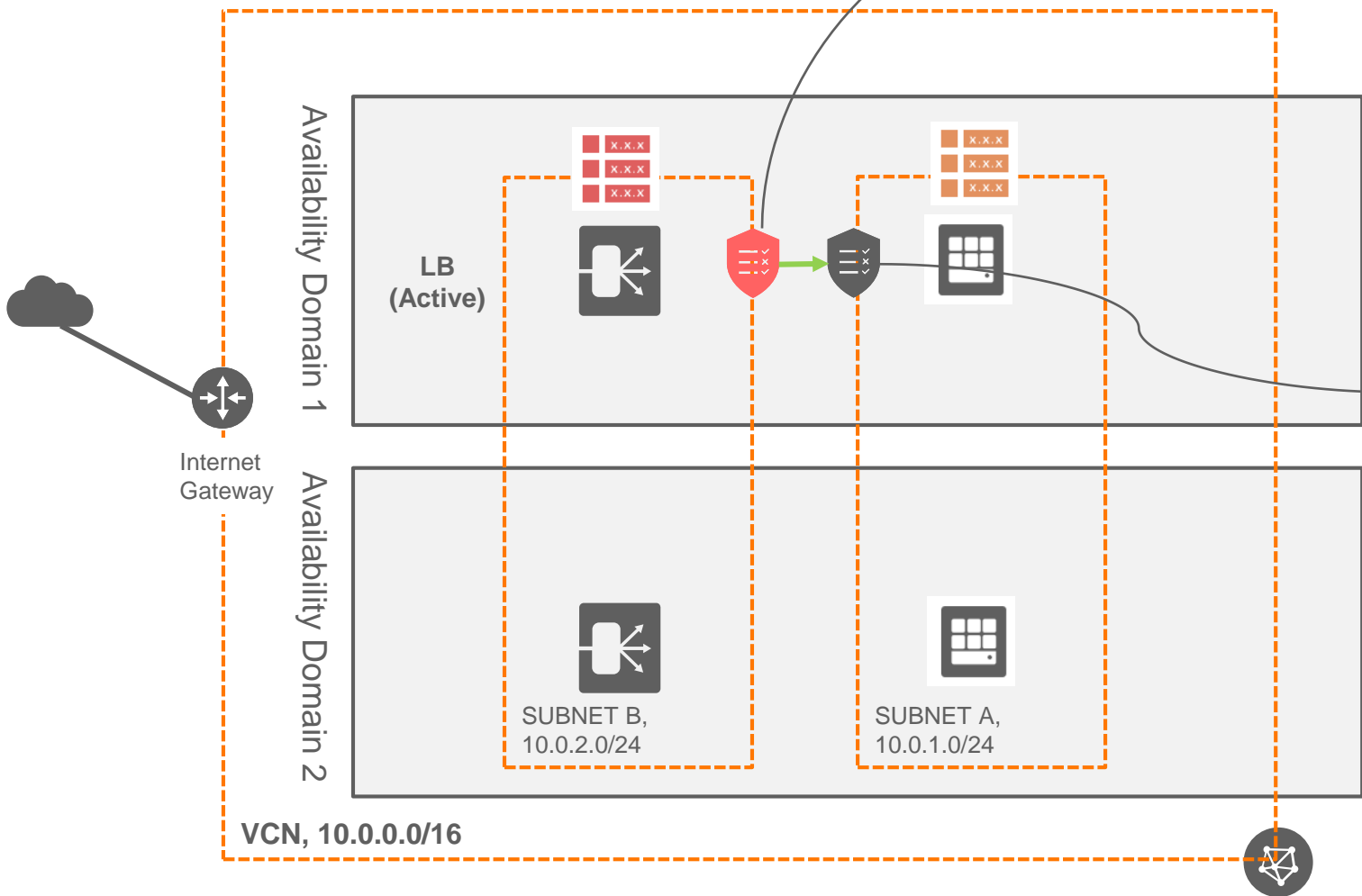
## 4.3 Add a Listener



- Add a Listener
  - Name
  - Protocol, Port, SSL
  - Backend Set

# 5: LB SL and Backend SL

Type	CIDR	Protocol	Source Port	Dest Port
Egress	10.0.1.0/24	TCP	All	80

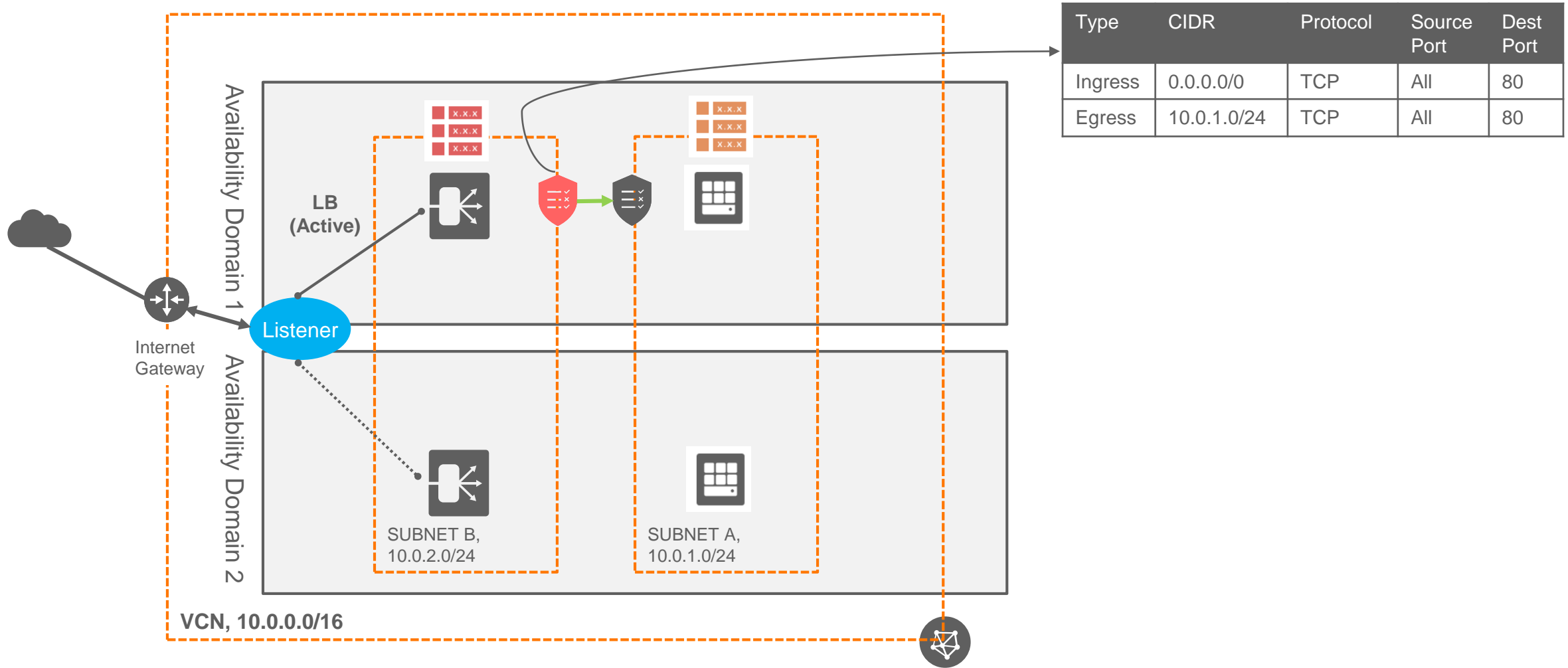


Type	CIDR	Protocol	Source Port	Dest Port
Ingress	10.0.2.0/24	TCP	All	80
Egress	0.0.0.0/0	All	All	





# 6: Allow traffic to Listener



# Pricing

OCI Load Balancer is priced as below

	Metric	Pay as You Go	Monthly Flex
OCI - 100 Mbps Load Balancer	Load Balancer Hour	\$0.0213	\$0.0213
OCI - 400 Mbps Load Balancer	Load Balancer Hour	\$0.0850	\$0.0850
OCI - 8000 Mbps Load Balancer	Load Balancer Hour	\$1.7	\$1.7

There is no separate charge for any data handled at the Load Balancer tier

[http://esource.oraclecorp.com/portal/page/portal/ROOTFOLDER/PRICE/PRICING\\_ENGINE/CURR\\_PRICE\\_LIST/ORACLE%20PAAS%20AND%20IAAS%20PUBLIC%20CLOUD%20GLOBAL%20PRICE%20LIST.PDF](http://esource.oraclecorp.com/portal/page/portal/ROOTFOLDER/PRICE/PRICING_ENGINE/CURR_PRICE_LIST/ORACLE%20PAAS%20AND%20IAAS%20PUBLIC%20CLOUD%20GLOBAL%20PRICE%20LIST.PDF)

# Summary

- OCI Load Balancer provides automated traffic distribution from one entry point to multiple servers in VCN; Improves resource utilization, facilitates scaling, and helps ensure high availability
- OCI Load Balancer is a regional service and supports Public or Private Load Balancers
- Supports multiple Protocols – TCP, HTTP/1.0, HTTP/1.1, HTTP/2, WebSocket
- Supports SSL Offloading – SSL Termination, End-to-End SSL, SSL Tunneling
- Supports session persistence and content based routing
- Key differentiators
  - Private or Public Load Balancer (with Public IP address)
  - Provisioned bandwidth – 100 Mbps, 400 Mbps, 8 Gbps
  - Single load balancer for TCP (layer 4) and HTTP (layer 7) traffic

**ORACLE<sup>®</sup>**  
Cloud Infrastructure

[cloud.oracle.com/iaas](https://cloud.oracle.com/iaas)

[cloud.oracle.com/tryit](https://cloud.oracle.com/tryit)