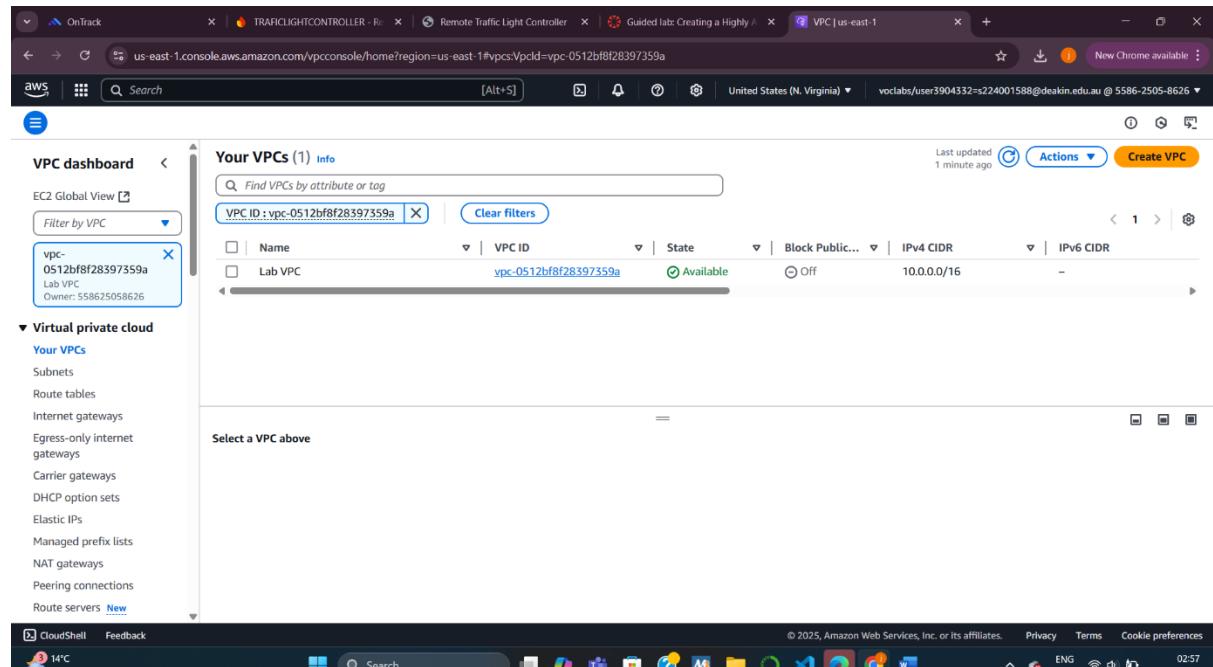


TASK 7.1P

Creating a Highly Available Environment

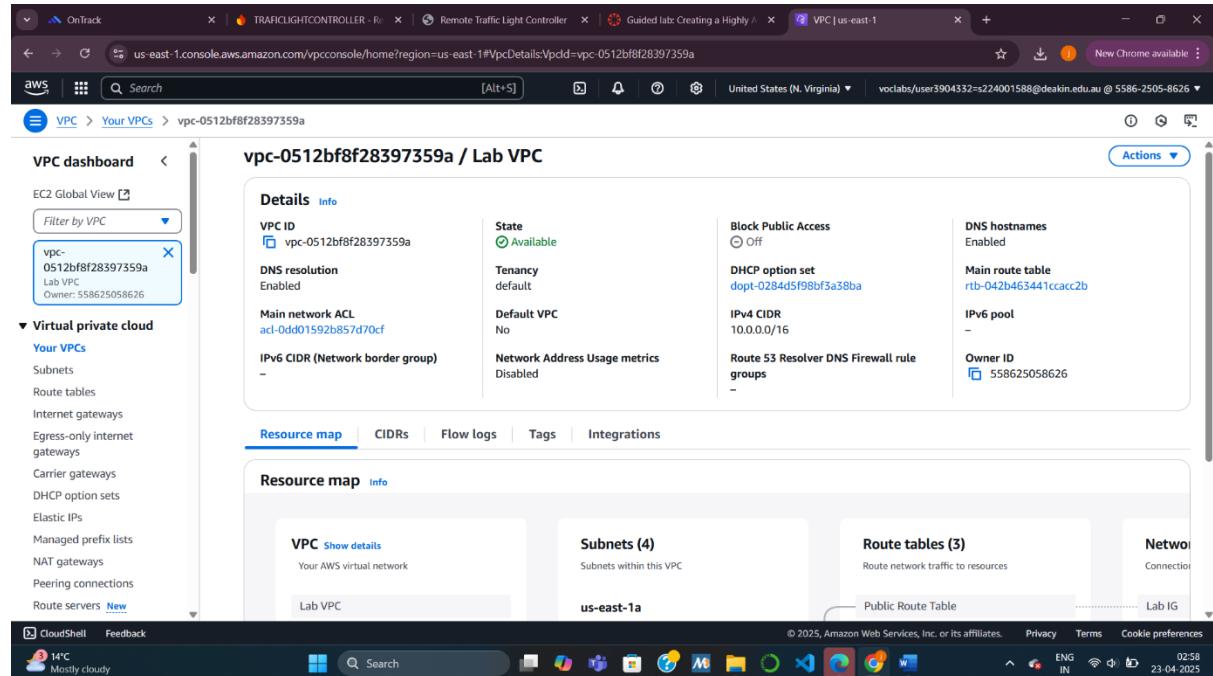
Task 1: Inspecting Your VPC

In this task, I reviewed the pre-configured Virtual Private Cloud (VPC) and its components. The VPC included public and private subnets across two Availability Zones, an internet gateway for public subnets, a NAT gateway for outbound internet access from private subnets, and an Amazon RDS instance located within a private subnet. This inspection helped me understand the network architecture, including routing tables, network ACLs, and existing security groups. The key takeaway was to ensure that only required components have internet access while maintaining private communication between tiers.



The screenshot shows the AWS VPC dashboard. In the left sidebar, under 'Virtual private cloud', 'Your VPCs' is selected. A table lists one VPC entry:

Name	VPC ID	State	Block Public Access	IPv4 CIDR	IPv6 CIDR
Lab VPC	vpc-0512bf8f28397359a	Available	Off	10.0.0.0/16	-



The screenshot shows the details for the 'Lab VPC' (vpc-0512bf8f28397359a). The 'Details' tab is selected, displaying the following configuration:

VPC ID vpc-0512bf8f28397359a	State Available	Block Public Access Off	DNS hostnames Enabled
DNS resolution Enabled	Tenancy default	DHCP option set dopt-0284d5f98bf3a38ba	Main route table rtb-042b465441ccacc2b
Main network ACL acl-0dd01592b857d70cf	Default VPC No	IPv4 CIDR 10.0.0.0/16	IPv6 pool -
IPv6 CIDR (Network border group) -	Network Address Usage metrics Disabled	Route 53 Resolver DNS Firewall rule groups -	Owner ID 558625058626

The 'Resource map' section provides a visual overview of the VPC's components:

- VPC**: Show details, Your AWS virtual network, Lab VPC
- Subnets (4)**: Subnets within this VPC, us-east-1a, Public Route Table
- Route tables (3)**: Route network traffic to resources, Lab IG
- Network interface (1)**: Connection

OnTrack | **TRAFLIGHTCONTROLLER - Re** | **Remote Traffic Light Controller** | **Guided lab: Creating a Highly A** | **VPC | us-east-1** | + | - | X

us-east-1.console.aws.amazon.com/vpcconsole/home?region=us-east-1#SubnetDetails:subnetId=subnet-024e76557f414d926

aws Search [Alt+S] United States (N. Virginia) vocabs/user3904332=s24001588@deakin.edu.au @ 5586-2505-8626 ▾

VPC > Subnets > subnet-024e76557f414d926

subnet-024e76557f414d926 / Public Subnet 1

Actions

Details

Subnet ID subnet-024e76557f414d926	Subnet ARN arn:aws:ec2:us-east-1:558625058626:subnet/subnet-024e76557f414d926	State Available	Block Public Access <input type="radio"/> Off
IPv4 CIDR 10.0.0.0/24	IPv6 CIDR -	Network border group us-east-1	IPv6 CIDR association ID -
Availability Zone us-east-1a	Default subnet No	Auto-assign public IPv4 address Yes	VPC vpc-0512bf8f28397359a Lab VPC
Route table rtb-0d27c821e87a8ed98 Public Route Table	Customer-owned IPv4 pool -	Outpost ID -	Auto-assign public IPv4 address Yes
Auto-assign IPv6 address No	IPv6-only No	Hostname type IP name	Owner 558625058626
IPv4 CIDR reservations -	IPv6 CIDR reservations -	Resource name DNS AAAA record Disabled	
Resource name DNS A record Disabled	DNS64 Disabled		

Flow logs | **Route table** | **Network ACL** | **CIDR reservations** | **Sharing** | **Tags**

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences 02:58 ENG IN 23-04-2025

OnTrack | TRAFLIGHTCONTROLLER - Re | Remote Traffic Light Controller | Guided lab: Creating a Highly A | VPC | us-east-1 | + | - | X

us-east-1.console.aws.amazon.com/vpcconsole/home?region=us-east-1#SubnetDetails:subnetId=subnet-024e76557f414d926

aws Search [Alt+S] United States (N. Virginia) vocabs/user3904332=s24001588@deakin.edu.au @ 5586-2505-8626 ▾

VPC > Subnets > subnet-024e76557f414d926

subnet-024e76557f414d926 / Public Subnet 1

Details

Subnet ID subnet-024e76557f414d926	Subnet ARN arn:aws:ec2:us-east-1:558625058626:subnet/subnet-024e76557f414d926	State Available	Block Public Access <input type="radio"/> Off
IPv4 CIDR 10.0.0.0/24	IPv6 CIDR -	Network border group us-east-1	IPv6 CIDR association ID -
Availability Zone us-east-1a	Default subnet No	Auto-assign public IPv4 address Yes	VPC vpc-0512bf8f28397359a Lab VPC
Route table rtb-0d27c821e87a8ed98 Public Route Table	Customer-owned IPv4 pool -	Outpost ID -	Auto-assign public IPv4 address Yes
Auto-assign IPv6 address No	IPv6-only No	Hostname type IP name	Owner 558625058626
IPv4 CIDR reservations -	IPv6 CIDR reservations -	Resource name DNS AAAA record Disabled	
Resource name DNS A record Disabled	DNS64 Disabled		

Flow logs | **Route table** | **Network ACL** | **CIDR reservations** | **Sharing** | **Tags**

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences 02:58 ENG IN 23-04-2025

OnTrack | TRAFLIGHTCONTROLLER - Re | Remote Traffic Light Controller | Guided lab: Creating a Highly A | VPC | us-east-1 | + | - | X

us-east-1.console.aws.amazon.com/vpcconsole/home?region=us-east-1#SubnetDetails:subnetId=subnet-024e76557f414d926

aws Search [Alt+S] United States (N. Virginia) vocabs/user3904332=s24001588@deakin.edu.au @ 5586-2505-8626 ▾

VPC > Subnets > subnet-024e76557f414d926

Route table: rtb-0d27c821e87a8ed98 / Public Route Table

Edit route table association

Routes (2)

Destination	Target
10.0.0.0/16	local
0.0.0.0/0	igw-06dc7ad9c9498e954

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences 02:59 ENG IN 23-04-2025

Screenshot of the AWS VPC console showing the Network ACL configuration for a subnet.

Subnet Details:

- Resource name DNS A record: Disabled
- IPv6 CIDR reservations: -
- DNS64: Disabled
- Owner: 558625058626

Network ACL: acl-0dd01592b857d70cf

Inbound rules (2):

Rule number	Type	Protocol	Port range	Source	Allow/Deny
100	All traffic	All	All	0.0.0.0/0	Allow
*	All traffic	All	All	0.0.0.0/0	Deny

Outbound rules (2):

Rule number	Type	Protocol	Port range	Destination	Allow/Deny
100	All traffic	All	All	0.0.0.0/0	Allow
*	All traffic	All	All	0.0.0.0/0	Deny

Internet Gateways (1) Info:

Name	Internet gateway ID	State	VPC ID	Owner
Lab IG	igw-06dc7ad9c9498e954	Attached	vpc-0512bf8f28397359a Lab VPC	558625058626

Select an internet gateway above:

Screenshot of the AWS VPC console showing the Internet gateway details for igw-06dc7ad9c9498e954.

Details

Internet gateway ID	igw-06dc7ad9c9498e954	State	Attached
		VPC ID	ypc-0512bf8f28397359a Lab VPC
		Owner	558625058626

Tags

Key	Value
Name	Lab IG
aws:cloudfor...	arn:aws:cloudformation:us-east-1:558625058626:stack/c144539a3736923l10063377t1w558625058626/c46b9e90-1ff9-11f0-9185-0affe6784705
aws:cloudfor...	IGW
aws:cloudfor...	c144539a3736923l10063377t1w558625058626
cloudlab	c144539a3736923l10063377t1w558625058626

Security Groups (3)

Name	Security group ID	Security group name	VPC ID	Description
-	sg-0befd43e53dcf017e	default	ypc-0512bf8f28397359a	default VPC secur
Inventory-DB	sg-00522d8c69448435c	Inventory-DB	ypc-0512bf8f28397359a	Enable access to I
Inventory-App	sg-0b6a32a5892c433b1	Inventory-App	ypc-0512bf8f28397359a	Enable access to /

Select a security group

Screenshot of the AWS VPC Security Groups console showing the configuration of a security group named "Inventory-DB".

Details:

- Security group name:** Inventory-DB
- Security group ID:** sg-00522d8c69448435c
- Description:** Enable access to MySQL
- VPC ID:** vpc-0512bf8f28397359a
- Owner:** 558625058626
- Inbound rules count:** 1 Permission entry
- Outbound rules count:** 1 Permission entry

Inbound rules (1):

Name	Security group rule ID	IP version	Type	Protocol	Port range
-	sgr-08e20cff19c9c4321	IPv4	MySQL/Aurora	TCP	3306

Outbound rules (1):

version	Type	Protocol	Port range	Source	Description
v4	MySQL/Aurora	TCP	3306	10.0.0.0/16	-

The screenshot shows the AWS VPC Security Groups console. On the left, a sidebar lists various VPC-related services. The main area displays the details for a security group named 'sg-00522d8c69448435c - Inventory-DB'. The 'Outbound rules' tab is selected, showing one rule: 'sgr-05aea79fee405f604' allowing IPv4 traffic of all types on all ports from the security group itself.

Name	Security group rule ID	IP version	Type	Protocol	Port range
-	sgr-05aea79fee405f604	IPv4	All traffic	All	All

Task 2: Creating an Application Load Balancer (ALB)

Here, I set up an Application Load Balancer (ALB) to distribute traffic across multiple instances. I configured the ALB to operate in two public subnets (each in different Availability Zones), ensuring resilience against data center failures. I also created a new security group (Inventory-LB) to allow HTTP and HTTPS traffic. A Target Group (Inventory-App) was configured with health check settings to continuously monitor the health of EC2 instances. This task ensured that incoming traffic would only be directed to healthy application servers, contributing to the high availability of the application.

Screenshot of the AWS Lambda console showing the 'Load balancers' page. The sidebar on the left shows various AWS services like EC2, Images, Elastic Block Store, Network & Security, Load Balancing, Auto Scaling, and more. The main content area displays a table titled 'Load balancers' with a single row: 'No load balancers'. A message below the table says 'You don't have any load balancers in us-east-1'. A blue 'Create load balancer' button is located at the bottom right of the table.

Screenshot of the AWS Lambda console showing the 'Create Application Load Balancer' wizard. The first step, 'IP pools - new', is displayed. It asks if you want to use an IPAM pool for public IPv4 addresses. There is a checkbox labeled 'Use IPAM pool for public IPv4 addresses'. Below this, there is a section for 'Availability Zones and subnets'. Under 'us-east-1a (use1-az2)', it lists a subnet: 'subnet-024e7655f414d926' (IPv4 subnet CIDR: 10.0.0.0/24). Under 'us-east-1b (use1-az4)', it lists another subnet: 'subnet-07302aedd32135d70' (IPv4 subnet CIDR: 10.0.1.0/24). At the bottom, there is a 'Security groups' section with a note about selecting firewall rules for the load balancer.

The screenshot shows the AWS CloudShell interface with multiple tabs open. The active tab is titled "Create security group".

Basic details

Security group name [Info](#)
Inventory-LB
Name cannot be edited after creation.

Description [Info](#)
Allows SSH access to developers

VPC Info
vpc-045c156ccb515dc2b

Inbound rules [Info](#)

This security group has no inbound rules.

[Add rule](#)

Outbound rules [Info](#)

This security group has no outbound rules.

The screenshot shows the AWS CloudShell interface with two browser windows open:

- Left Window:** Displays the "Create security group" page for a VPC. It shows the "Inbound rules" section with one rule: "HTTP" on port "80" from "Anywhere" (0.0.0.0/0) to "0.0.0.0/0". A warning message states: "⚠ Rules with source of 0.0.0.0/0 or ::/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only." The "Outbound rules" section shows "All traffic" to "Custom" (0.0.0.0/0).
- Right Window:** Displays the "Step 1 Create target group" page. It shows the "Specify group details" section with the note: "Your load balancer routes requests to the targets in a target group and performs health checks on the targets." Under "Basic configuration", it says: "Settings in this section can't be changed after the target group is created." The "Choose a target type" section has "Instances" selected, with the following benefits:
 - Supports load balancing to instances within a specific VPC.
 - Facilitates the use of [Amazon EC2 Auto Scaling](#) to manage and scale your EC2 capacity.Other options shown but not selected are "IP addresses", "Lambda function", and "Application Load Balancer".

Screenshot of the AWS CloudShell interface showing the creation of a target group named "Inventory-App".

EC2 > Target groups > Create target group

Override

Healthy threshold: The number of consecutive health checks successes required before considering an unhealthy target healthy. 2-10

Unhealthy threshold: The number of consecutive health check failures required before considering a target unhealthy. 2-10

Timeout: The amount of time, in seconds, during which no response means a failed health check. 10 seconds 2-120

Interval: The approximate amount of time between health checks of an individual target. 30 seconds 5-300

Success codes: The HTTP codes to use when checking for a successful response from a target. You can specify multiple values (for example, "200,202") or a range of values (for example, "200-299").

Target group details

Inventory-App successfully created.

Details

Target type	Protocol : Port	Protocol version	VPC
Instance	HTTP: 80	HTTP1	vpc-0512bf8f28397359a
IP address type	Load balancer None associated		
IPv4			

Total targets	Healthy	Unhealthy	Unused	Initial	Draining
0	0	0	0	0	0
	0 Anomalous				

Targets | Monitoring | Health checks | Attributes | Tags

Registered targets (0)

Anomaly mitigation: Not applicable

Deregister | Register targets

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences ENG IN 03:45 23-04-2025

Listeners and routing Info

A listener is a process that checks for connection requests using the port and protocol you configure. The rules that you define for a listener determine how the load balancer routes requests to its registered targets.

Listener HTTP:80

Protocol	Port
HTTP	: 80 1-65535

Default action Info

Forward to	Inventory-App Target type: Instance, IPv4
HTTP	(Edit)

Listener tags - optional
Consider adding tags to your listener. Tags enable you to categorize your AWS resources so you can more easily manage them.

Add listener tag
You can add up to 50 more tags.

Add listener

CloudShell Feedback

Successfully created load balancer: Inventory-LB
It might take a few minutes for your load balancer to fully set up and route traffic. Targets will also take a few minutes to complete the registration process and pass initial health checks.

Application Load Balancers now support public IPv4 IP Address Management (IPAM)
You can get started with this feature by configuring IP pools in the Network mapping section.

Inventory-LB

Details

Load balancer type	Status	VPC	Load balancer IP address type
Application	Provisioning	vpc-0512bf8f28397359a	IPv4
Scheme	Hosted zone	Availability Zones	Date created
Internet-facing	Z355XD0TRQ7X7K	subnet-07302aeedd52135d70 us-east-1b (use1-az2) subnet-024e6557f414d926 us-east-1a (use1-az2)	April 23, 2025, 03:16 (UTC+10:00)
Load balancer ARN	DNS name		
arn:aws:elasticloadbalancing:us-east-1:558625058626:loadbalancer/app/Inventory-LB/fd9cff4ebd210138	Inventory-LB-971235181.us-east-1.elb.amazonaws.com (A Record)		

Task 3: Creating an Auto Scaling Group

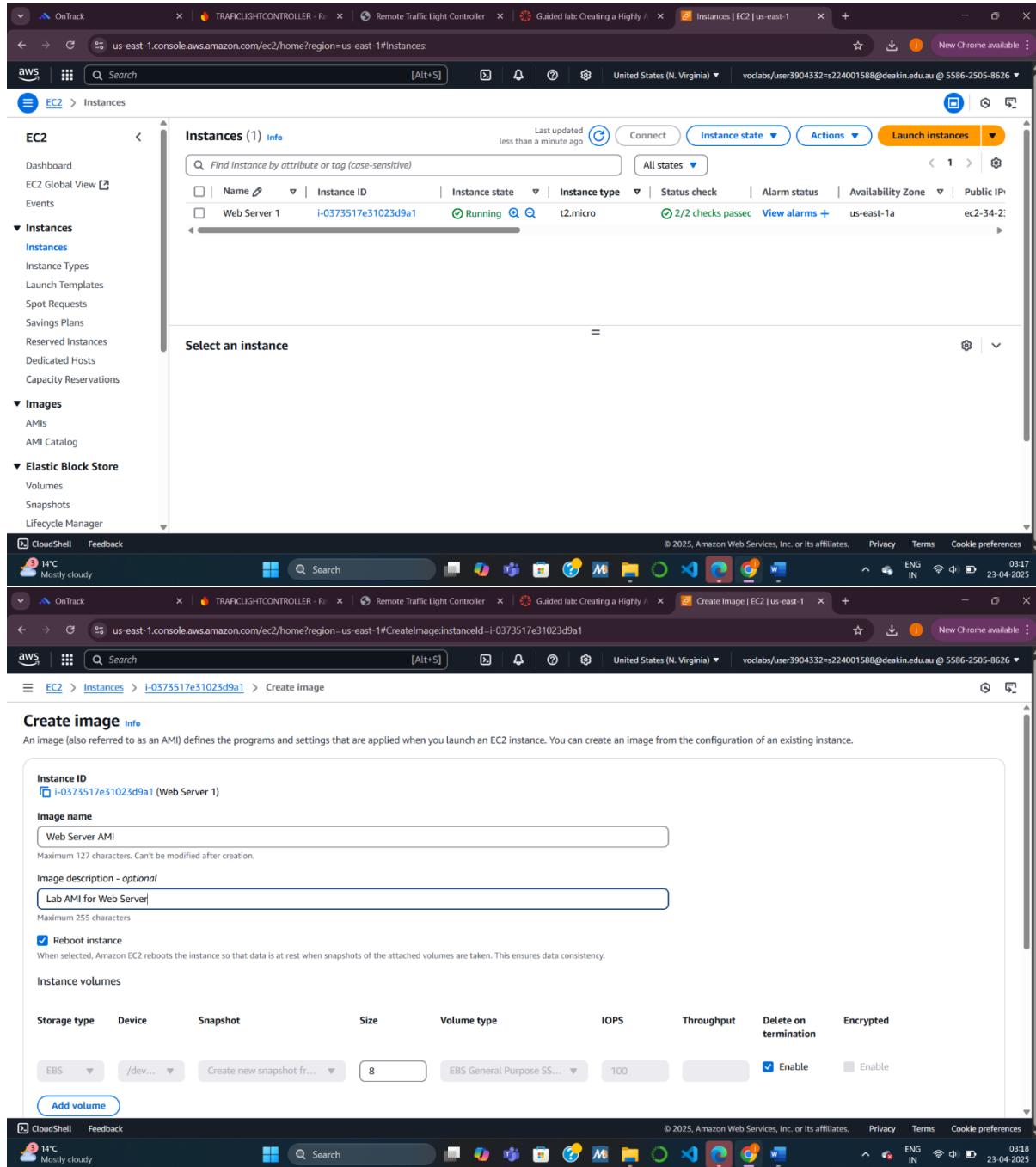
Task 3.1: Creating an AMI for Auto Scaling

I created an Amazon Machine Image (AMI) from the existing Web Server 1. This AMI captured the configuration of the instance, allowing me to launch identical instances through the Auto Scaling group later.

Task 3.2: Creating a Launch Template and an Auto Scaling Group

Using the AMI, I created a Launch Template (Inventory-LT) specifying details like instance type, key pair, IAM role, and user data to automate server setup. Then, I created an Auto Scaling Group (Inventory-ASG) configured to maintain two EC2 instances across two private subnets. The group was attached to the previously created Target Group (Inventory-App). This ensured that instances were

automatically launched, monitored, and replaced in case of failure, supporting continuous application availability.



The screenshot shows the AWS CloudWatch Metrics Insights interface. A query is being run against the 'Latency' metric. The results are displayed in a table with columns for 'Time' and 'Latency'. The data shows a significant increase in latency starting around 2023-04-23T12:00:00Z, peaking at approximately 1000 ms around 2023-04-23T12:30:00Z, and then gradually decreasing. The 'Time' column shows the timestamp of each data point, and the 'Latency' column shows the measured latency in milliseconds.

Time	Latency
2023-04-23T12:00:00Z	~100 ms
2023-04-23T12:15:00Z	~200 ms
2023-04-23T12:30:00Z	~1000 ms
2023-04-23T12:45:00Z	~500 ms
2023-04-23T13:00:00Z	~300 ms
2023-04-23T13:15:00Z	~200 ms
2023-04-23T13:30:00Z	~150 ms
2023-04-23T13:45:00Z	~100 ms

The screenshot displays two side-by-side views of the AWS EC2 console.

Left View: Instances

- EC2** menu:
 - Dashboard
 - EC2 Global View
 - Events
 - Instances**
 - Instances
 - Instance Types
 - Launch Templates
 - Spot Requests
 - Savings Plans
 - Reserved Instances
 - Dedicated Hosts
 - Capacity Reservations
 - Images
 - AMIs
 - AMI Catalog
- Instances (1/1) Info** table:

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IP
Web Server 1	i-0373517e31023d9a1	Running	t2.micro	2/2 checks passed	View alarms +	us-east-1a	ec2-34-2-
- i-0373517e31023d9a1 (Web Server 1)** instance details:
 - Details** tab selected.
 - Instance summary** table:

Instance ID	Public IPv4 address	Private IPv4 addresses
i-0373517e31023d9a1	34.238.154.119 open address	10.0.0.58
 - IPv6 address**: -
 - Instance state**: Running
 - Public IPv4 DNS**: ec2-34-238-154-119.compute-1.amazonaws.com | open address

Right View: Launch Templates

- EC2** menu:
 - Dashboard
 - EC2 Global View
 - Events
 - Launch Templates**
 - Spot Requests
 - Savings Plans
 - Reserved Instances
 - Dedicated Hosts
 - Capacity Reservations
 - Images
 - AMIs
 - AMI Catalog
- Compute** section:

EC2 launch templates

Streamline, simplify and standardize instance launches

Use launch templates to automate instance launches, simplify permission policies, and enforce best practices across your organization. Save launch parameters in a template that can be used for on-demand launches and with managed services, including EC2 Auto Scaling and EC2 Fleet. Easily update your launch parameters by creating a new launch template version.

New launch template button

Benefits and features

 - Streamline provisioning**: Minimize steps to provision instances. With EC2 Auto Scaling, updates to a launch template can be automatically passed to an Auto Scaling group. [Learn more](#)
 - Simplify permissions**: Create shorter, easier to manage IAM policies. [Learn more](#)
 - Documentation**:
 - [Documentation](#)
 - [API reference](#)

Screenshot of the AWS Cloud console showing the 'Create launch template' wizard.

Template version description: A prod webserver for MyApp

Auto Scaling guidance: Select this if you intend to use this template with EC2 Auto Scaling. Provide guidance to help me set up a template that I can use with EC2 Auto Scaling

Template tags:

Source template:

Launch template contents: Application and OS Images (Amazon Machine Image) - required

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below.

Free tier: In your first year of opening an AWS account, you get 750 hours per month of t2.micro instance usage (or t3.micro where t2.micro isn't available) when used with free tier AMIs, 750 hours per month of public IPv4 address usage, 30 GiB of EBS storage, 2 million I/Os, 1 GB of snapshots, and 100 GB of bandwidth to the internet.

Description: Lab AMI for Web Server

Architecture: x86_64 **AMI ID:** ami-0ac20700bb3170372

Instance type: t2.micro (Free tier eligible)

Key pair (login): Don't include in launch template

Summary:

- Software Image (AMI): Lab AMI for Web Server
- Virtual server type (instance type): t2.micro
- Firewall (security group):
- Storage (volumes): 1 volume(s) - 8 GiB

Create launch template

Screenshot of the AWS Cloud Console showing the 'Create launch template' wizard.

The page title is 'Create launch template' under 'Launch templates' in the EC2 service.

Network settings:

- Subnet:** Don't include in launch template (selected).
- Create new subnet:** Create new subnet.
- Firewall (security groups):** Select existing security group (Inventory-App selected).
- Security groups:** Select security groups (Inventory-App sg-0b6a32a5892c433b1 selected).

Advanced network configuration: Advanced network configuration is collapsed.

Storage (volumes):

- EBS Volumes:** Hide details.

Summary:

- Software Image (AMI):** Lab AMI for Web Server ami-0ac20700bb3170372
- Virtual server type (instance type):** t2.micro
- Firewall (security group):** Inventory-App
- Storage (volumes):** 1 volume(s) - 8 GiB

A tooltip for the Free tier is displayed: "Free tier: In your first year of opening an AWS account, you get 750 hours per month of t2.micro instance usage (or t3.micro where t2.micro isn't available) when used with free tier AMIs, 750 hours per month of public IPv4 address usage, 30 GiB of EBS storage, 2 million I/Os, 1 GB of snapshots, and 100 GB of bandwidth to the internet."

At the bottom right are 'Cancel' and 'Create Launch template' buttons.

The browser status bar shows: CloudShell Feedback, 14°C Mostly cloudy, © 2025, Amazon Web Services, Inc. or its affiliates., Privacy, Terms, Cookie preferences, ENG IN, 03:21, 23-04-2025.

The screenshot shows the AWS CloudShell interface with a browser window titled "Create launch template". The user is configuring a new launch template with the following settings:

- Software Image (AMI):** Lab AMI for Web Server (ami-0ac20700bb3170372)
- Virtual server type (instance type):** t2.micro
- Firewall (security group):** Inventory-App
- Storage (volumes):** 1 volume(s) - 8 GiB

A tooltip provides information about the Free tier: "In your first year of opening an AWS account, you get 750 hours per month of t2.micro instance usage (or t3.micro where t2.micro isn't available) when used with free tier AMIs, 750 hours per month of public IPv4 address usage, 30 GiB of EBS storage, 2 million I/Os, 1 GB of snapshots, and 100 GB of bandwidth to the internet."

The user data section contains the following script:

```
#!/bin/bash
# Install Apache Web Server and PHP
yum install -y httpd mysql
amazon-linux-extras install -y php7.2
# Download Lab files
wget https://aws-te-largeobjects.s3.us-west-2.amazonaws.com/CUR-TF-200-ACACAD-3-115230/12-lab-mod10-guided-Scaling/s3/scripts/inventory-app.zip
unzip inventory-app.zip -d /var/www/html/
# Download and install the AWS SDK for PHP
wget https://github.com/aws/aws-sdk-php/releases/download/3.62.3/aws.zip
unzip aws -d /var/www/html
# Turn on web server
chkconfig httpd on
service httpd start
```

User data has already been base64 encoded

Create Launch template

The screenshot shows the AWS CloudShell interface with a browser window titled "Create launch template". A green success message box appears, stating: "Success Successfully created Inventory-LT(lt-0d3ddcbdf4d5df488)."

Actions log

Next Steps

Launch an instance
With On-Demand Instances, you pay for compute capacity by the second (for Linux, with a minimum of 60 seconds) or by the hour (for all other operating systems) with no long-term commitments or upfront payments. Launch an On-Demand instance from your launch template.

Launch instance from this template

Create an Auto Scaling group from your template
Amazon EC2 Auto Scaling helps you maintain application availability and allows you to scale your Amazon EC2 capacity up or down automatically according to conditions you define. You can use Auto Scaling to help ensure that you are running your desired number of Amazon EC2 instances during demand spikes to maintain performance and decrease capacity during lulls to reduce costs.

Create Auto Scaling group

Create Spot Fleet
A Spot Instance is an unused EC2 instance that is available for less than the On-Demand price. Because Spot Instances enable you to request unused EC2 instances at steep discounts, you can lower your Amazon EC2 costs significantly. The hourly price for a Spot Instance (of each instance type in each Availability Zone) is set by Amazon EC2, and adjusted gradually based on the long-term supply of and demand for Spot Instances. Spot instances are well-suited for data-analysis, batch jobs, background processing, and optional tasks.

Create Spot Fleet

CloudShell Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences ENG IN 03:23 23-04-2025

Inventory-LT (lt-0d3ddcbdf4d5df488)

Launch template details

Launch template ID lt-0d3ddcbdf4d5df488	Launch template name Inventory-LT	Default version 1
--	--------------------------------------	----------------------

Actions ▾ **Delete template**

- Launch instance from template
- Modify template (Create new version)
- Delete template version
- Set default version
- Manage tags
- Create Spot Fleet
- Create Auto Scaling group

Launch template version details

Version 1 (Default)	Description	Date created 2025-04-22T17:23:33.000Z	Created by arn:aws:sts::558625058626:assumed-role/voclabs/user3904332=s224001588@deakin.edu.au
Instance details		Storage	Resource tags
AMI ID ami-0ac20700bb3170372	Instance type t2.micro	Availability Zone	Key pair name vockey
Security groups		Security group IDs sg-0b6a32a5892c435b1	

Actions ▾ **Delete template version**

Choose launch template Info

Specify a launch template that contains settings common to all EC2 instances that are launched by this Auto Scaling group.

Name

Auto Scaling group name
Enter a name to identify the group.
Inventory-ASG

Must be unique to this account in the current Region and no more than 255 characters.

Launch template Info

For accounts created after May 31, 2023, the EC2 console only supports creating Auto Scaling groups with launch templates. Creating Auto Scaling groups with launch configurations is not recommended but still available via the CLI and API until December 31, 2023.

Launch template
Choose a launch template that contains the instance-level settings, such as the Amazon Machine Image (AMI), instance type, key pair, and security groups.
Inventory-LT

Create a launch template Info

Version

Default (1)

Create a launch template version

Network Info

For most applications, you can use multiple Availability Zones and let EC2 Auto Scaling balance your instances across the zones. The default VPC and default subnets are suitable for getting started quickly.

VPC
Choose the VPC that defines the virtual network for your Auto Scaling group.
 10.0.0.0/16 C

[Create a VPC](#) C

Availability Zones and subnets
Define which Availability Zones and subnets your Auto Scaling group can use in the chosen VPC.
 C

us-east-1a | subnet-071f6831265326dfs (Private Subnet 1) 10.0.2.0/23 X

us-east-1b | subnet-099b4641095c3048b (Private Subnet 2) 10.0.4.0/23 X

[Create a subnet](#) C

Availability Zone distribution - new
Auto Scaling automatically balances instances across Availability Zones. If launch failures occur in a zone, select a strategy.

Balanced best effort
If launches fail in one Availability Zone, Auto Scaling will attempt to launch in another healthy Availability Zone.

Balanced only
If launches fail in one Availability Zone, Auto Scaling will continue to attempt to launch in the unhealthy Availability Zone to preserve balanced distribution.

Load balancing Info

Use the options below to attach your Auto Scaling group to an existing load balancer, or to a new load balancer that you define.

No load balancer
Traffic to your Auto Scaling group will not be fronted by a load balancer.

Attach to an existing load balancer
Choose from your existing load balancers.

Attach to a new load balancer
Quickly create a basic load balancer to attach to your Auto Scaling group.

Attach to an existing load balancer
Select the load balancers that you want to attach to your Auto Scaling group.

Choose from your load balancer target groups
This option allows you to attach Application, Network, or Gateway Load Balancers.

Choose from Classic Load Balancers

Existing load balancer target groups
Only instance target groups that belong to the same VPC as your Auto Scaling group are available for selection.
 C

Inventory-App | HTTP Application Load Balancer: Inventory-LB X

VPC Lattice integration options Info

To improve networking capabilities and scalability, integrate your Auto Scaling group with VPC Lattice. VPC Lattice facilitates communications between AWS services and helps you connect and manage your applications across compute services in AWS.

Select VPC Lattice service to attach

Screenshot of the AWS CloudWatch Metrics Metrics Insights page showing a log stream for a Lambda function named "HelloWorld" with the message "Hello, World!". The log stream is displayed in a table format with columns for timestamp, log group, log stream, and log message.

The "Metrics" tab is selected, showing a chart of CPU Utilization over time. The chart has two data series: "CPU Utilization" and "CPU Utilization (estimated)". The chart shows a sharp increase in CPU utilization starting around 2023-04-23T10:00:00Z, peaking at approximately 100% before returning to baseline levels.

The "Logs" tab is also visible, showing the raw log entries for the Lambda function.

Screenshot of the AWS CloudWatch Metrics Metrics Insights page showing a log stream for a Lambda function named "HelloWorld" with the message "Hello, World!". The log stream is displayed in a table format with columns for timestamp, log group, log stream, and log message.

The "Metrics" tab is selected, showing a chart of CPU Utilization over time. The chart has two data series: "CPU Utilization" and "CPU Utilization (estimated)". The chart shows a sharp increase in CPU utilization starting around 2023-04-23T10:00:00Z, peaking at approximately 100% before returning to baseline levels.

The "Logs" tab is also visible, showing the raw log entries for the Lambda function.

The screenshot displays the AWS CloudShell interface with two identical screenshots of the Auto Scaling groups page. Both screenshots show the following details:

- Auto Scaling groups (1) Info**
- Last updated: less than a minute ago
- Launch configurations: [Launch configurations](#)
- Launch templates: [Launch templates](#)
- Actions: [Actions](#)
- Create Auto Scaling group: [Create Auto Scaling group](#)
- Search your Auto Scaling groups:
- Table headers: Name, Launch template/configuration, Instances, Status, Desired capacity, Min, Max, Availability Zones.
- Table data:
 - Name: Inventory-ASG
 - Launch template/configuration: [Inventory-LT](#) | Version Default
 - Instances: 0
 - Status: Updating capacity...
 - Desired capacity: 2
 - Min: 2
 - Max: 2
 - Availability Zones: us-east-1a, us-east-1b

Both screenshots also show a message at the top: **Inventory-ASG created successfully. Group metrics collection is enabled.**

Task 4: Updating Security Groups

Task 4.1: Configuring the Load Balancer Security Group

I verified that the Load Balancer's Security Group (Inventory-LB) was set to allow HTTP and HTTPS traffic from external sources. This group facilitates client communication with the application.

Task 4.2: Configuring the Application Security Group

The Application Security Group (Inventory-App) was configured to accept incoming HTTP traffic only from the Load Balancer Security Group. This prevents direct access to the EC2 instances from the internet, enforcing a secure tiered architecture.

Task 4.3: Configuring the Database Security Group

I updated the Database Security Group (Inventory-DB) to only allow MySQL (port 3306) traffic from the Application Security Group. This enforced the principle of least privilege by restricting database access solely to the application layer.

The image consists of three vertically stacked screenshots of the AWS EC2 Security Groups 'Edit inbound rules' interface. Each screenshot shows a different security group configuration:

- Screenshot 1 (Top):** Shows a rule for port 80 (HTTP) from a load balancer (sg-00522d8c69448435c) to the instance. The rule has a description 'Traffic from load balancer'. Buttons for 'Add rule', 'Cancel', 'Preview changes', and 'Save rules' are visible.
- Screenshot 2 (Middle):** Shows a rule for port 3306 (MySQL/Aurora) from the IP range 10.0.0.0/16 to the instance. The rule has a description '10.0.0.0/16'. Buttons for 'Add rule', 'Cancel', 'Preview changes', and 'Save rules' are visible.
- Screenshot 3 (Bottom):** A blank screenshot showing the same interface but with no rules listed.

The screenshot displays the AWS CloudShell interface with two open browser tabs:

- Top Tab:** Shows the "Edit inbound rules" page for a security group. It lists a single rule: "sg-0b6a32a5892c433b1" (Traffic from application server) with "sg-0b6a32a5892c433b1" as the source. The rule type is MySQL/Aurora, protocol is TCP, and port range is 3306.
- Bottom Tab:** Shows the "Inventory-LB" load balancer details page. The page includes a sidebar with navigation links like EC2 Catalog, Elastic Block Store, Network & Security, Load Balancing, and Auto Scaling. The main content area displays the load balancer's configuration, including its ARN, status (Active), scheme (Internet-facing), VPC (vpc-0512bf8f28397359a), and availability zones (us-east-1a, us-east-1b). The DNS name is listed as "Inventory-LB-971235181.us-east-1.elb.amazonaws.com (A Record)".

Task 5: Testing the Application

In this task, I verified that the application was properly deployed and reachable through the Load Balancer's DNS. I confirmed that the Target Group registered two healthy EC2 instances and tested the Load Balancer by accessing the application's web page. Reloading the page showed load distribution between the two instances, proving correct setup and functionality.

Additionally, To simulate a failure scenario, I terminated one of the application's EC2 instances. The Load Balancer detected the failure and routed traffic to the remaining healthy instance without service interruption. The Auto Scaling Group automatically launched a replacement instance to maintain the desired instance count. This test validated the effectiveness of the high availability setup.

Screenshot of the AWS Management Console showing the EC2 Instances details for an instance named "i-0d3ae17d76541b3c4".

Instance summary for i-0d3ae17d76541b3c4 (Inventory-App)

Instance ID	i-0d3ae17d76541b3c4	Public IPv4 address	—
IPv6 address	—	Instance state	Terminated
Hostname type	—	Instance type	t2.micro
Answer private resource DNS name	—	VPC ID	—
Auto-assigned IP address	—	Subnet ID	—
IAM Role	—	Instance ARN	arn:aws:ec2:us-east-1:558625058626:instance/i-0d3ae17d76541b3c4
IMDSv2	Optional	Managed	false
EC2 recommends setting IMDSv2 to required Learn more		Elastic IP addresses	—
		Public IPv4 DNS	—
		AWS Compute Optimizer finding	Opt-in to AWS Compute Optimizer for recommendations Learn more
		Auto Scaling Group name	Inventory-ASG

Guided Lab: Creating a Highly Available Environment

Lab overview and objectives

Critical business systems should be deployed as highly available applications; that is, applications remain operational even when some components fail. To achieve high availability in Amazon Web Services (AWS), you run services across multiple Availability Zones.

Many AWS services are inherently highly available, such as load balancers. Many AWS services can also be configured for high availability, such as deploying Amazon Elastic Compute Cloud (Amazon EC2) instances in multiple Availability Zones.

In this lab, you start with an application that runs on a single EC2 instance and then make it highly available.

After completing this lab, you should be able to do the following:

- [Task 2A] Inventory-LB was created
- [Task 2B] Inventory-LB is an ALB
- [Task 2C] Inventory-LB is in Lab VP
- [Task 2D] ALB is using Public Subnet
- [Task 2E] ALB is using Public Subnet
- [Task 2F] ALB is using the correct Target Type
- [Task 2G] ALB is using correct Target Type
- [Task 3A] Inventory-LT Launch Template
- [Task 3B] Launch Template Instance
- [Task 3C] Launch Template Security Groups
- [Task 3D] Auto Scaling Group Size
- [Task 4A] Inventory-App Protocol

Guided Lab: Creating a Highly Available Environment

Lab overview and objectives

Critical business systems should be deployed as highly available applications; that is, applications remain operational even when some components fail. To achieve high availability in Amazon Web Services (AWS), you run services across multiple Availability Zones.

Many AWS services are inherently highly available, such as load balancers. Many AWS services can also be configured for high availability, such as deploying Amazon Elastic Compute Cloud (Amazon EC2) instances in multiple Availability Zones.

In this lab, you start with an application that runs on a single EC2 instance and then make it highly available.

After completing this lab, you should be able to do the following:

-LB is in Lab VPC	5/5
-ng Public Subnet 1	5/5
-ng Public Subnet 2	5/5
-ng the correct Security Group	0/5
-ng correct Target Group	5/5
-LT Launch Template Created	5/5
-mplate Instance Type	5/5
-mplate Security Group	5/5
-ing Group Size	5/5
-App Protocol Type	5/5
-App Source	5/5
-DB Group Rule	5/5
Server is accessible	0/5

WHAT WENT WRONG ?

Problem 1: Load Balancer Security Group is Incorrect

What Went Wrong:

- The Load Balancer (Inventory-LB) was **not assigned the correct Security Group**.
- The Security Group attached to the Load Balancer is **not named or configured as "Inventory-LB" Security Group** as expected.
- Security Groups act like virtual firewalls for your instances and Load Balancer — if these aren't configured correctly, the application won't be reachable.

How to Rectify :

1. Identify the Correct Security Group:

- Make sure there is a dedicated Security Group created for the Load Balancer named properly (e.g., Inventory-LB).
- This Security Group should allow **incoming HTTP/HTTPS traffic (TCP port 80 and/or 443)** from anywhere (0.0.0.0/0) or your allowed IP ranges.

2. Check Security Group Attachment:

- When creating or configuring the Load Balancer, explicitly **attach the correct Security Group** (Inventory-LB group).
- Verify this in the AWS Console under **EC2 → Load Balancers → Select your LB → Description → Security Groups**.

3. Double-Check Inbound and Outbound Rules:

- Inbound Rule:
 - Allow HTTP (Port 80) and/or HTTPS (Port 443).
- Outbound Rule:
 - Allow all outbound traffic (default is usually fine unless restricted).

Problem 2: Application is Not Reachable (Possible Subnet or Instance Issue)

What Went Wrong:

- Even though your Load Balancer is in **public subnets**, the system reports that **the application is not reachable**.
- The issue may relate to **subnet configuration, Security Group settings, or unhealthy target instances**.

How to Rectify :

1. Verify Public Subnet Settings:

- Ensure the subnets used by the Load Balancer are:
 - Marked as **public** (i.e., have a Route Table with a route to the **Internet Gateway**).
- Check subnet settings under **VPC → Subnets → Route Table → Routes**.

2. Check Auto Scaling Group and Target Group Health:

- Confirm that **two healthy instances** are running in the Auto Scaling Group.
- Check that the **Target Group health check path** is correct (e.g., /index.html or /health endpoint).
- Instances must respond successfully to the health check; otherwise, they will be marked as **unhealthy**.

3. Instance Security Group Rules:

- The Security Group attached to your EC2 instances (through Launch Template or Auto Scaling Group) **must allow inbound traffic from the Load Balancer Security Group** on the application port (usually TCP port 80).
- Example:
 - Inbound Rule:
 - Type: HTTP
 - Source: **Inventory-LB Security Group ID**

4. Confirm Web Server is Running:

- Ensure that the EC2 instances are actually **running the web server** (like Apache, Nginx, etc.) and listening on the correct port.

TO SUMMARIZE:

- * Ensure the correct Security Group is attached to the Load Balancer with proper inbound rules (HTTP/HTTPS).
- * Verify public subnets are properly routed through an Internet Gateway.
- * Confirm that the Auto Scaling Group is launching healthy instances.
- * Double-check that instances accept traffic from the Load Balancer's Security Group on the correct port.
- * Validate the web server is running and reachable on the target instances.
- * Always test instance health and Load Balancer reachability before submission.

Task Questions

- What is an AWS Region and Availability Zone? How are they related?
An AWS Region is a geographical area containing multiple isolated **Availability Zones (AZs)**.
 An AZ is one or more separate data centers with independent power, networking, and cooling within a Region. They are related because multiple AZs within a Region enable high availability and fault tolerance for applications.
- What are the different types of Elastic Load Balancer?
 The three types of Elastic Load Balancers (ELB) in AWS are:
Application Load Balancer (ALB): Used for HTTP/HTTPS traffic with Layer 7 features like path-based and host-based routing.
Network Load Balancer (NLB): Designed for TCP/UDP traffic, providing high performance and ultra-low latency at Layer 4.
Gateway Load Balancer (GWLB): Used for deploying and managing third-party virtual appliances like firewalls, operating at Layer 3.
- When should you use Application Load Balancer?
 Use **Application Load Balancer** when your application requires:
Advanced request routing (based on URL path or hostname).
HTTP/HTTPS traffic handling with SSL/TLS termination.
WebSocket support and routing to containers (ECS/EKS).
Sticky sessions or need for Layer 7 features like redirects or fixed responses.
 It is ideal for **web applications** where content-based routing is needed.
- How do you make an RDS highly available?
 To make Amazon RDS highly available, use **Multi-AZ deployments**. This automatically creates a standby replica in a different Availability Zone and enables failover in case of instance failure, ensuring continuous database operation.

KNOWLEDGE CHECK :

OnTrack TRAFFICLIGHTCONTROLLER - Re Module 10 Knowledge Check ChatGPT - jass cloud computin... New Chrome available

awsacademy.instructure.com/courses/104153/assignments/114753?module_item_id=9702743

ACAv3EN-US-LTI13-104153 > Assignments > Module 10 Knowledge Check

Home Modules Discussions Grades Lucid (Whiteboard)

Calendar Inbox History Help

Module 10 Knowledge Check

Due No Due Date Points 100 Submitting an external tool

KEYBOARD NAVIGATION

Knowledge check results

Your score: 100% (100 points)
Required score: 70% (70 points)

Result: Congratulations! You have completed this knowledge check.
To continue, choose Next in the lower-right corner.

aws © 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Tomorrow's high To break record

Search ENG IN 23-04-2025

