LOAD BALANCER (OCI)

1. What is OCI?

- OCI, or Oracle Cloud Infrastructure, is Oracle's cloud computing platform. It offers a range of services for deploying and scaling applications with high performance and security. OCI is known for its global presence, robust infrastructure, and support for diverse workloads, making it a preferred choice for enterprises.

2. What is OCI Security?

- OCI security involves protective measures in Oracle Cloud Infrastructure to safeguard data and resources from unauthorized access and cyber threats, using features like encryption and identity management.

3. What is Load Balancer?

- A load balancer in Oracle Cloud Infrastructure (OCI) is a networking service designed to evenly distribute incoming traffic across multiple servers or instances to ensure optimal performance, reliability, and resource utilization. It acts as a traffic cop, directing requests from clients to backend servers based on predefined algorithms.

Technical Example:

In a technical scenario, suppose you have a web application deployed on multiple virtual machines in OCI. When users access your application, the load balancer efficiently distributes incoming requests across these VMs, ensuring that no single server becomes overwhelmed with traffic. For instance, if one server is experiencing high load, the load balancer intelligently redirects new requests to other available servers, maintaining smooth operation of the application.

Non-Technical Example:

Think of a load balancer as a traffic conductor for your online services. Imagine you're hosting a popular shopping website with multiple checkout counters. Instead of having all customers queue up at a single counter, the load balancer ensures that incoming shoppers are directed to the least busy checkout counter. This way, customers experience faster checkout times, and the overall shopping experience remains smooth, even during peak hours.

4. Importance of Load Balancer?

- The importance of a load balancer in Oracle Cloud Infrastructure (OCI) lies in its ability to enhance the performance, availability, and scalability of cloud-based applications and services. Here are key reasons highlighting the importance of a load balancer:

Optimized Resource Utilization:

High Availability:

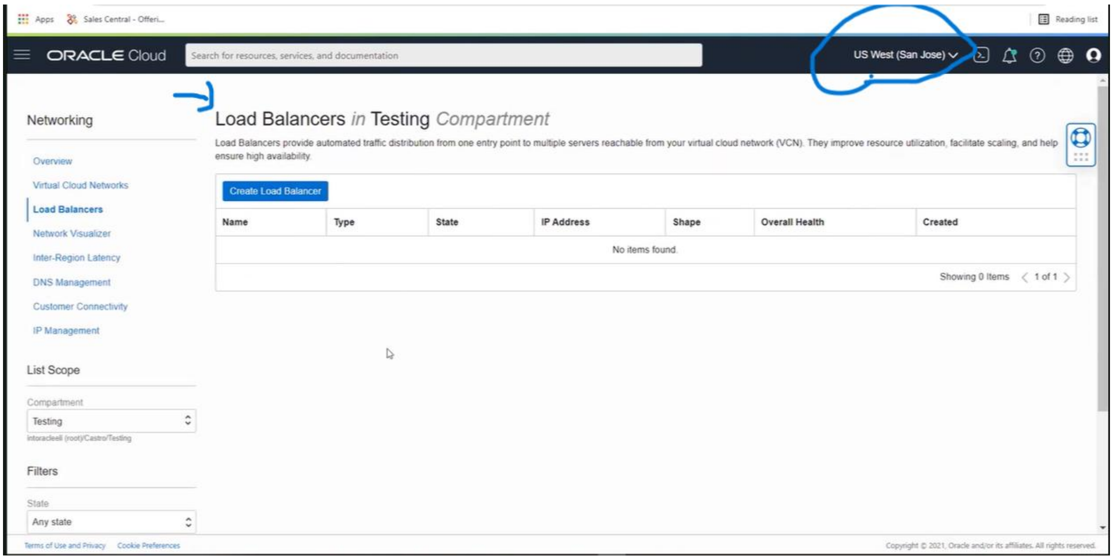
Scalability:

Fault Tolerance:

Security:

Cost Efficiency:

STEP 1 :- In our tenancy which is in US we will create a Load Balancer



Load Balancers *in* Testing *Compartment*

Load Balancers provide automated traffic distribution from one entry point to multiple servers reachable from your virtual cloud network (VCN). They improve resource utilization, facilitate scaling, and help ensure high availability.

Name	Type	State	IP Address	Shape	Overall Health	Created
No items found.						

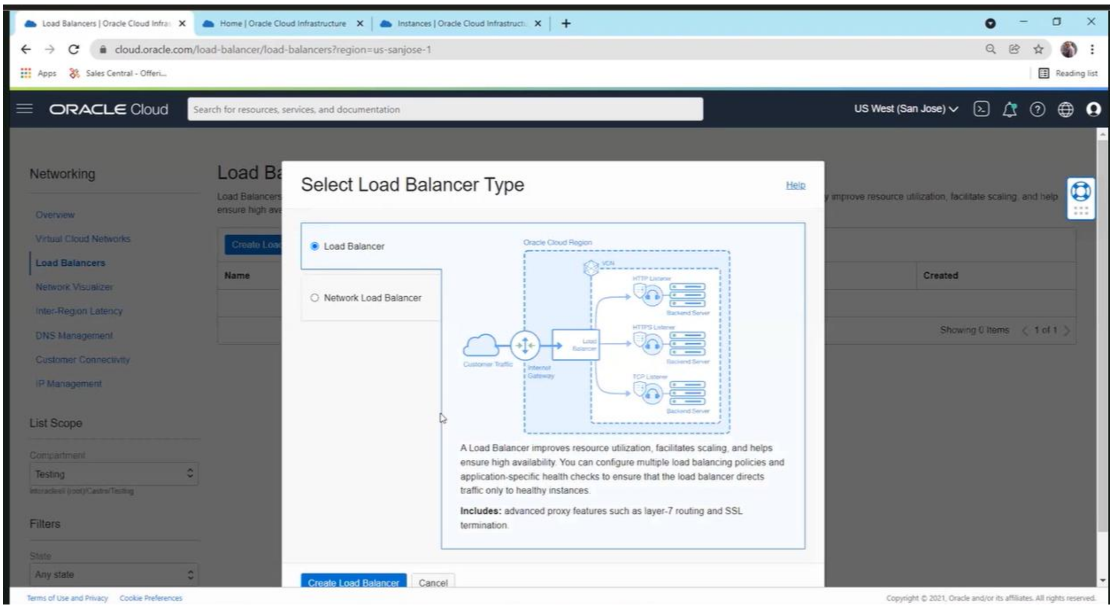
Showing 0 Items 〈 1 of 1 〉

Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

STEP 2:- Click on create Loan balancer and choose the 1st option

Load balancer = Application layer (layer 7)

Network Load balancer = Network Layer (layer 3)



Select Load Balancer Type

● Load Balancer

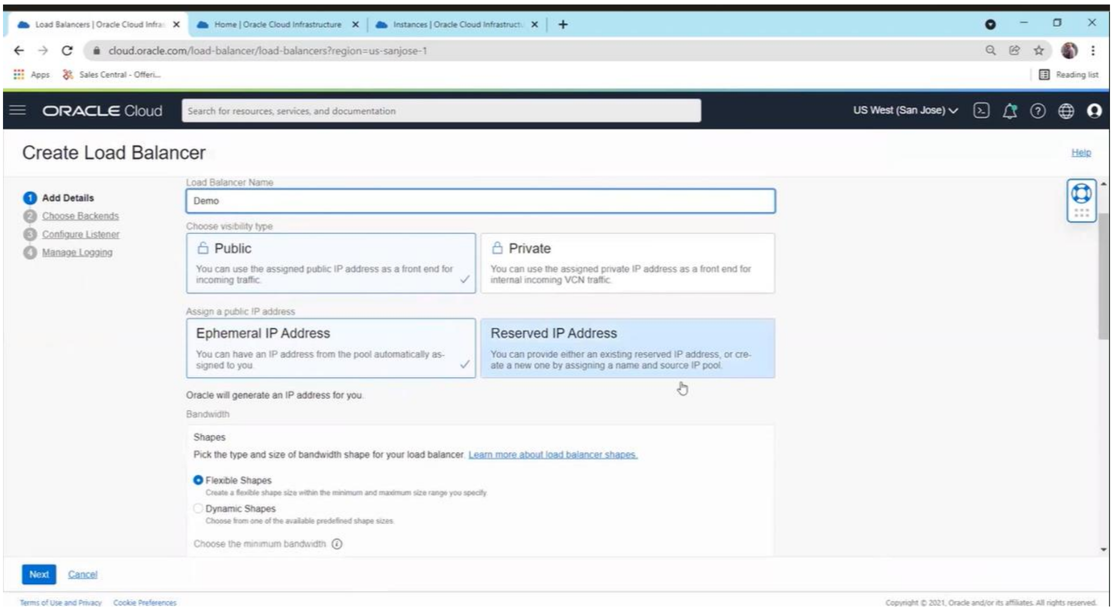
○ Network Load Balancer

A Load Balancer improves resource utilization, facilitates scaling, and helps ensure high availability. You can configure multiple load balancing policies and application-specific health checks to ensure that the load balancer directs traffic only to healthy instances.

Includes: advanced proxy features such as layer-7 routing and SSL termination.

Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

STEP 3 :- Give the name as required, choose between public and private, and assign IP from Ephemeral IP or Reserved IP


Load Balancers | Oracle Cloud Infra
Home | Oracle Cloud Infrastructure
Instances | Oracle Cloud Infrastructu
cloud.oracle.com/load-balancer/load-balancers?region=us-sanjose-1
Apps
Sales Central - Offeri...
Reading list
ORACLE Cloud
Search for resources, services, and documentation
US West (San Jose)
Help
Terms of Use and Privacy
Cookie Preferences
Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

Create Load Balancer

1 Add Details
2 Choose Backends
3 Configure Listener
4 Manage Logging

Load Balancer Name: Demo

Choose visibility type

- Public ✓ — You can use the assigned public IP address as a front end for incoming traffic.
- Private — You can use the assigned private IP address as a front end for internal incoming VCN traffic.

Assign a public IP address

- Ephemeral IP Address ✓ — You can have an IP address from the pool automatically assigned to you.
- Reserved IP Address — You can provide either an existing reserved IP address, or create a new one by assigning a name and source IP pool.

Oracle will generate an IP address for you.

Bandwidth

Shapes

Pick the type and size of bandwidth shape for your load balancer. Learn more about load balancer shapes.

● Flexible Shapes
Create a flexible shape size within the minimum and maximum size range you specify

○ Dynamic Shapes
Choose from one of the available predefined shape sizes.

Choose the minimum bandwidth

Next Cancel

Public vs. Private Load Balancer:

Public Load Balancer: This type of load balancer is accessible from the internet, allowing external users to send requests to your applications or services. It's commonly used for public-facing applications or services that need to be accessed by users over the internet.

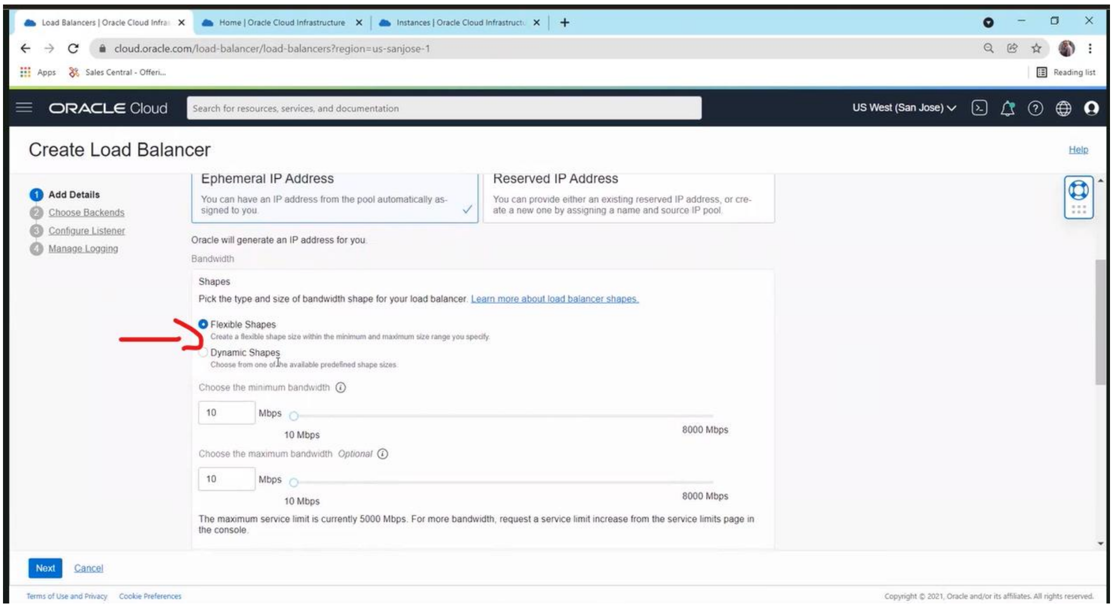
Private Load Balancer: Unlike a public load balancer, a private load balancer is not directly accessible from the internet. It's deployed within a private subnet and is primarily used for internal communication between resources within your virtual cloud network (VCN). Private load balancers are suitable for applications or services that do not need to be exposed to the public internet, enhancing security by limiting external access.

Ephemeral IP vs. Reserved IP:

Ephemeral IP: An ephemeral IP address is dynamically assigned to the load balancer and is temporary in nature. It's typically used for short-term deployments or testing purposes. Ephemeral IPs are released when the associated load balancer is terminated, and they cannot be reserved for long-term use.

Reserved IP: A reserved IP address is a static, persistent IP address that you can reserve and assign to your load balancer. Unlike ephemeral IPs, reserved IPs remain associated with the load balancer even if it's terminated and can be reused across different deployments. Reserved IPs are suitable for production environments or scenarios where consistent IP addressing is required for routing traffic.

STEP 4 :- Keep in mind While creating chooser **Flexible Shape** cause web application firewall policy will only work with Flexible shape.



Create Load Balancer

1 Add Details
2 Choose Backends
3 Configure Listener
4 Manage Logging

Ephemeral IP Address
You can have an IP address from the pool automatically assigned to you.

Reserved IP Address
You can provide either an existing reserved IP address, or create a new one by assigning a name and source IP pool.

Oracle will generate an IP address for you.

Bandwidth

Shapes
Pick the type and size of bandwidth shape for your load balancer. Learn more about load balancer shapes.

● Flexible Shapes
Create a flexible shape size within the minimum and maximum size range you specify.

○ Dynamic Shapes
Choose from one of the available predefined shape sizes.

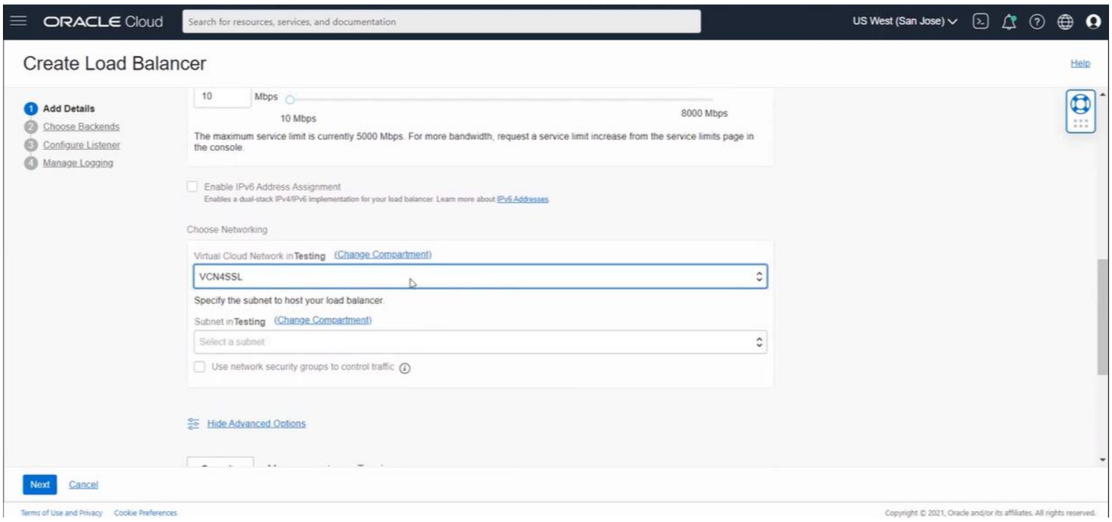
Choose the minimum bandwidth: 10 Mbps (10 Mbps – 8000 Mbps)

Choose the maximum bandwidth *Optional*: 10 Mbps (10 Mbps – 8000 Mbps)

The maximum service limit is currently 5000 Mbps. For more bandwidth, request a service limit increase from the service limits page in the console.

Load Balancers | Oracle Cloud Infra
Home | Oracle Cloud Infrastructure
Instances | Oracle Cloud Infrastructu
cloud.oracle.com/load-balancer/load-balancers?region=us-sanjose-1
Apps
Sales Central - Offeri...
Reading list
ORACLE Cloud
Search for resources, services, and documentation
US West (San Jose)
Help
Next
Cancel
Terms of Use and Privacy
Cookie Preferences
Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

STEP 5 :- Choose the VCN and Subnet in which we want to create the Load balancer



Create Load Balancer

1 Add Details
2 Choose Backends
3 Configure Listener
4 Manage Logging

10 Mbps (10 Mbps – 8000 Mbps)

The maximum service limit is currently 5000 Mbps. For more bandwidth, request a service limit increase from the service limits page in the console.

□ Enable IPv6 Address Assignment
Enables a dual-stack IPv4/IPv6 implementation for your load balancer. Learn more about IPv6 Addresses.

Choose Networking

Virtual Cloud Network in **Testing** (Change Compartment)
VCN4SSL

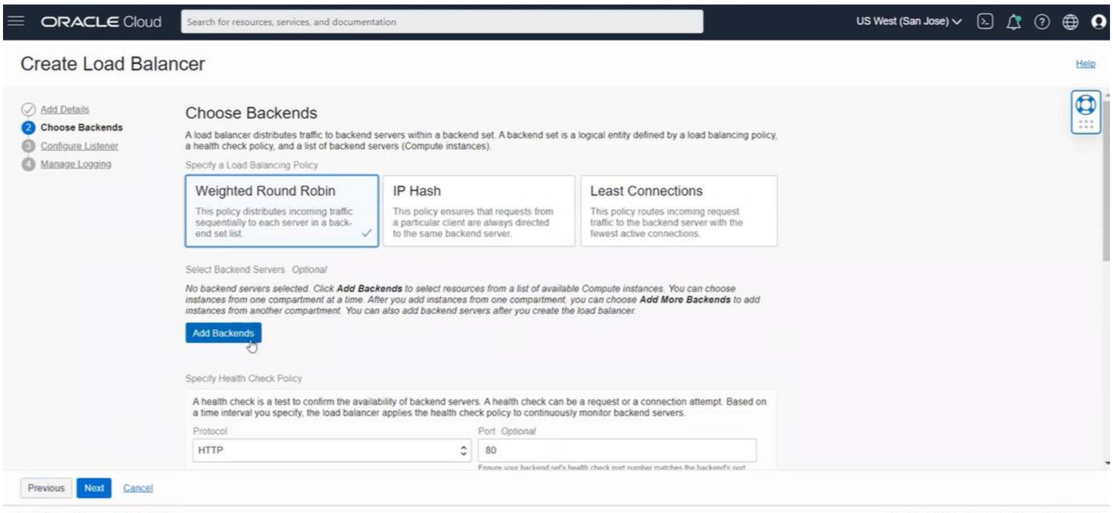
Specify the subnet to host your load balancer.

Subnet in **Testing** (Change Compartment)
Select a subnet

□ Use network security groups to control traffic

ORACLE Cloud
Search for resources, services, and documentation
US West (San Jose)
Help
Hide Advanced Options
Next
Cancel
Terms of Use and Privacy
Cookie Preferences
Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

STEP 6 :- Click on next and choose from Weighted Round Robin, IP Hash, Least Connections



Create Load Balancer

- Add Details
- 2 Choose Backends
- 3 Configure Listener
- 4 Manage Logging

Choose Backends

A load balancer distributes traffic to backend servers within a backend set. A backend set is a logical entity defined by a load balancing policy, a health check policy, and a list of backend servers (Compute instances).

Specify a Load Balancing Policy

Weighted Round Robin ✓
This policy distributes incoming traffic sequentially to each server in a back-end set list.

IP Hash
This policy ensures that requests from a particular client are always directed to the same backend server.

Least Connections
This policy routes incoming request traffic to the backend server with the fewest active connections.

Select Backend Servers *Optional*

*No backend servers selected. Click **Add Backends** to select resources from a list of available Compute instances. You can choose instances from one compartment at a time. After you add instances from one compartment, you can choose **Add More Backends** to add instances from another compartment. You can also add backend servers after you create the load balancer.*

Add Backends

Specify Health Check Policy

A health check is a test to confirm the availability of backend servers. A health check can be a request or a connection attempt. Based on a time interval you specify, the load balancer applies the health check policy to continuously monitor backend servers.

Protocol: HTTP

Port *Optional*: 80

Ensure your backend set's health check port number matches the backend's port

Previous Next Cancel

Weighted Round Robin:

This algorithm assigns a weight to each server based on its capacity or performance.

Traffic is distributed to servers in a cyclic manner according to their weights.

Servers with higher weights receive more traffic compared to those with lower weights.

Useful when you want to prioritize certain servers over others based on their capabilities.

IP Hash:

In this algorithm, the source IP address of the client is used to determine which server will handle the request.

The hash function generates a unique identifier from the client's IP address, and this identifier is used to select the server.

Ensures that requests from the same client are always routed to the same server.

Useful for session persistence or when maintaining state information between the client and server.

Least Connection:

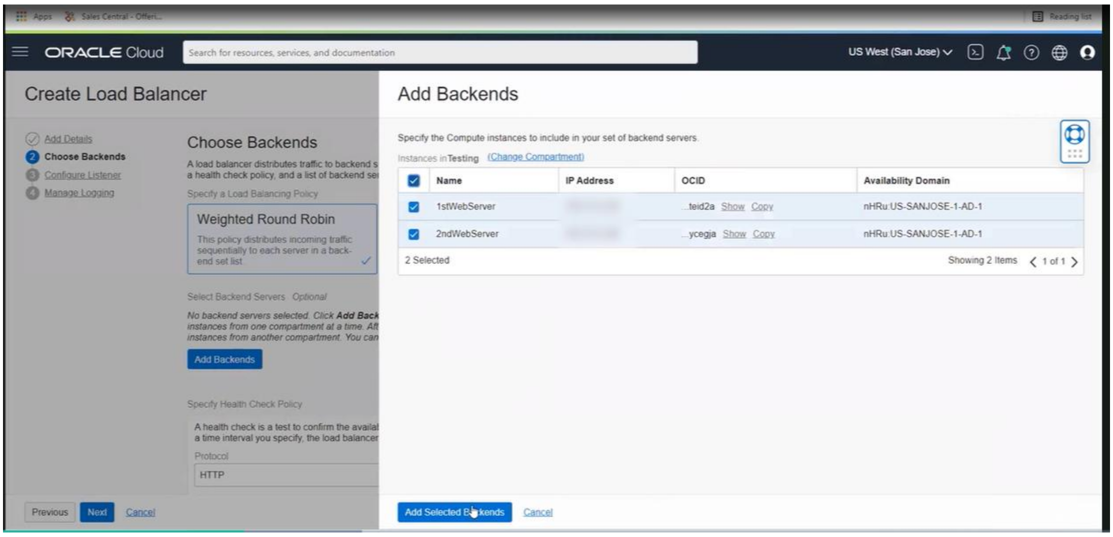
This algorithm directs traffic to the server with the fewest active connections at the time of the request.

It dynamically adjusts the load distribution based on the current workload of each server.

New connections are sent to the server with the least number of active connections.

Helpful in evenly distributing the load among backend servers and preventing overloading of any single server.

STEP 7 :- Add Back-end Servers



Apps
Sales Central - Offeri...
Reading list
ORACLE Cloud
Search for resources, services, and documentation
US West (San Jose)

Create Load Balancer

Choose Backends

A load balancer distributes traffic to backend s
a health check policy, and a list of backend se

Specify a Load Balancing Policy

Weighted Round Robin ✓

This policy distributes incoming traffic sequentially to each server in a back-end set list

Select Backend Servers Optional

*No backend servers selected. Click **Add Back** instances from one compartment at a time. Aft instances from another compartment. You can*

Specify Health Check Policy

A health check is a test to confirm the availal
a time interval you specify, the load balancer

Protocol: HTTP

Add Backends

Specify the Compute instances to include in your set of backend servers.

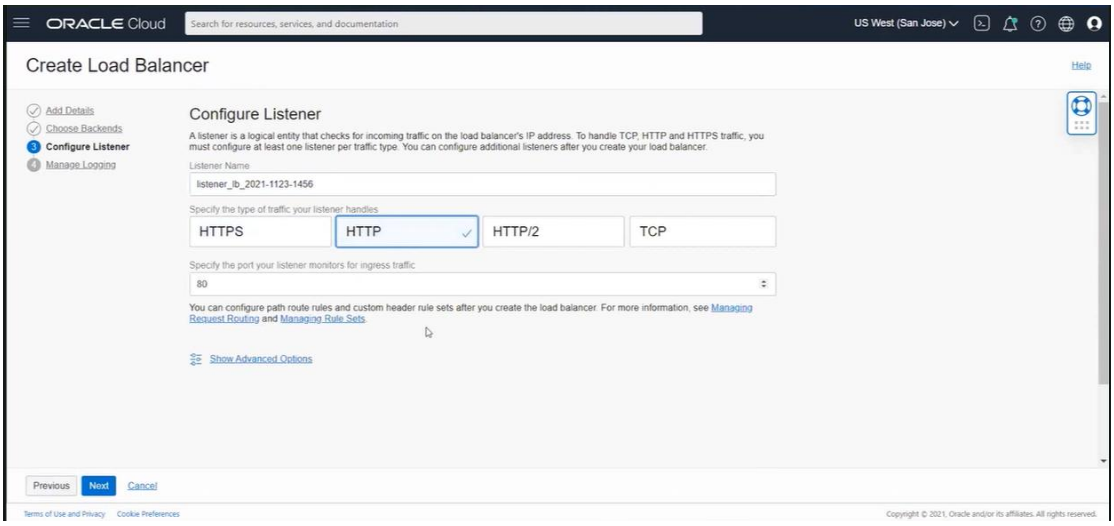
Instances in **Testing** (Change Compartment)

☑	Name	IP Address	OCID	Availability Domain
☑	1stWebServer	[redacted]	...teid2a Show Copy	nHRu:US-SANJOSE-1-AD-1
☑	2ndWebServer	[redacted]	...ycegja Show Copy	nHRu:US-SANJOSE-1-AD-1

2 Selected

Showing 2 Items < 1 of 1 >

STEP 8 :- Click on next and choose on which protocol you want to configure Listener



Create Load Balancer

Configure Listener

A listener is a logical entity that checks for incoming traffic on the load balancer's IP address. To handle TCP, HTTP and HTTPS traffic, you must configure at least one listener per traffic type. You can configure additional listeners after you create your load balancer.

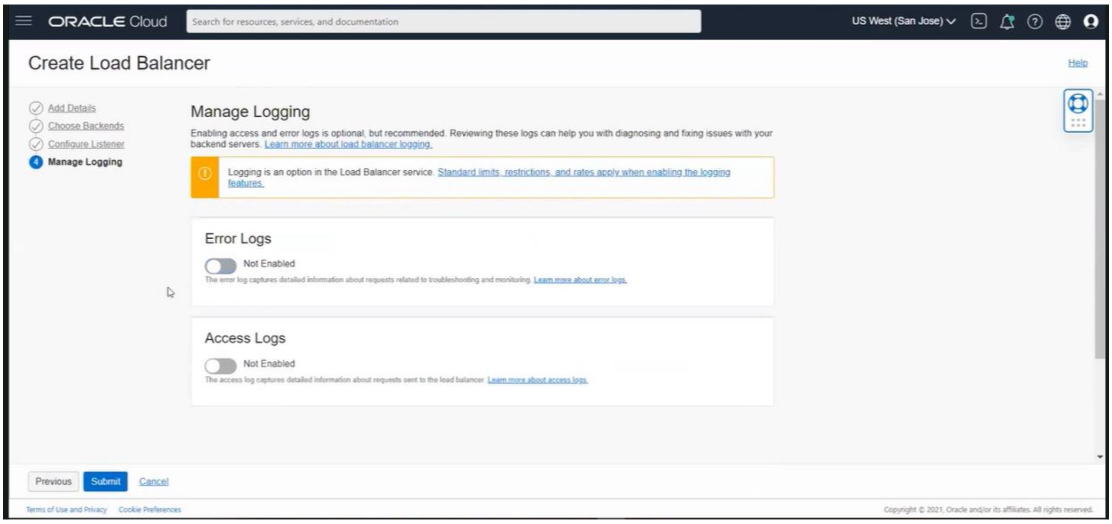
Listener Name: listener_lb_2021-1123-1456

Specify the type of traffic your listener handles: HTTPS [✓ HTTP] HTTP/2 TCP

Specify the port your listener monitors for ingress traffic: 80

You can configure path route rules and custom header rule sets after you create the load balancer. For more information, see Managing Request Routing and Managing Rule Sets.

STEP 9 :- Click next ad Choose if you want enable or disable log and then click on submit



Create Load Balancer

Manage Logging

Enabling access and error logs is optional, but recommended. Reviewing these logs can help you with diagnosing and fixing issues with your backend servers. Learn more about load balancer logging.

Logging is an option in the Load Balancer service. Standard limits, restrictions, and rates apply when enabling the logging features.

Error Logs

Not Enabled

The error log captures detailed information about requests related to troubleshooting and monitoring. Learn more about error logs.

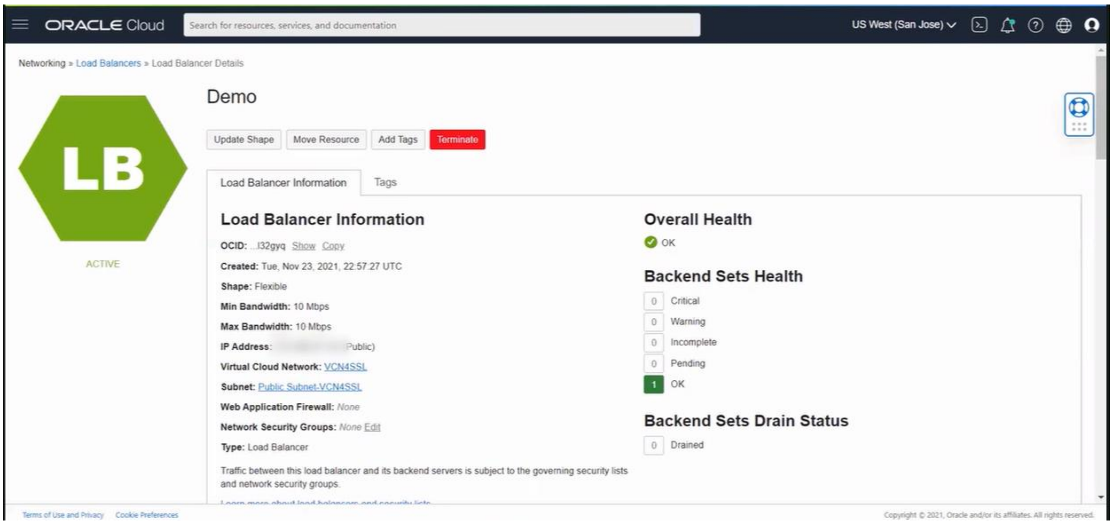
Access Logs

Not Enabled

The access log captures detailed information about requests sent to the load balancer. Learn more about access logs.

Copyright © 2021, Oracle and/or its affiliates. All rights reserved.

STEP 10 :- Successfully Created Load Balancer



Demo

Load Balancer Information

OCID: ...l32gyq Show Copy

Created: Tue, Nov 23, 2021, 22:57:27 UTC

Shape: Flexible

Min Bandwidth: 10 Mbps

Max Bandwidth: 10 Mbps

IP Address: [redacted] Public)

Virtual Cloud Network: VCN4SSL

Subnet: Public Subnet-VCN4SSL

Web Application Firewall: *None*

Network Security Groups: *None* Edit

Type: Load Balancer

Traffic between this load balancer and its backend servers is subject to the governing security lists and network security groups.

Overall Health

✓ OK

Backend Sets Health

0	Critical
0	Warning
0	Incomplete
0	Pending
1	OK

Backend Sets Drain Status

0 Drained

Copyright © 2021, Oracle and/or its affiliates. All rights reserved.