

# Assignment 1

Vedant Marwadi<sup>1</sup>, Alyssa Pausanos<sup>2</sup>, Cong Tran Quoc<sup>3</sup>, Kirushnaamoni  
Ramakrishnan<sup>4</sup>, and Lin Aung Thet<sup>5</sup>

<sup>1</sup>ID: 23208466

<sup>2</sup>ID: 18021379

<sup>3</sup>ID: 19078245

<sup>4</sup>ID: 23195283

<sup>5</sup>ID: 22175616

## 1. Introduction

A start-up education software company has a software application that serves as a crucial platform for knowledge sharing, managing student information, and facilitating learning through the sharing of documents, books, and videos. In response to the evolving needs of the education software company and the growing demand for scalable solutions, the decision was made to migrate the existing on-premises IT infrastructure to a cloud-based architecture.

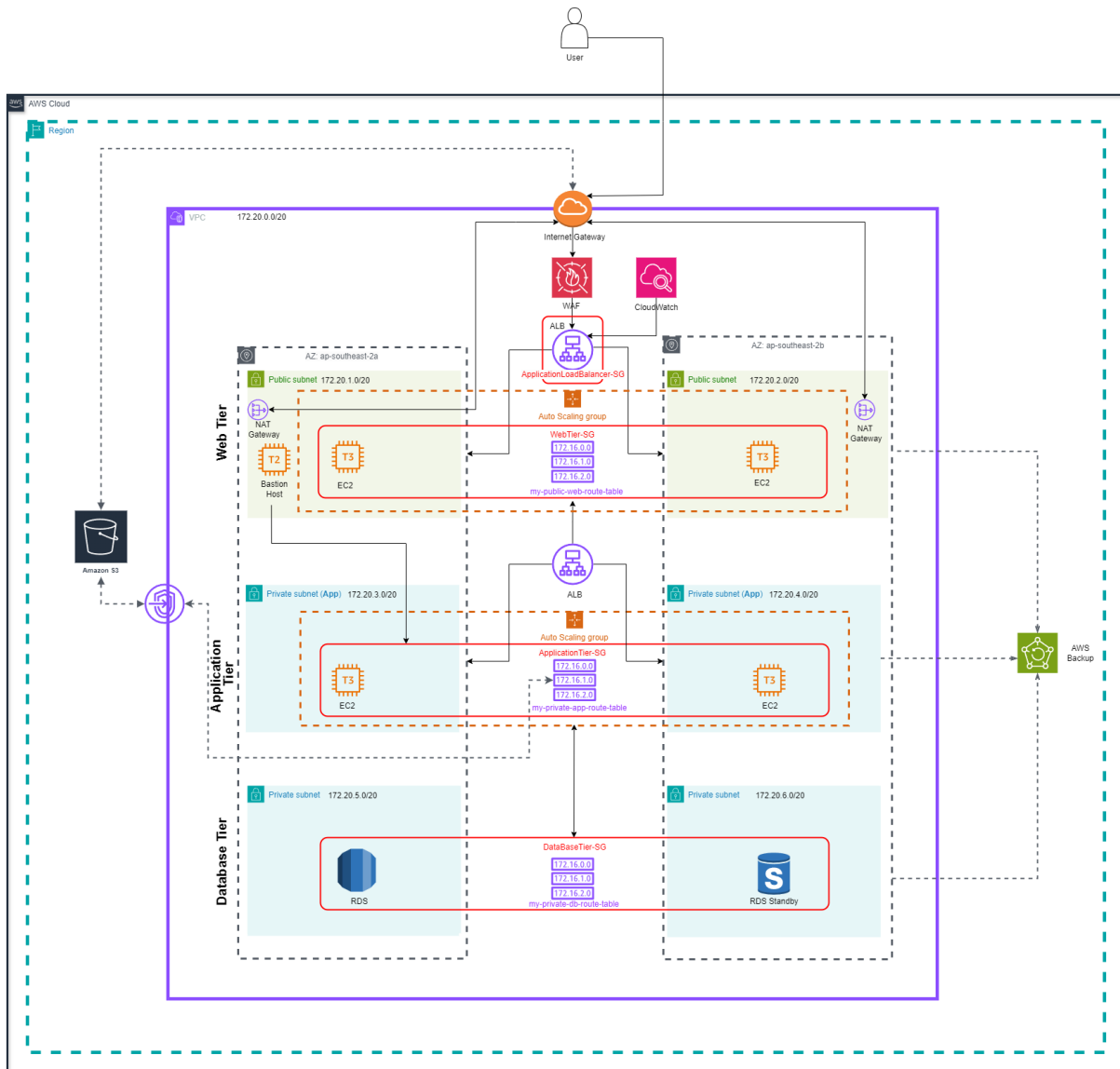
To address these requirements, a Three-Tier architecture solution is designed, leveraging Amazon Web Services. This architecture encompasses distinct tiers, assuring a modular and flexible architecture that can adapt to fluctuating demands and accommodate future growth seamlessly.

In this report, a detailed information of the architecture solution is provided, outlining all the components that make up each tier. In addition to this, a cost analysis is conducted to evaluate the financial implications of the migration to the cloud. Additionally, the report also assess how the solution aligns with key principles of elasticity, security, reliability, and cost optimization without over-provisioning resources.

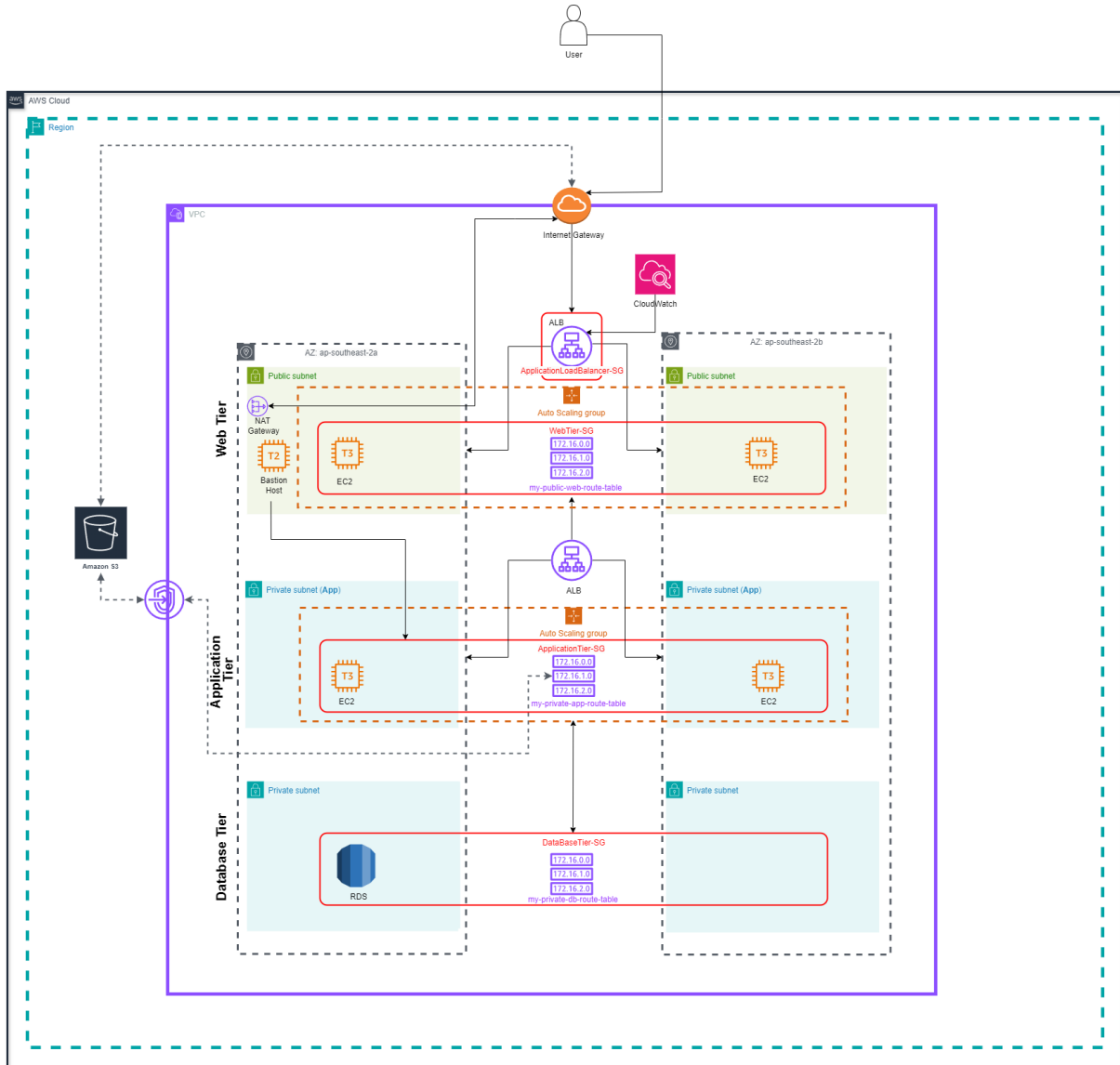
The goal of this report is to outline all the steps taken to embrace the cloud technology and to implement a robust architecture tailored to meet the specific needs of the education software company.

## 2. Architecture Solution

### 2.1 Production Environment



## 2.2 Test Environment



## 3. Detailed Solution for Each Component

### 3.1 Network

A Virtual Private Cloud is established within the Sydney region (ap-southeast-2) of AWS because the main users of this application are in Australia and New Zealand. The chosen CIDR block for the VPC is 172.20.0.0/20. This CIDR block allows for a maximum of 4096 IP addresses and is configured to provide a substantial address space to accommodate the present and future resource requirements of the business.

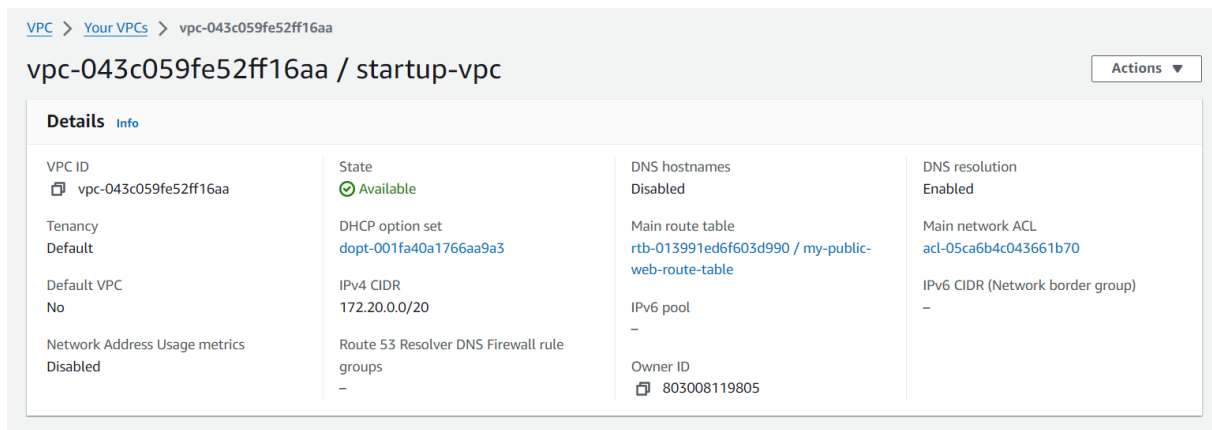


Figure 1: VPC

To control traffic flow within the VPC and isolate resources, two public subnets are established for the Web Tier. The application tier consists of two private subnets. These subnets house the application logic, guaranteeing their isolation from direct internet access. Similar to the application tier, the database tier also comprises two private subnets. These subnets are dedicated to hosting database instances. All these subnets are spread across two Availability Zones to establish redundancy and resilience against potential failures in any single zone.

Name	Subnet ID	State	VPC	IPv4 CIDR
my-private-app-subnet-2	subnet-04d0a30e2e92a5535	Available	vpc-043c059fe52ff16aa   startu...	172.20.4.0/24
my-private-db-subnet-1	subnet-0d847d6e5ea23a5b3	Available	vpc-043c059fe52ff16aa   startu...	172.20.5.0/24
my-public-web-subnet-2	subnet-03d05db6c08049600	Available	vpc-043c059fe52ff16aa   startu...	172.20.2.0/24
my-public-web-subnet-1	subnet-035f65eeede51925b	Available	vpc-043c059fe52ff16aa   startu...	172.20.1.0/24
my-private-db-subnet-2	subnet-0246a4c6b5aeaf92	Available	vpc-043c059fe52ff16aa   startu...	172.20.6.0/24
my-private-app-subnet-1	subnet-058f5e896741fe632	Available	vpc-043c059fe52ff16aa   startu...	172.20.3.0/24

Figure 2: Subnets

Furthermore, route tables are created for Web Tier, Application Tier, and Database Tier to segregate traffic based on its destination and purpose. The route table for the web tier is named *my-public-web-route-table* and is connected to the Internet Gateway. An IGW named *my-interent-gateway* is created.

VPC > Route tables > rtb-013991ed6f603d990

## rtb-013991ed6f603d990 / my-public-web-route-table Actions ▼

**Details** Info

Route table ID rtb-013991ed6f603d990	Main Yes	Explicit subnet associations <a href="#">2 subnets</a>	Edge associations -
VPC <a href="#">vpc-043c059fe52ff16aa</a>   startup-vpc	Owner ID 803008119805		

[Routes](#)
[Subnet associations](#)
[Edge associations](#)
[Route propagation](#)
[Tags](#)

**Routes (2)**
Both ▼
Edit routes

< 1 > ⚙

Destination ▼	Target ▼	Status ▼	Propagated ▼
0.0.0.0/0	<a href="#">igw-08d48a1a2dd528622</a>	Active	No
172.20.0.0/20	local	Active	No

Figure 3: Route Table For Web Tier

VPC > Internet gateways > igw-08d48a1a2dd528622

## igw-08d48a1a2dd528622 / my-internet-gateway Actions ▼

**Details** Info

Internet gateway ID igw-08d48a1a2dd528622	State Attached	VPC ID <a href="#">vpc-043c059fe52ff16aa</a>   startup-vpc	Owner 803008119805
--	-------------------	---	-----------------------

**Tags**
Manage tags

< 1 > ⚙

Key	Value
Name	my-internet-gateway

Figure 4: Internet Gateway

The route table for the Database Tier is named *my-private-db-route-table* and is configured to route traffic through a NAT Gateway. By directing traffic through a NAT Gateway, instances within the database tier can securely access resources outside the VPC, such as updates or patches from external repositories, without exposing their private IP addresses to the internet.

VPC > Route tables > rtb-0849afff243d73918															
rtb-0849afff243d73918 / my-private-db-route-table															
<div>Details Info</div> <table> <tr> <td>Route table ID rtb-0849afff243d73918</td><td>Main No</td><td>Explicit subnet associations 2 subnets</td><td>Edge associations -</td></tr> <tr> <td>VPC vpc-043c059fe52ff16aa   startup-vpc</td><td>Owner ID 803008119805</td><td></td><td></td></tr> </table>				Route table ID rtb-0849afff243d73918	Main No	Explicit subnet associations 2 subnets	Edge associations -	VPC vpc-043c059fe52ff16aa   startup-vpc	Owner ID 803008119805						
Route table ID rtb-0849afff243d73918	Main No	Explicit subnet associations 2 subnets	Edge associations -												
VPC vpc-043c059fe52ff16aa   startup-vpc	Owner ID 803008119805														
<div>Routes Subnet associations Edge associations Route propagation Tags</div> <div> <div>Routes (2)</div> <div> <input type="text" value="Filter routes"/> <div>Both Edit routes</div> <div>&lt; 1 &gt; ⚙</div> <table> <tr> <th>Destination</th><th>Target</th><th>Status</th><th>Propagated</th></tr> <tr> <td>0.0.0.0/0</td><td>nat-022457a3939092a1d</td><td>Active</td><td>No</td></tr> <tr> <td>172.20.0.0/20</td><td>local</td><td>Active</td><td>No</td></tr> </table> </div> </div>				Destination	Target	Status	Propagated	0.0.0.0/0	nat-022457a3939092a1d	Active	No	172.20.0.0/20	local	Active	No
Destination	Target	Status	Propagated												
0.0.0.0/0	nat-022457a3939092a1d	Active	No												
172.20.0.0/20	local	Active	No												

Figure 5: Route Table For Database Tier

Two NAT Gateways, named *my-nat-gateway-1* and *my-nat-gateway-2*, are created and nestled inside the two public subnets.

VPC > NAT gateways > nat-0ed800f9435091c7f															
nat-0ed800f9435091c7f / my-nat-gateway-1															
<div>Details</div> <table> <tr> <td>NAT gateway ID nat-0ed800f9435091c7f</td><td>Connectivity type Public</td><td>State Available</td><td>State message -</td></tr> <tr> <td>NAT gateway ARN arn:aws:ec2:ap-southeast-2:803008119805:natgateway/nat-0ed800f9435091c7f</td><td>Primary public IPv4 address 3.105.115.251</td><td>Primary private IPv4 address 172.20.1.128</td><td>Primary network interface ID eni-046f5f65f7929bfa9</td></tr> <tr> <td>VPC vpc-043c059fe52ff16aa   startup-vpc</td><td>Subnet subnet-035f65eede51925b / my-public-web-subnet-1</td><td>Created Thursday, April 18, 2024 at 10:13:03 GMT+12</td><td>Deleted -</td></tr> </table>				NAT gateway ID nat-0ed800f9435091c7f	Connectivity type Public	State Available	State message -	NAT gateway ARN arn:aws:ec2:ap-southeast-2:803008119805:natgateway/nat-0ed800f9435091c7f	Primary public IPv4 address 3.105.115.251	Primary private IPv4 address 172.20.1.128	Primary network interface ID eni-046f5f65f7929bfa9	VPC vpc-043c059fe52ff16aa   startup-vpc	Subnet subnet-035f65eede51925b / my-public-web-subnet-1	Created Thursday, April 18, 2024 at 10:13:03 GMT+12	Deleted -
NAT gateway ID nat-0ed800f9435091c7f	Connectivity type Public	State Available	State message -												
NAT gateway ARN arn:aws:ec2:ap-southeast-2:803008119805:natgateway/nat-0ed800f9435091c7f	Primary public IPv4 address 3.105.115.251	Primary private IPv4 address 172.20.1.128	Primary network interface ID eni-046f5f65f7929bfa9												
VPC vpc-043c059fe52ff16aa   startup-vpc	Subnet subnet-035f65eede51925b / my-public-web-subnet-1	Created Thursday, April 18, 2024 at 10:13:03 GMT+12	Deleted -												

Figure 6: NAT Gateway 1

VPC > NAT gateways > nat-022457a3939092a1d															
nat-022457a3939092a1d / my-nat-gateway-2															
<div>Details</div> <table> <tr> <td>NAT gateway ID nat-022457a3939092a1d</td><td>Connectivity type Public</td><td>State Available</td><td>State message -</td></tr> <tr> <td>NAT gateway ARN arn:aws:ec2:ap-southeast-2:803008119805:natgateway/nat-022457a3939092a1d</td><td>Primary public IPv4 address 13.211.218.53</td><td>Primary private IPv4 address 172.20.2.221</td><td>Primary network interface ID eni-081070dbc6c8a6837</td></tr> <tr> <td>VPC vpc-043c059fe52ff16aa   startup-vpc</td><td>Subnet subnet-03d05db6c08049600 / my-public-web-subnet-2</td><td>Created Thursday, April 18, 2024 at 10:13:52 GMT+12</td><td>Deleted -</td></tr> </table>				NAT gateway ID nat-022457a3939092a1d	Connectivity type Public	State Available	State message -	NAT gateway ARN arn:aws:ec2:ap-southeast-2:803008119805:natgateway/nat-022457a3939092a1d	Primary public IPv4 address 13.211.218.53	Primary private IPv4 address 172.20.2.221	Primary network interface ID eni-081070dbc6c8a6837	VPC vpc-043c059fe52ff16aa   startup-vpc	Subnet subnet-03d05db6c08049600 / my-public-web-subnet-2	Created Thursday, April 18, 2024 at 10:13:52 GMT+12	Deleted -
NAT gateway ID nat-022457a3939092a1d	Connectivity type Public	State Available	State message -												
NAT gateway ARN arn:aws:ec2:ap-southeast-2:803008119805:natgateway/nat-022457a3939092a1d	Primary public IPv4 address 13.211.218.53	Primary private IPv4 address 172.20.2.221	Primary network interface ID eni-081070dbc6c8a6837												
VPC vpc-043c059fe52ff16aa   startup-vpc	Subnet subnet-03d05db6c08049600 / my-public-web-subnet-2	Created Thursday, April 18, 2024 at 10:13:52 GMT+12	Deleted -												

Figure 7: NAT Gateway 2

The route table for the Application Tier is named *my-private-app-route-table*. Similar to the database tier, it is connected to a NAT Gateway. Additionally, it is linked to a VPC Endpoint for Amazon S3. This configuration certifies that instances within the

application tier can securely access S3 buckets without traversing the internet, enhancing security and reducing latency.

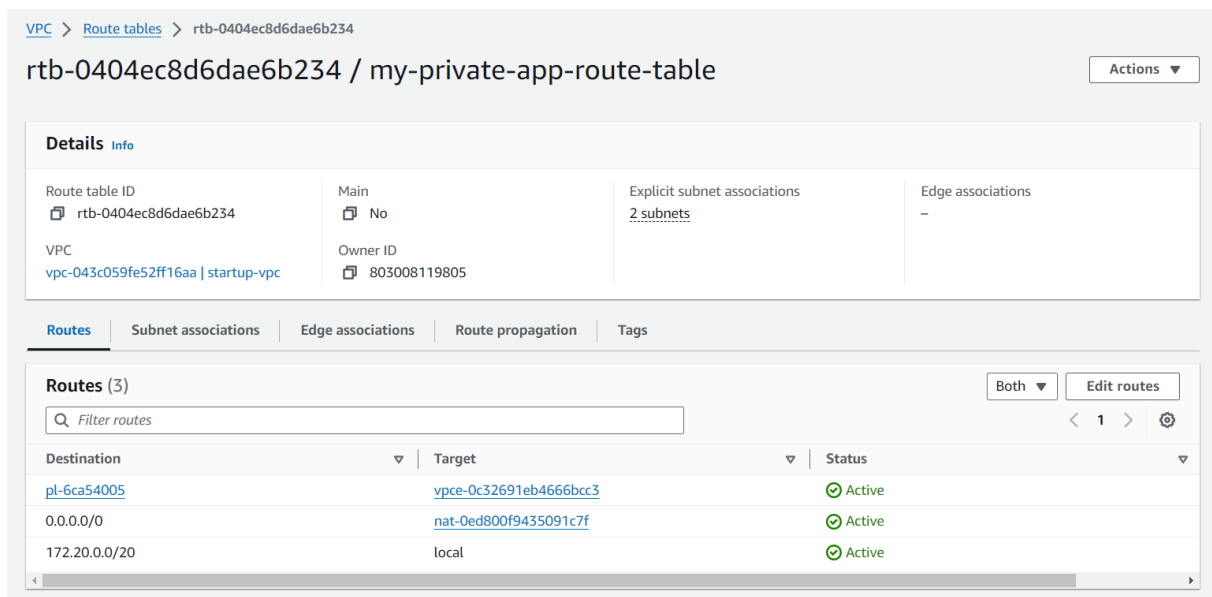


Figure 8: Route Table For Application Tier

In the testing environment, all aspects mirror the production setup, including the establishment of a VPC in the Sydney region, subnet distribution across two Availability Zones, and route table configurations for each tier. However, to economize, only one NAT Gateway is deployed instead of two to assure secure access to external resources while minimizing costs. This setup meets current needs and supports future growth.

## 3.2 Web Tier

Firstly, the web tier of the three-tier architecture leverages an Application Load Balancer named *WebTier-ALB* for high availability and scalability. It is configured as internet-facing to allow external users to access the web application. It operates within the public subnets: "my-public-web-subnet-1" and "my-public-web-subnet-2" and spans across two Availability Zones: "ap-southeast-2a" and "ap-southeast-2b". It also ensures that only healthy instances receive traffic.

### Summary

Review and confirm your configurations. [Estimate cost](#)

Basic configuration [Edit](#)

WebTier - ALB

- Internet-facing
- IPv4

Security groups [Edit](#)

- ApplicationLoadBalancer - SG [sg-04fac140bb5b31258](#)

Network mapping [Edit](#)

VPC [vpc-043c059fe52ff16aa](#)  
startup-vpc

- ap-southeast-2a [subnet-035f65eede51925b](#)  
my-public-web-subnet-1
- ap-southeast-2b [subnet-03d05db6c08049600](#)  
my-public-web-subnet-2

Listeners and routing [Edit](#)

- HTTP:80 defaults to [WebTier-TargetGroup](#)

Service integrations [Edit](#)

AWS WAF: None  
AWS Global Accelerator: None

Tags [Edit](#)

None

Figure 9: Web Tier - Application Load Balancer

Secondly, to warrant robust security, a security group named *ApplicationLoadBalancer-SG* is established for *WebTier-ALB*, which allows SSH access solely from the IP address of the designated device and allows HTTP and HTTPS traffic from anywhere.

[EC2](#) > [Security Groups](#) > sg-04fac140bb5b31258 - ApplicationLoadBalancer - SG

sg-04fac140bb5b31258 - ApplicationLoadBalancer - SG

Actions

Details

Security group name  
ApplicationLoadBalancer - SG

Security group ID  
sg-04fac140bb5b31258

Description  
Security Group for ApplicationLoadBalancer

VPC ID  
[vpc-043c059fe52ff16aa](#)

Owner  
803008119805

Inbound rules count  
3 Permission entries

Outbound rules count  
1 Permission entry

Inbound rules

Outbound rules

Tags

Inbound rules (3)

< 1 >

⊗

<input type="checkbox"/>	Name	Security group rule...	IP version	Type	Protocol	Port range	Source	Description
<input type="checkbox"/>	-	sgr-00bae048c55a810f8	IPv4	SSH	TCP	22	114.23.218.180/32	-
<input type="checkbox"/>	-	sgr-0e5fb9cd250c4e895	IPv4	HTTPS	TCP	443	0.0.0.0/0	-
<input type="checkbox"/>	-	sgr-02c5fb2b43a0f27b1	IPv4	HTTP	TCP	80	0.0.0.0/0	-

Figure 10: Web Tier Application Load Balancer Security Group

Additionally, another security group for the Web Tier instances is created named *WebTier-SG*. It too restricts SSH access to the IP of the designated device and permits HTTP and HTTPS traffic exclusively from the *ApplicationLoadBalancer-SG*.

8



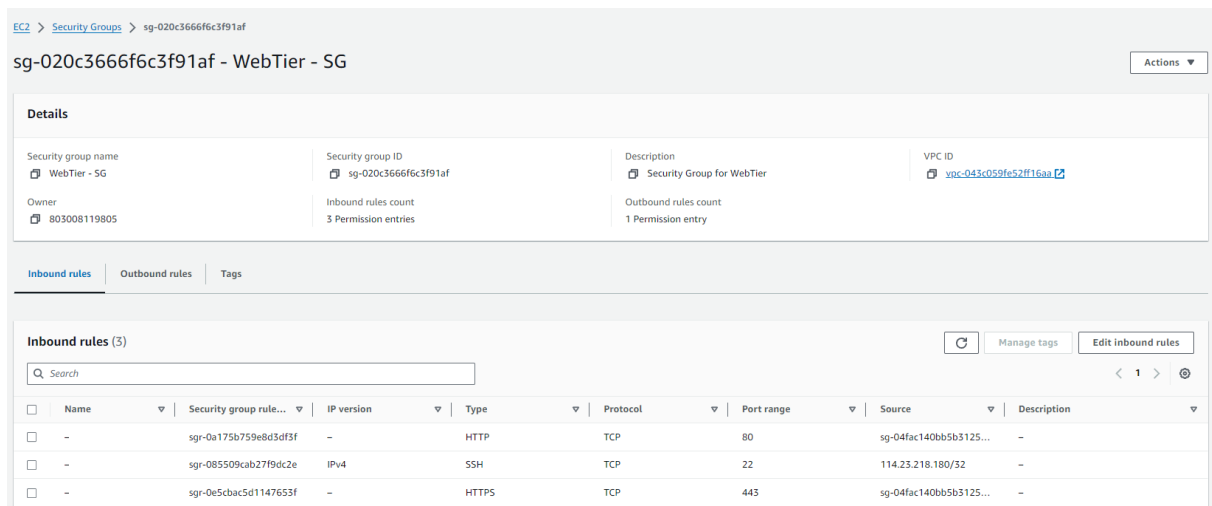


Figure 11: Web Tier Security Group

This setup asserts that users can only access the front-end tier through the ALB and not directly via instance public IPs. If users try to access the website using the public IP of the instance, they'll encounter an error screen.

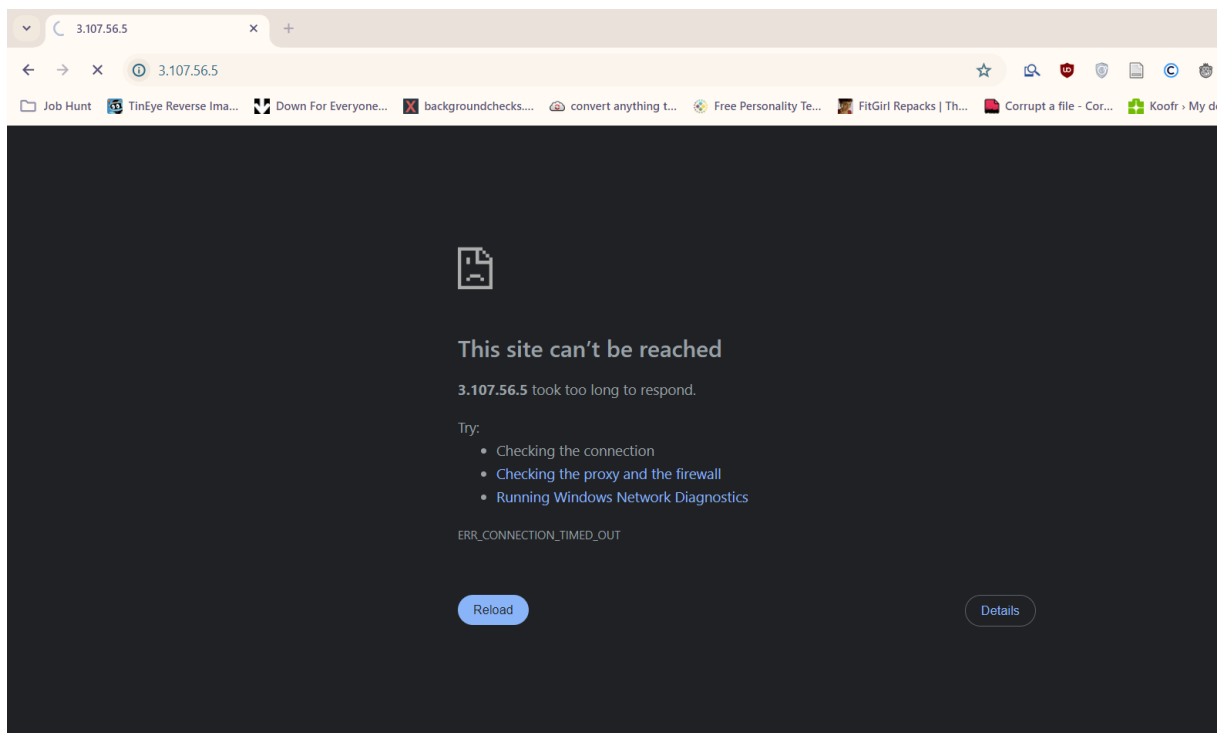


Figure 12: Error

However, accessing the website through the ALB will lead to successful loading of the application.

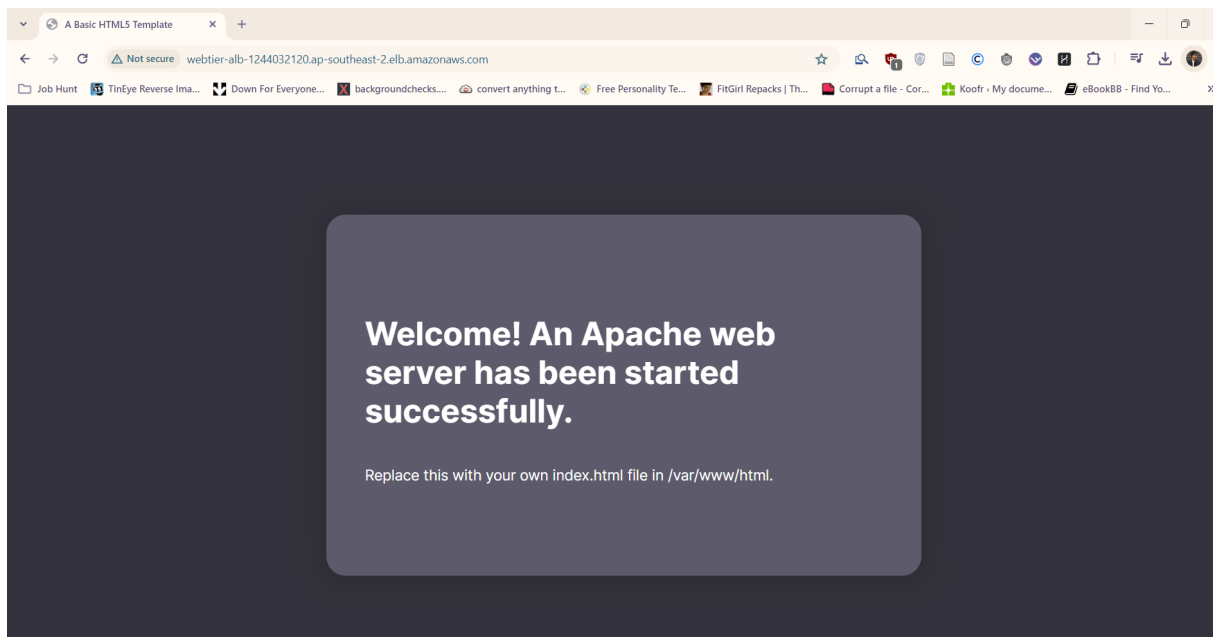


Figure 13: Success

Lastly, to insure the elasticity of the Web Tier, an Auto Scaling Group named *WebTier-ASG* is implemented. To configure the blueprint of the instances that it creates, a launch template named *WebTier-LaunchTemplate* is created. For compatibility with the current production environment, Red Hat Enterprise Linux 9 (RHEL 9) is chosen as the base Amazon Machine Image (AMI). This selection Guarantees continuity with the business's existing infrastructure while leveraging the benefits and features offered by RHEL 9.

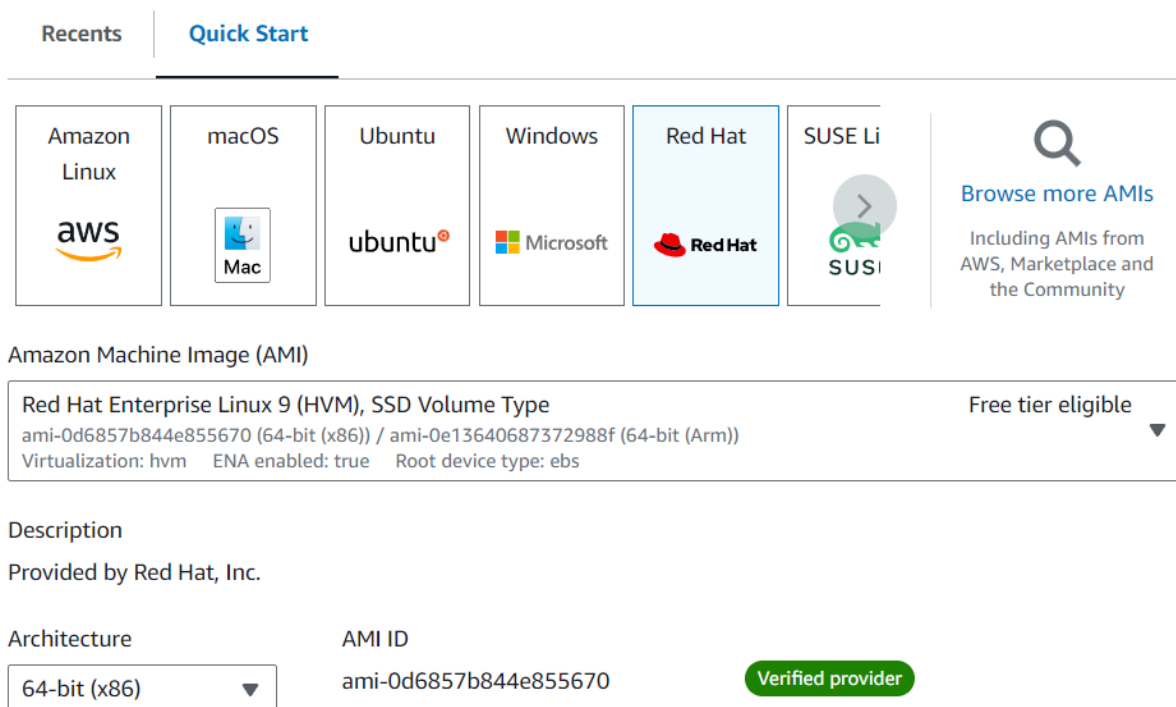


Figure 14: Launch Template AMI

The instance type selected for the Web Tier instances is the t3.xlarge. This instance type is chosen to align with the computing requirements of the current production environment. With 4 vCPUs and 16 GiB of memory, the t3.xlarge instance type matches the specifications of the existing servers, ensuring seamless transition and consistent performance.

▼ Instance type
[Info](#)
[Get advice](#)

Advanced

Instance type

t3.xlarge  
Family: t3 4 vCPU 16 GiB Memory Current generation: true  
On-Demand Linux base pricing: 0.2112 USD per Hour  
On-Demand RHEL base pricing: 0.2712 USD per Hour  
On-Demand SUSE base pricing: 0.2675 USD per Hour  
On-Demand Windows base pricing: 0.2848 USD per Hour

▼

☒ All generations  
[Compare instance types](#)

Additional costs apply for AMIs with pre-installed software

Figure 15: Launch Template Instance Type

*WebTier-LaunchTemplate* along with *WebTier-ALB* is attached to the *WebTier-ASG*.

Group details

Auto Scaling group name

WebTier-ASG

Launch template

Launch template	Version	Description
<a href="#">WebTier-LaunchTemplate</a> lt-01b23bf3cc3c05a66	Default	a launch template for web tier

Figure 16: Auto Scaling Group Launch Template

Load balancing

Load balancer 1

Name	Type	Target group
<a href="#">WebTier-ALB</a>	Application/HTTP	<a href="#">WebTier-TargetGroup</a>

Figure 17: Auto Scaling Group Load Balancer

*WebTier-ASG* is deployed across two public subnets, 172.20.1.0/24 (ap-southeast-2a) and 172.20.2.0/24 (ap-southeast-2b).

Network		
<b>Network</b> VPC <a href="#">vpc-043c059fe52ff16aa</a>		
Availability Zone	Subnet	
ap-southeast-2b	<a href="#">subnet-03d05db6c08049600</a>	172.20.2.0/24
ap-southeast-2a	<a href="#">subnet-035f65eeede51925b</a>	172.20.1.0/24

Figure 18: Auto Scaling Group Network

Furthermore, the ASG is configured with a minimum desired capacity of 1 instance which assures that the web tier always maintains at least one operational instance. However, to handle traffic spikes, the maximum capacity is set to 5 instances. This promises efficient resource utilization while avoiding overprovisioning, and thus optimizing cost-effectiveness. This scaling triggers at an average CPU utilization of 50%. Additionally, a scale-in policy is enabled that allows the architecture to terminate excess instances during periods of reduced demand or low resource utilization. This approach helps in cost optimization by eliminating unnecessary instances while promising adequate capacity to handle incoming requests.

Scaling		
Minimum desired capacity 1	Maximum desired capacity 5	
Target tracking policy Policy type Target tracking scaling	Scaling policy name Target Tracking Policy	Execute policy when As required to maintain Average CPU utilization at 50
Take the action Add or remove capacity units as required	Instances need 300 seconds to warm up before including in metric	Scale in Enabled

Figure 19: Auto Scaling Group Scaling Policy

In the testing environment, all components mirror the production setup for the web tier. However, to optimize costs, the instance type chosen for the Web Tier instances is t3.medium instead of t3.xlarge, maintaining compatibility while reducing resource overhead. These measures collectively address the present requirements while laying a robust foundation for future expansion.

### 3.3 Application Tier

One of the critical components of the Application Tier is the *Bastion Host*. In the architecture, it serves as a secure gateway for accessing the application servers deployed within the private subnets.

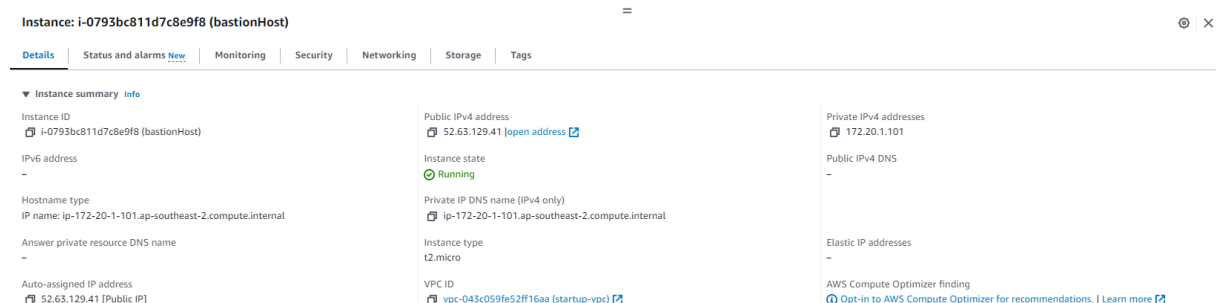


Figure 20: Bastion Host

To further enhance security, a dedicated security group is configured for the bastion host, restricting inbound SSH traffic solely from the IP address of the device.

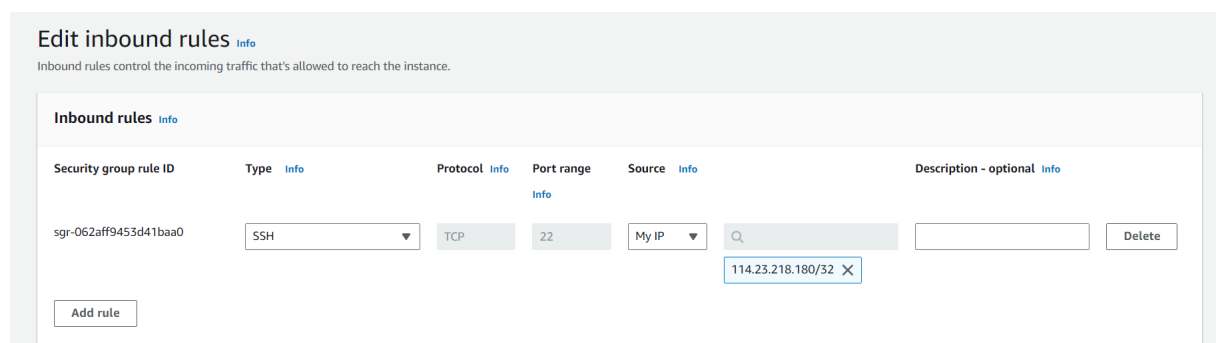


Figure 21: Bastion Host

Amazon Linux is opted as the operating system due to its reliability and compatibility for the *Bastion Host*.

Recents

Quick Start

Amazon Linux

aws

macOS

Mac

Ubuntu

ubuntu

Windows

Microsoft

Red Hat

Red Hat

SUSE Li

SUSE

Q

Browse more AMIs

Including AMIs from AWS, Marketplace and the Community

Amazon Machine Image (AMI)

Amazon Linux 2023 AMI

Free tier eligible

ami-04e1ec3aaf06b9060 (64-bit (x86), uefi-preferred) / ami-0f904469a29ed43ed (64-bit (Arm), uefi)

Virtualization: hvm    ENA enabled: true    Root device type: ebs

Description

Amazon Linux 2023 AMI 2023.4.20240416.0 x86\_64 HVM kernel-6.1

Architecture

64-bit (x86)

Boot mode

uefi-preferred

AMI ID

ami-04e1ec3aaf06b9060

Verified provider

Figure 22: Bastion Host Security Group

For the instance type t2.micro is selected because it aligns with the need for a cost-effective solution without compromising performance.

▼ Instance type

Info | Get advice

Instance type

t2.micro

Free tier eligible

Family: t2    1 vCPU    1 GiB Memory    Current generation: true

On-Demand SUSE base pricing: 0.0146 USD per Hour

On-Demand Linux base pricing: 0.0146 USD per Hour

On-Demand Windows base pricing: 0.0192 USD per Hour

On-Demand RHEL base pricing: 0.0746 USD per Hour

All generations

Compare instance types

Additional costs apply for AMIs with pre-installed software

Figure 23: Bastion Host Instance Type

The Bastion Host is placed within the first public subnet: "my-public-subnet-1".

▼ Network settings

Info

VPC - required

Info

vpc-043c059fe52ff16aa (startup-vpc)

172.20.0.0/20

▼

↻

Subnet

Info

subnet-035f65eede51925b

my-public-web-subnet-1

▼

↻

Create new subnet

VPC: vpc-043c059fe52ff16aa

Owner: 803008119805

Availability Zone: ap-southeast-2a

IP addresses available: 248

CIDR: 172.20.1.0/24)

Auto-assign public IP

Info

Enable

▼

Figure 24: Bastion Host Network

Subsequently, a security group for the application tier named *ApplicationTier-SG* is configured with three inbound rules. SSH access is restricted to only allow connections originating from the security group associated with the Bastion Host. This confirms that only authorized users with access to the Bastion Host can initiate SSH connections to the application tier instances for administrative purposes. ICMP traffic is permitted from the security group associated with the web tier. Inbound traffic on the MySQL/Aurora port is allowed exclusively from the security group associated with the database tier. This restricts database access solely to the application tier, enforcing data isolation and minimizing the risk of unauthorized access or malicious activity.

EC2 > Security Groups > sg-0c4c05aaba22ac59

sg-0c4c05aaba22ac59 - ApplicationTier - SG

Actions ▼

Details

<div>Security group name</div> <div>ApplicationTier - SG</div>	<div>Security group ID</div> <div>sg-0c4c05aaba22ac59</div>	<div>Description</div> <div>ssh and icmp from web server</div>	<div>VPC ID</div> <div>vpc-043c059fe52ff16aa</div>
<div>Owner</div> <div>803008119805</div>	<div>Inbound rules count</div> <div>3 Permission entries</div>	<div>Outbound rules count</div> <div>2 Permission entries</div>	

Figure 25: Security Group For Application Tier Host Network

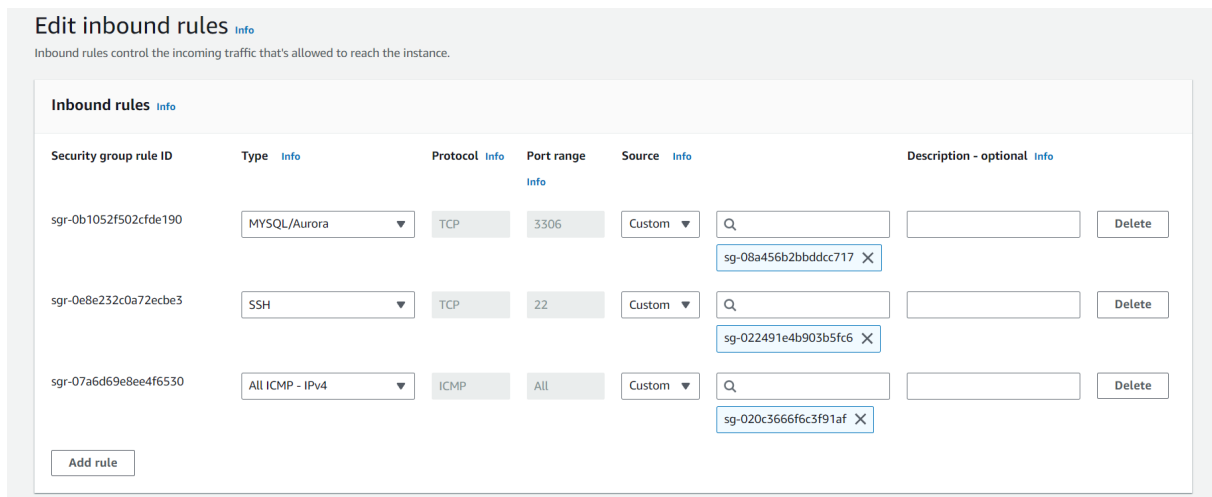


Figure 26: Inbound Rules

Moreover, to establish seamless scalability and manageability of the application tier, a launch template named *ApplicationTier-LaunchTemplate* is created for the Application Load Balancer. This launch template is designed to provision instances with Red Hat Enterprise Linux 9 as the Amazon Machine Image (AMI). This choice is made to maintain consistency with the current production environment which utilizes Red Hat Enterprise Linux 7.5.

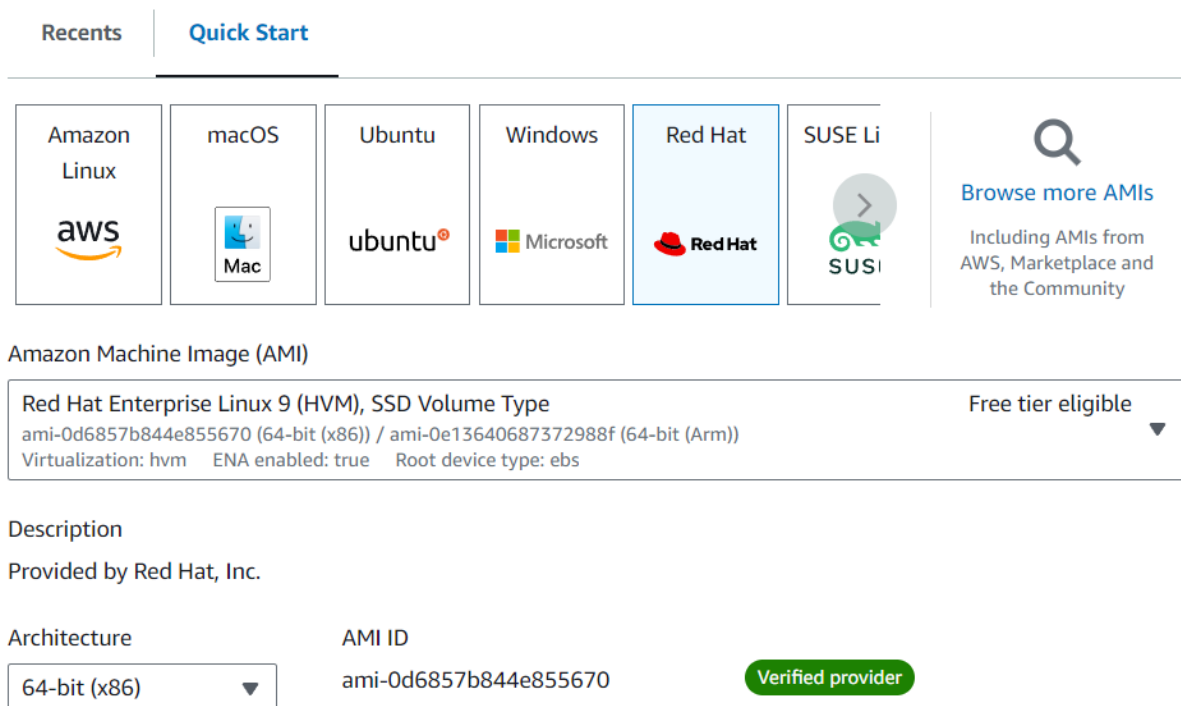


Figure 27: Application Tier Launch Template AMI

The instance type chosen for the application tier is t3.2xlarge. This instance type offers 8 vCPUs and 32GB of memory, providing ample resources for handling the application's



computational demands. Additionally, a 5000GB EBS volume (5TB) is attached to cater to the storage requirements. This configuration closely matches the specifications of the existing production environment, which comprises four physical servers with 8 CPUs, 32GB memory, and 5TB storage.

▼ Instance type
[Info](#)
[Get advice](#)

Advanced

Instance type

t3.2xlarge  
Family: t3 8 vCPU 32 GiB Memory Current generation: true  
On-Demand SUSE base pricing: 0.5474 USD per Hour  
On-Demand RHEL base pricing: 0.5524 USD per Hour  
On-Demand Windows base pricing: 0.5696 USD per Hour  
On-Demand Linux base pricing: 0.4224 USD per Hour

▼

☐ All generations  
[Compare instance types](#)

Additional costs apply for AMIs with pre-installed software

Figure 28: Application Tier Launch Template Instance Type

▼ Storage (volumes) [Info](#)

EBS Volumes [Hide details](#)

▼ Volume 1 (AMI Root) (Custom)

Storage type [Info](#)

EBS

Device name - required [Info](#)

/dev/xvda

Snapshot [Info](#)

snap-09e3ee230ee6826d1

Size (GiB) [Info](#)

5000

Volume type [Info](#)

gp3 ▼

IOPS [Info](#)

3000

Delete on termination [Info](#)

Yes ▼

Encrypted [Info](#)

Not encrypted ▼

KMS key [Info](#)

Don't include in launch tem... ▼

KMS keys are only applicable when encryption is set on this volume.

Throughput [Info](#)

125

Figure 29: Application Tier Launch Template Storage

On top of that, to distribute incoming traffic across the instances within the application tier, internal-facing load balancer named *ApplicationTier-LoadBalancer* is deployed. This load balancer is placed within two private subnets: "my-private-app-subnet-1" and "my-private-app-subnet-2". Putting the load balancer in a private subnet strengthens security by restricting direct access from the internet. The *ApplicationTier-SG* is attached to the

load balancer to enforce that only authorized traffic is allowed to communicate with the instances in the application tier.

Summary			
Review and confirm your configurations. <a href="#">Estimate cost</a>			
<b>Basic configuration</b> <a href="#">Edit</a> ApplicationTier-LoadBalancer <ul style="list-style-type: none"> <li>Internal</li> <li>IPv4</li> </ul>	<b>Security groups</b> <a href="#">Edit</a> <ul style="list-style-type: none"> <li>ApplicationTier - SG <a href="#">sg-0c4c05aabae22ac59</a></li> </ul>	<b>Network mapping</b> <a href="#">Edit</a> VPC <a href="#">vpc-043c059fe52ff16aa</a> startup-vpc <ul style="list-style-type: none"> <li>ap-southeast-2a <a href="#">subnet-058f5e896741fe632</a> my-private-app-subnet-1</li> <li>ap-southeast-2b <a href="#">subnet-04d0a30e2e92a5535</a> my-private-app-subnet-2</li> </ul>	<b>Listeners and routing</b> <a href="#">Edit</a> <ul style="list-style-type: none"> <li>HTTP:80 defaults to <a href="#">ApplicationTier-TargetGroup</a></li> </ul>
<b>Service integrations</b> <a href="#">Edit</a> AWS WAF: None AWS Global Accelerator: None		<b>Tags</b> <a href="#">Edit</a> None	

Figure 30: Application Tier Load Balancer

Additionally, to achieve high availability and fault tolerance, an auto scaling group named *ApplicationTier-ASG* is created. The *ApplicationTier-LaunchTemplate* is attached to this ASG.

Group details		
Auto Scaling group name		
ApplicationTier - ASG		
Launch template		
Launch template	Version	Description
<a href="#">ApplicationTier-LaunchTemplate</a>	Default	a launch template for the application tier
lt-0e02a66c2be5b4735		

Figure 31: Application Tier ASG

The ASG is configured to operate within two subnets, 172.20.3.0/24 (ap-southeast-2a) and 172.20.4.0/24 (ap-southeast-2a), thereby distributing the instances across multiple Availability Zones for improved resilience.

Network		
Network		
VPC		
<a href="#">vpc-043c059fe52ff16aa</a>		
Availability Zone	Subnet	
ap-southeast-2b	<a href="#">subnet-04d0a30e2e92a5535</a>	172.20.4.0/24
ap-southeast-2a	<a href="#">subnet-058f5e896741fe632</a>	172.20.3.0/24

Figure 32: Application Tier ASG Network

The *ApplicationTier-LoadBalancer* is associated with the ASG to facilitate the distribution of incoming traffic.

Load balancing		
Load balancer 1		
Name	Type	Target group
<a href="#">ApplicationTier-LoadBalancer</a>	Application/HTTP	<a href="#">ApplicationTier-TargetGroup</a>

Figure 33: Application Tier ASG Load Blancer

The scaling policy implemented for the Application Tier ASG mirrors that of the web tier ASG, warranting uniform scalability across the entire architecture

Lastly, The business requires application administrators to be notified via email when the application tier experiences a high volume of HTTP 404 errors (Not Found) per minute in the application. For this purpose, A CloudWatch alarm is created for the application tier load balancer. The alarm monitors the metric `HTTPCode_Target_404_Count`.

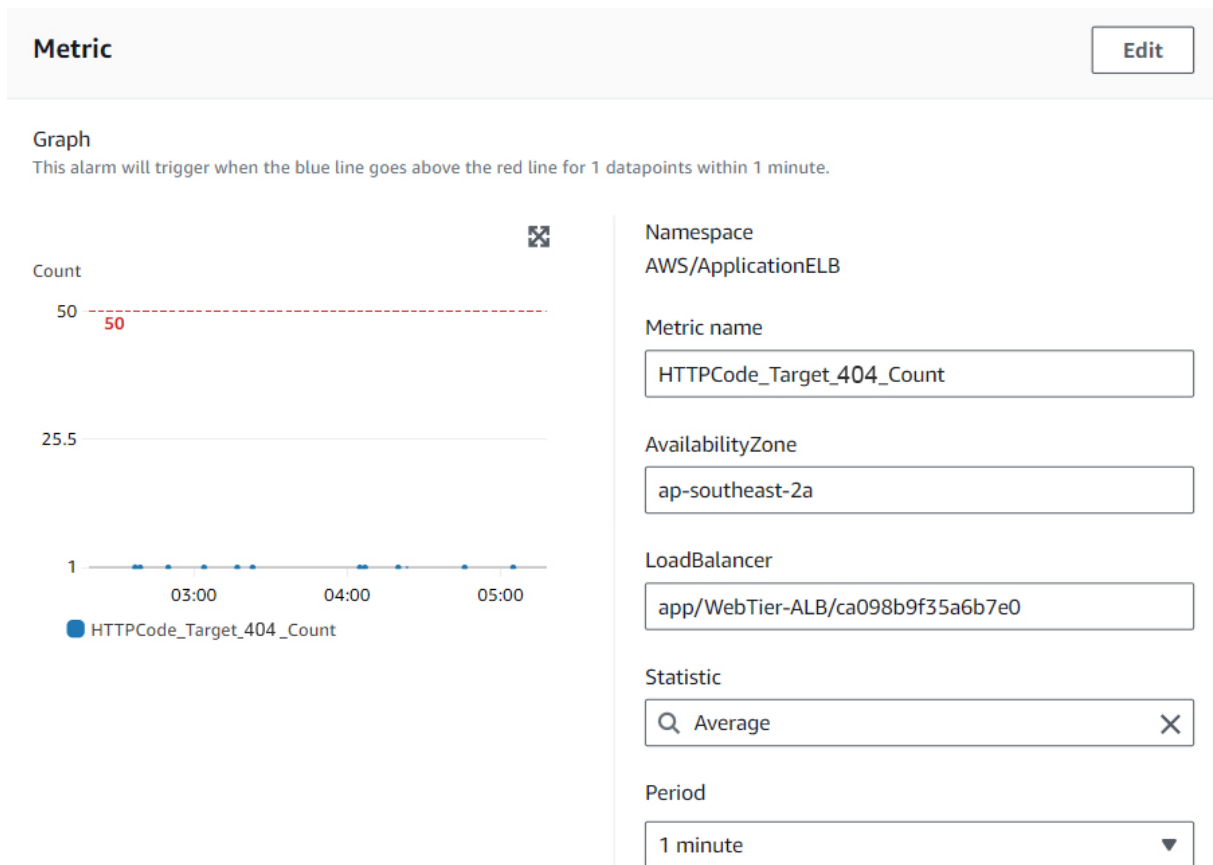


Figure 34: Cloud Watch Alarm Metric

The alarm is set to evaluate the metric over a 1-minute period. The alarm triggers when the HTTPCode\_Target\_404\_Count metric exceeds 50 within a 1-minute window. The alarm is configured to send notification emails to applicationadministrators@gmail.com.

**Conditions**

Threshold type

☒ Static  
Use a value as a threshold

☐ Anomaly detection  
Use a band as a threshold

Whenever HTTPCode\_Target\_4XX\_Count is...

Define the alarm condition.

☒ Greater  
> threshold

☐ Greater/Equal  
>= threshold

☐ Lower/Equal  
<= threshold

☐ Lower  
< threshold

than...

Define the threshold value.

50

Must be a number

► Additional configuration

Figure 35: Cloud Watch Alarm Conditions

In the testing environment, all aspects of the Application Tier mirror the production setup for security, scalability, and monitoring. However, to economize, the selected instance type for the application tier is t3.medium instead of t3.2xlarge, and a 1TB EBS volume is used instead of 5TB. These adjustments promises cost savings while maintaining functionality and compatibility with the production environment. By leveraging these AWS services, the current configuration effectively serves present demands while being adaptable for future growth requirements.

### 3.4 Database Tier

Firstly, a dedicated Security Group named *DataBaseTier-SG* is created.

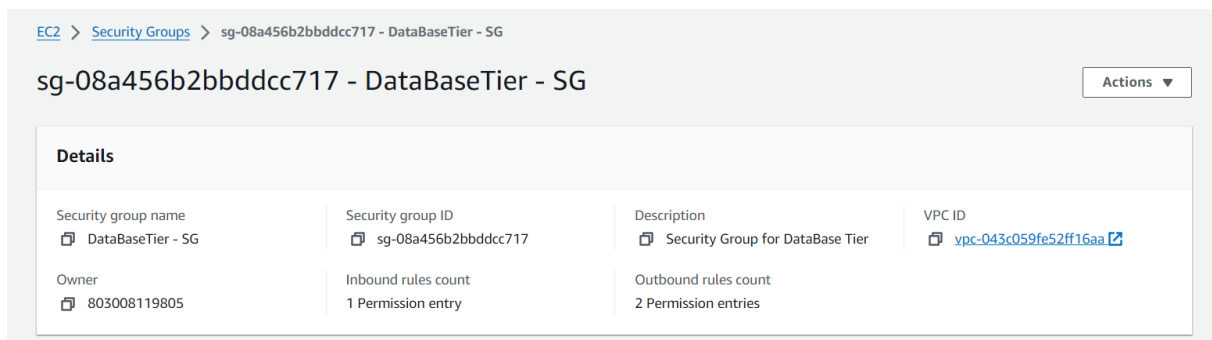


Figure 36: Security Group For Database Tier

This Security Group is configured with the following inbound and outbound rules.

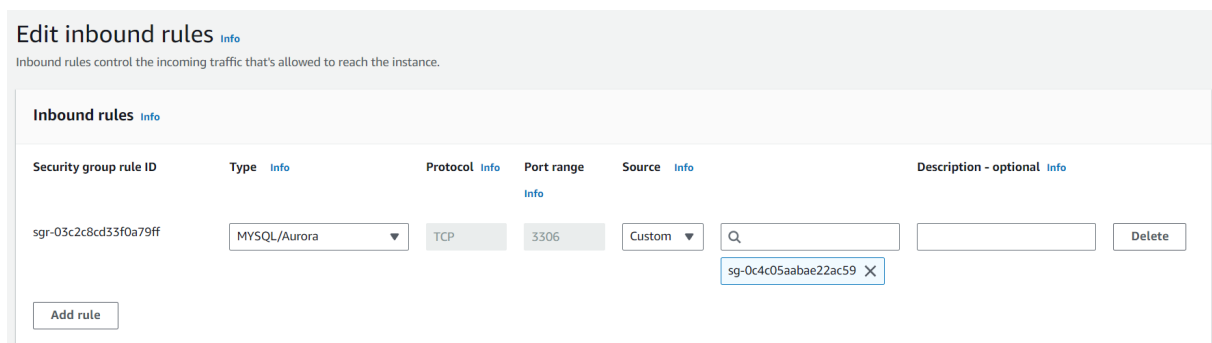


Figure 37: Inbound Rules

Edit outbound rules
[Info](#)

Outbound rules control the outgoing traffic that's allowed to leave the instance.

Outbound rules
[Info](#)

Security group rule ID	Type <a href="#">Info</a>	Protocol <a href="#">Info</a>	Port range <a href="#">Info</a>	Destination <a href="#">Info</a>	Description - optional <a href="#">Info</a>	
sgr-0969a8eb5adc9f10c	All traffic	All	All	Custom	Q 0.0.0.0/0	Delete
sgr-0b4bf59bfdadeb0b2	MySQL/Aurora	TCP	3306	Custom	Q sg-0c4c05aaba22ac59	Delete

Add rule

Figure 38: Outbound Rules

Secondly, MySQL 8.0.35 engine is selected for the database. This aligns with the current production environment which uses a MySQL 5.7.22 database cluster.

Engine Version

MySQL 8.0.35

Figure 39: Database Engine

The instance class selected is db.m5.2xlarge. This class provides processing power and memory comparable to the current production environment's two physical servers.

**Instance configuration**

The DB instance configuration options below are limited to those supported by the engine that you selected above.

DB instance class
[Info](#)

▼ Hide filters

☐ Show instance classes that support Amazon RDS Optimized Writes
[Info](#)
  
Amazon RDS Optimized Writes improves write throughput by up to 2x at no additional cost.

☐ Include previous generation classes

☒ Standard classes (includes m classes)

☐ Memory optimized classes (includes r and x classes)

☐ Burstable classes (includes t classes)

db.m5.2xlarge (supports Amazon RDS Optimized Writes)
  
8 vCPUs   32 GiB RAM   Network: 4,750 Mbps

Figure 40: Database Instance

Additionally, 5000 GB of storage is allocated to the database instance, matching the 5 TB storage available in the current production environment.

**Storage**

Storage type [Info](#)

Provisioned IOPS SSD (io2) storage volumes are now available.

Provisioned IOPS SSD (io1)


Flexibility in provisioning I/O

Allocated storage [Info](#)

5000

GiB

The minimum value is 100 GiB and the maximum value is 65,536 GiB




After you modify the storage for a DB instance, the status of the DB instance will be in storage-optimization. Your instance will remain available as the storage-optimization operation completes. [Learn more](#) 

Figure 41: Database Storage

Lastly, to establish high availability and durability of the database infrastructure, a Multi-AZ (Availability Zone) deployment strategy is employed. In the event of a failure or outage in one Availability Zone, the system seamlessly redirects traffic to a standby replica in another zone, thereby maintaining uninterrupted availability and safeguarding data integrity.

**Availability and durability**

Deployment options [Info](#)

The deployment options below are limited to those supported by the engine you selected above.

☐

**Multi-AZ DB Cluster**

Creates a DB cluster with a primary DB instance and two readable standby DB instances, with each DB instance in a different Availability Zone (AZ). Provides high availability, data redundancy and increases capacity to serve read workloads.

☒

**Multi-AZ DB instance**

Creates a primary DB instance and a standby DB instance in a different AZ. Provides high availability and data redundancy, but the standby DB instance doesn't support connections for read workloads.

☐

**Single DB instance**

Creates a single DB instance with no standby DB instances.

Figure 42: Database Availability And Durability

In the testing environment, all configurations for the Database Tier closely replicate those of the production setup while optimizing costs. However, to achieve cost efficiency, the instance type selected is db.r5b.large instead of db.m5.2xlarge, providing adequate processing power while reducing expenses. Additionally, a Single DB instance deployment strategy is chosen instead of Multi-AZ, and the storage is scaled down to 100GB from 5000GB, which assures cost savings efficiency while maintaining preserving functionality and compatibility with the production environment.

## 3.5 Storage

For storage purposes Amazon S3 service is utilized and S3 bucket named *startup-s3-bucket* is chosen for storing user assets (documents and pictures) due to its high durability, scalability, and cost-effectiveness.

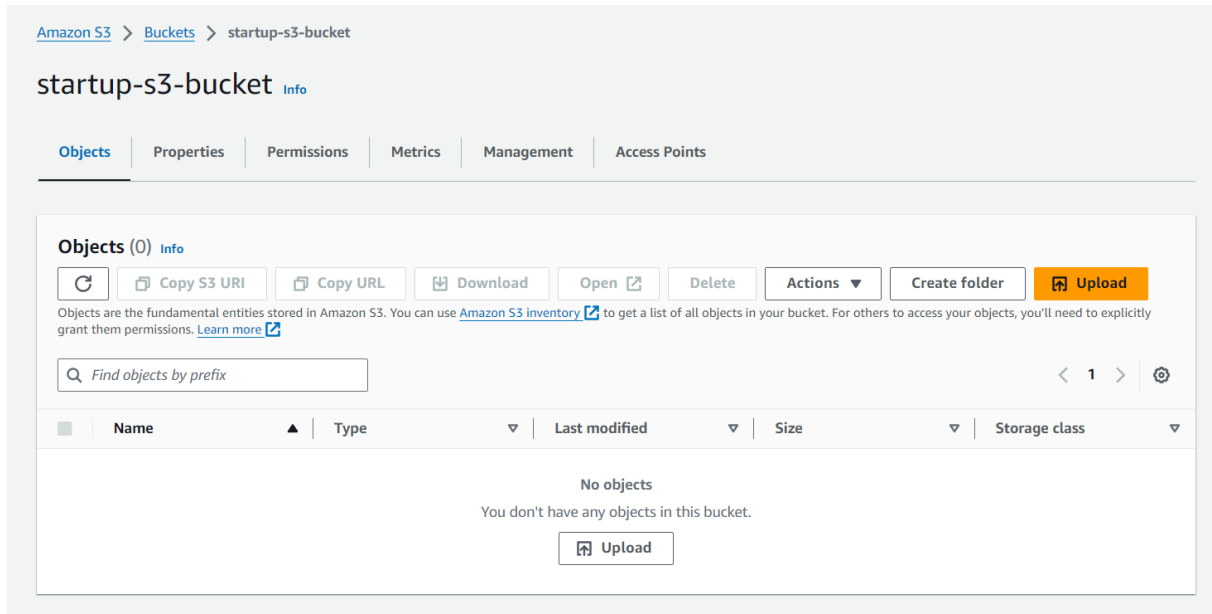


Figure 43: S3 Bucket

This bucket is configured with a lifecycle rule named *startup-s3-lifecycle-rule*.



## Lifecycle rule configuration

Lifecycle rule name

startup-s3-lifecycle-rule

Up to 255 characters

Choose a rule scope

- ☐ Limit the scope of this rule using one or more filters
- ☒ Apply to all objects in the bucket



### Apply to all objects in the bucket

If you want the rule to apply to specific objects, you must use a filter to identify those objects. Choose "Limit the scope of this rule using one or more filters". [Learn more](#)

☒ I acknowledge that this rule will apply to all objects in the bucket.

## Lifecycle rule actions

Choose the actions you want this rule to perform. Per-request fees apply. [Learn more](#) or see [Amazon S3 pricing](#)

- ☒ Move current versions of objects between storage classes
- ☐ Move noncurrent versions of objects between storage classes
- ☒ Expire current versions of objects
- ☐ Permanently delete noncurrent versions of objects
- ☐ Delete expired object delete markers or incomplete multipart uploads
- These actions are not supported when filtering by object tags or object size.

Figure 44: S3 Bucket Life cycle

This rule implements a tiered storage strategy. Objects are automatically transitioned to Glacier Instant Retrieval Storage after 90 days of creation. This storage class offers lower costs compared to standard S3 storage while still allowing retrieval within milliseconds. Objects are permanently deleted from Glacier after 5 years, fulfilling the business requirement of retaining user assets for that duration.

### Transition current versions of objects between storage classes

Choose transitions to move current versions of objects between storage classes based on your use case scenario and performance access requirements. These transitions start from when the objects are created and are consecutively applied. [Learn more](#)

Choose storage class transitions

Glacier Instant Retrieval

Days after object creation

90

Remove

Add transition

### Expire current versions of objects

For version-enabled buckets, Amazon S3 adds a delete marker and the current version of an object is retained as a noncurrent version. For non-versioned buckets, Amazon S3 permanently removes the object. [Learn more](#)

Days after object creation

1826

Figure 45: S3 Bucket Life cycle

### Review transition and expiration actions

Current version actions

Day 0

- Objects uploaded

↓

Day 90

- Objects move to Glacier Instant Retrieval

↓

Day 1826

- Objects expire

Noncurrent versions actions

Day 0

No actions defined.

Figure 46: S3 Bucket Life cycle

This strategy balances cost and access needs. User assets are unlikely to be accessed frequently after three months, making Glacier a suitable option. However, occasional access is still required, and Glacier Instant Retrieval provides low-latency retrieval when needed.

Additionally, a VPC endpoint named *s3-bucket-endpoint* is established which allows application tier instances residing within a private subnet of the VPC to securely access S3 storage without traversing the public internet

26

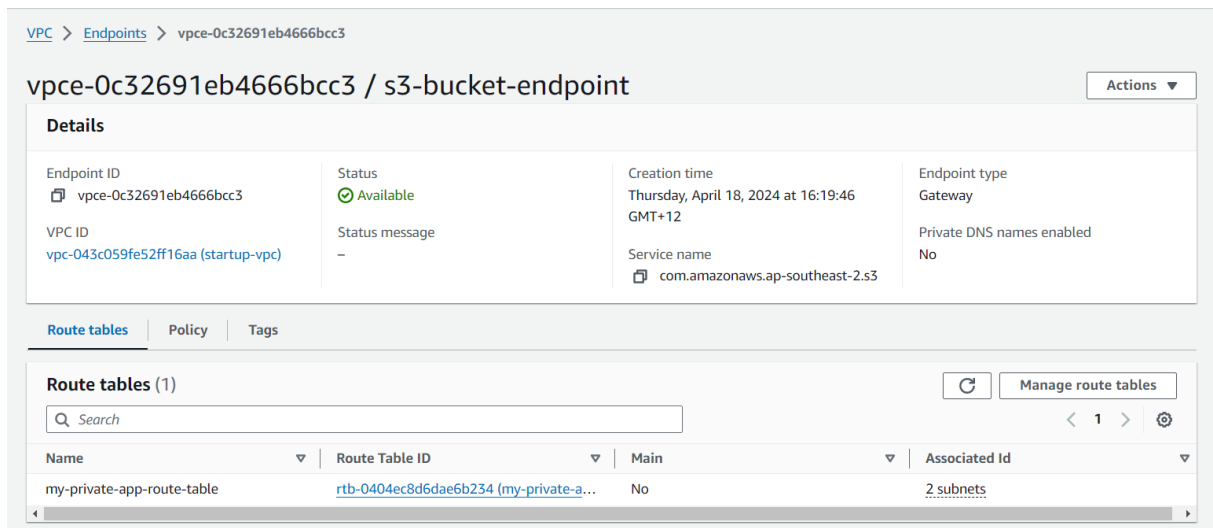


Figure 47: VPC Endpoint

In the testing environment, the same approach is adopted for storage utilizing Amazon S3 service.

### 3.6 Backup

The business requirements mentioned that a disaster recovery plan should be considered. Therefore, a backup strategy for the 3-tier AWS architecture is implemented using the AWS Backup service, and an AWS Backup rule named *startup-backup* is created to automate backup processes.

## Backup rule configuration [Info](#)

### Schedule

Backup rule name

startup-backup

Backup rule name is case sensitive. Must contain from 1 to 50 alphanumeric or '-' characters.

Backup vault [Info](#)

Default

Create new Backup vault

Backup frequency [Info](#)

Weekly

on ☐ Mon ☐ Tue ☐ Wed ☐ Thu ☐ Fri ☒ Sat ☐ Sun

### Backup window [Info](#)

Start time

Specify the time of day the backups will start. For **hourly** frequency, start time is the time the first backup is taken in a day. Where applicable, time will adjust to daylight savings time so that it retains the same local time all year.

00

:

30

Pacific/Auckland (UTC+12:00)

Start within [Info](#)

Specify period of time in which the backup plan starts if it doesn't start at the specified time.

8 hours

Complete within [Info](#)

7 days

Figure 48: Backup Rule Configuration

This rule is configured to retain backups for 5 days in warm storage and 95 days in cold storage. This guarantees quick access to recent backups while reducing costs for long-term retention.

## Lifecycle [Info](#)

### Cold storage [Info](#)

#### ☒ Move backups from warm to cold storage

Available for CloudFormation, DynamoDB with advanced features, EFS, SAP HANA, Timestream, and VMware virtual machines. Some resource types convert incremental backups to full backups. Requires at least 90 days of retention.

### Time in warm storage [Info](#)

Days



Recommended minimum is 8 days

### Cold storage for Amazon EBS [Info](#)

Archive Amazon EBS snapshots is available when cold storage is enabled and backup frequency is at least monthly.

### Total retention period [Info](#)

Tell AWS Backup how long to store your backups.

Days



### Total retention (days)

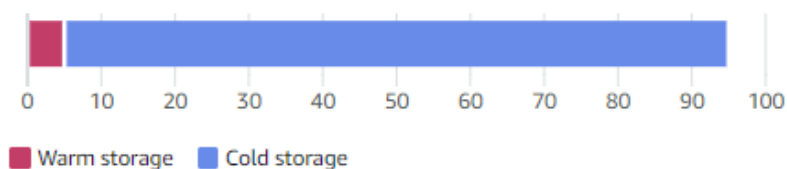


Figure 49: Backup Lifecycle

Moreover, the backup rule is set to include all resources associated with the application.

## Resource selection [Info](#)

Assign resources to this Backup plan using tags and resource IDs.

### 1. Define resource selection [Info](#)

Protect all resources or specify resources by type or ID.

#### ☒ Include all resource types

Protect all resource types that are enabled in your account.

#### ☐ Include specific resource types

Choose resources by type or specify individual resources by ID.

Figure 50: Backup Resource

In the testing environment, a disaster recovery plan utilizing the AWS Backup service was not implemented as in the production setup.

### 3.7 DDoS Protection

To fortify the infrastructure against Distributed Denial of Service (DDoS) attacks, Amazon Web Application Firewall (WAF) service is deployed. Within this framework, a Web ACL named *DDoS-Protection-WebACL* is established and linked with the WebTier Application Load Balancer (ALB).

**Web ACL details**

**Resource type**  
Choose the type of resource to associate with this web ACL. Changing this setting will reset the page.  
☐ Amazon CloudFront distributions  
☒ Regional resources (Application Load Balancers, Amazon API Gateway REST APIs, Amazon App Runner services, AWS AppSync GraphQL APIs, Amazon Cognito user pools and AWS Verified Access Instances)

**Region**  
Choose the AWS Region to create this web ACL in. Changing this setting will reset the page.  
Asia Pacific (Sydney) ▼

**Name**  
DDoS-Protection-WebACL  
The name must have 1-128 characters. Valid characters: A-Z, a-z, 0-9, - (hyphen), and \_ (underscore).

**Description - optional**  
  
The description can have 1-256 characters.

**CloudWatch metric name**  
DDoS-Protection-WebACL  
The name must have 1-128 characters. Valid characters: A-Z, a-z, 0-9, - (hyphen), and \_ (underscore).

Figure 51: Web ACL Details

**Associated AWS resources - optional (1)** Remove Add AWS resources

< 1 > ⚙

<input type="checkbox"/>	Name	Resource type	Region
<input type="checkbox"/>	WebTier-ALB	Application Load Balancer	Asia Pacific (Sydney)

Figure 52: Web ACL Resources

This Web ACL incorporates three key protection rules: the Amazon IP Reputation List, the Anonymous IP List, and the Bot Control rule.

<b>Rules (3)</b> If a request matches a rule, take the corresponding action. The rules are prioritized in order they appear.		
Name	Capacity	Action
AWS-AWSManagedRulesAmazonIpReputationList	25	Use rule actions
AWS-AWSManagedRulesAnonymousIpList	50	Use rule actions
AWS-AWSManagedRulesBotControlRuleSet	50	Use rule actions

Figure 53: Web ACL Rules

These rules collectively shield the architecture from various forms of malicious activity, including bot traffic, requests from anonymizing services, and sources flagged by Amazon's threat intelligence.

By integrating DDoS protection measures into the architecture, the security posture is bolstered, guaranteeing the availability and reliability of the application, and mitigating potential risks associated with DDoS attacks.

In the testing environment, WAF is not implemented as in the production setup.

## 4. Coast Analysis

The cost analysis delineates expenses across various components for both production and testing environments. In the production setup, expenditures include 2 NAT Gateways, each is assumed to handle 500GB of data monthly, resulting in a total monthly data processing of 1TB. This processing volume would incur a cost of \$145.14 per month.

### ▼ Show calculations

730 hours in a month x 0.059 USD = 43.07 USD (Gateway usage hourly cost)  
 500 GB per month x 0.059 USD = 29.50 USD (NAT Gateway data processing cost)  
 43.07 USD + 29.50 USD = 72.57 USD (NAT Gateway processing and month hours)  
 2 NAT Gateways x 72.57 USD = 145.14 USD (Total NAT Gateway usage and data processing cost)  
**Total NAT Gateway usage and data processing cost (monthly): 145.14 USD**

Figure 54: NAT-Production

Conversely, the testing environment employs a single NAT Gateway, and is assumed to process a quarter of the data load of the production environment which would amount to \$58.17 per month.

▼ Show calculations

730 hours in a month x 0.059 USD = 43.07 USD (Gateway usage hourly cost)  
256 GB per month x 0.059 USD = 15.10 USD (NAT Gateway data processing cost)  
43.07 USD + 15.10 USD = 58.17 USD (NAT Gateway processing and month hours)  
1 NAT Gateways x 58.17 USD = 58.17 USD (Total NAT Gateway usage and data processing cost)  
**Total NAT Gateway usage and data processing cost (monthly): 58.17 USD**

Figure 55: NAT-Testing

While AWS doesn't impose charges for inbound data, outbound data transfer incurs costs. With an anticipated 10TB of inbound data from the internet monthly, projected outbound data transfer cost would amount to \$583.68 per month, assuming 5TB of data is destined for the internet.

▼ Show calculations

Unit conversions

Inbound:

Internet: 10 TB per month x 1024 GB in a TB = 10240 GB per month

Outbound:

Internet: 5 TB per month x 1024 GB in a TB = 5120 GB per month

Pricing calculations

Inbound:

Internet: 10240 GB x 0 USD per GB = 0.00 USD

Intra region:

(0 GB x 0.01 USD per GB outbound) + (0 GB x 0.01 USD per GB inbound) = 0.00 USD

Outbound:

Internet: 5120 GB x 0.114 USD per GB = 583.68 USD

**Data Transfer cost (monthly): 583.68 USD**

Figure 56: IG-Production

In the testing environment the outbound data requirements are significantly lower. Hence assuming that maximum of 5GB of data would go out monthly, the associated cost would be minimal, amounting to just \$0.57 per month.

▼ Show calculations

Unit conversions

Inbound:

Internet: 10 TB per month x 1024 GB in a TB = 10240 GB per month

Pricing calculations

Inbound:

Internet: 10240 GB x 0 USD per GB = 0.00 USD

Intra region:

(0 GB x 0.01 USD per GB outbound) + (0 GB x 0.01 USD per GB inbound) = 0.00 USD

Outbound:

Internet: 5 GB x 0.114 USD per GB = 0.57 USD

**Data Transfer cost (monthly): 0.57 USD**

Figure 57: IG-Testing

Application load balancers would contribute \$36.79 per month for both environments.



▼ Show calculations

2 load balancers x 0.0252 USD per hour x 730 hours in a month = 36.79 USD

**Application Load Balancer fixed hourly charges (monthly): 36.79 USD**

Figure 58: ALB

In the production environment, the Web Tier instance, which uses RHEL as the operating system and a t3.xlarge instance type, aligns with business requirements at a cost of \$129.65 per month.

**Payment options**

Estimated commitment price based on the following selections:  
Instance type: **t3.xlarge** Operating system: **RHEL**

Select the container and options to find your best price

<p><input checked="" type="radio"/> <b>Compute Savings Plans</b> One plan that automatically applies to all usage on EC2, Fargate, and Lambda. Up to 66% discount. <a href="#">Learn more</a></p> <p>Reservation term</p> <p><input type="radio"/> 1 year</p> <p><input checked="" type="radio"/> 3 year</p> <p>Payment Options</p> <p><input checked="" type="radio"/> No upfront</p> <p><input type="radio"/> Partial upfront</p> <p><input type="radio"/> All upfront</p> <hr/> <p>Upfront: 0.00</p> <p>Monthly: 129.65/Month</p>	<p><input type="radio"/> <b>EC2 Instance Savings Plans</b> Get deeper discount when you only need one instance family and region. Up to 72% discount. <a href="#">Learn more</a></p> <p>Reservation term</p> <p><input type="radio"/> 1 year</p> <p><input checked="" type="radio"/> 3 year</p> <p>Payment Options</p> <p><input checked="" type="radio"/> No upfront</p> <p><input type="radio"/> Partial upfront</p> <p><input type="radio"/> All upfront</p> <hr/> <p>Upfront: 0.00</p> <p>Monthly: 110.38/Month</p>	<p><input type="radio"/> <b>On-Demand</b> Maximize flexibility. <a href="#">Learn more</a></p> <p>Expected utilization Enter the expected usage of Amazon EC2 instances</p> <p>Usage</p> <p>100</p> <p>Usage type</p> <p>Utilization percent per month</p> <hr/> <p>Instance: 0.2712/Hour</p> <p>Monthly: 197.98/Month</p>	<p><input type="radio"/> <b>Spot Instances</b> Minimize cost by leveraging EC2's spare capacity. Recommended for fault tolerant and interruption tolerant applications. <a href="#">Learn more</a></p> <p>The historical average discount for t3.xlarge is 0%</p> <p>Assume percentage discount for my estimate</p> <p>0</p> <p><b>Actual spot instance pricing varies</b> With spot instances, you pay the spot price that's in effect for the time period your instance is running</p> <hr/> <p>Instance: 0.2712/Hour</p> <p>Monthly: 197.98/Month</p>
--	---	--	--

Figure 59: EC2-Web-Production

Meanwhile, for the testing environment, t3.medium instance type is selected, resulting in a reduced expense of \$65.26 per month.

**Payment options**

Estimated commitment price based on the following selections:  
Instance type: **t3.medium** Operating system: **RHEL**

Select the container and options to find your best price

<p><input checked="" type="radio"/> <b>Compute Savings Plans</b> One plan that automatically applies to all usage on EC2, Fargate, and Lambda. Up to 66% discount. <a href="#">Learn more</a></p> <p>Reservation term</p> <p><input type="radio"/> 1 year</p> <p><input checked="" type="radio"/> 3 year</p> <p>Payment Options</p> <p><input checked="" type="radio"/> No upfront</p> <p><input type="radio"/> Partial upfront</p> <p><input type="radio"/> All upfront</p> <hr/> <p>Upfront: 0.00</p> <p>Monthly: 65.26/Month</p>	<p><input type="radio"/> <b>EC2 Instance Savings Plans</b> Get deeper discount when you only need one instance family and region. Up to 72% discount. <a href="#">Learn more</a></p> <p>Reservation term</p> <p><input type="radio"/> 1 year</p> <p><input checked="" type="radio"/> 3 year</p> <p>Payment Options</p> <p><input checked="" type="radio"/> No upfront</p> <p><input type="radio"/> Partial upfront</p> <p><input type="radio"/> All upfront</p> <hr/> <p>Upfront: 0.00</p> <p>Monthly: 60.44/Month</p>	<p><input type="radio"/> <b>On-Demand</b> Maximize flexibility. <a href="#">Learn more</a></p> <p>Expected utilization Enter the expected usage of Amazon EC2 instances</p> <p>Usage</p> <p>100</p> <p>Usage type</p> <p>Utilization percent per month</p> <hr/> <p>Instance: 0.1128/Hour</p> <p>Monthly: 82.34/Month</p>	<p><input type="radio"/> <b>Spot Instances</b> Minimize cost by leveraging EC2's spare capacity. Recommended for fault tolerant and interruption tolerant applications. <a href="#">Learn more</a></p> <p>The historical average discount for t3.medium is 0%</p> <p>Assume percentage discount for my estimate</p> <p>-1</p> <p><b>Actual spot instance pricing varies</b> With spot instances, you pay the spot price that's in effect for the time period your instance is running</p> <hr/> <p>Instance: 0.1128/Hour</p> <p>Monthly: 83.17/Month</p>
---	--	---	--

Figure 60: EC2-Web-Testing

Similarly, in the Application Tier, RHEL-operated instances utilize a t3.2xlarge instance type to meet business needs, incurring a cost of \$266.52 per month.

**Payment options**

Estimated commitment price based on the following selections:  
Instance type: **t3.2xlarge** Operating system: **RHEL**

Select the container and options to find your best price

<p><input checked="" type="radio"/> <b>Compute Savings Plans</b> One plan that automatically applies to all usage on EC2, Fargate, and Lambda. Up to 66% discount. <a href="#">Learn more</a></p> <p>Reservation term <input type="radio"/> 1 year <input checked="" type="radio"/> 3 year</p> <p>Payment Options <input checked="" type="radio"/> No upfront <input type="radio"/> Partial upfront <input type="radio"/> All upfront</p> <hr/> <p>Upfront: 0.00 Monthly: 266.52/Month</p>	<p><input type="radio"/> <b>EC2 Instance Savings Plans</b> Get deeper discount when you only need one instance family and region. Up to 72% discount. <a href="#">Learn more</a></p> <p>Reservation term <input type="radio"/> 1 year <input checked="" type="radio"/> 3 year</p> <p>Payment Options <input checked="" type="radio"/> No upfront <input type="radio"/> Partial upfront <input type="radio"/> All upfront</p> <hr/> <p>Upfront: 0.00 Monthly: 228.13/Month</p>	<p><input type="radio"/> <b>On-Demand</b> Maximize flexibility. <a href="#">Learn more</a></p> <p>Expected utilization Enter the expected usage of Amazon EC2 instances</p> <p>Usage <input type="text" value="100"/></p> <p>Usage type <input type="text" value="Utilization percent per month"/></p> <hr/> <p>Instance: 0.5524/Hour Monthly: 403.25/Month</p>	<p><input type="radio"/> <b>Spot Instances</b> Minimize cost by leveraging EC2's spare capacity. Recommended for fault tolerant and interruption tolerant applications. <a href="#">Learn more</a></p> <p>The historical average discount for t3.2xlarge is 0%</p> <p>Assume percentage discount for my estimate <input type="text" value="-1"/></p> <div> <p><b>Actual spot instance pricing varies</b> With spot instances, you pay the spot price that's in effect for the time period your instance is running</p> </div> <hr/> <p>Instance: 0.5524/Hour Monthly: 407.28/Month</p>
--	---	---	--

Figure 61: EC2-App-Production

Additionally, 5TB of storage is essential for production, amounting to \$480 per month.

▼ Show calculations

730 total EC2 hours / 730 hours in a month = 1.00 instance months

5,000 GB x 1.00 instance months x 0.096 USD = 480.00 USD (EBS Storage Cost)

**EBS Storage Cost: 480.00 USD**

**Amazon Elastic Block Store (EBS) total cost (monthly): 480.00 USD**

Figure 62: EBS-Production

In the testing environment, a t3.medium instance paired with 500GB of EBS storage costs \$113.26 per month.

**Payment options**

Estimated commitment price based on the following selections:  
Instance type: **t3.medium** Operating system: **RHEL**

Select the container and options to find your best price

☒ **Compute Savings Plans**  
One plan that automatically applies to all usage on EC2, Fargate, and Lambda. Up to 66% discount. [Learn more](#)

Reservation term  
☐ 1 year  
☒ 3 year

Payment Options  
☒ No upfront  
☐ Partial upfront  
☐ All upfront

Upfront: 0.00  
Monthly: 65.26/Month

☐ **EC2 Instance Savings Plans**  
Get deeper discount when you only need one instance family and region. Up to 72% discount. [Learn more](#)

Reservation term  
☐ 1 year  
☒ 3 year

Payment Options  
☒ No upfront  
☐ Partial upfront  
☐ All upfront

Upfront: 0.00  
Monthly: 60.44/Month

☐ **On-Demand**  
Maximize flexibility. [Learn more](#)

Expected utilization  
Enter the expected usage of Amazon EC2 instances

Usage  
100

Usage type  
Utilization percent per month

Instance: 0.1128/Hour  
Monthly: 82.34/Month

☐ **Spot Instances**  
Minimize cost by leveraging EC2's spare capacity. Recommended for fault tolerant and interruption tolerant applications. [Learn more](#)

The historical average discount for t3.medium is 0%

Assume percentage discount for my estimate  
-1

**Actual spot instance pricing varies**  
With spot instances, you pay the spot price that's in effect for the time period your instance is running

Instance: 0.1128/Hour  
Monthly: 83.17/Month

Figure 63: EC2-App-Testing

▼ Show calculations

730 total EC2 hours / 730 hours in a month = 1.00 instance months  
500 GB x 1.00 instance months x 0.096 USD = 48.00 USD (EBS Storage Cost)  
**EBS Storage Cost: 48.00 USD**  
**Amazon Elastic Block Store (EBS) total cost (monthly): 48.00 USD**

Figure 64: EBS-Testing

The Bastion Host has a modest price of \$5.91 per month.

**Payment options**

Estimated commitment price based on the following selections:  
Instance type: **t2.micro** Operating system: **Linux**

Select the container and options to find your best price

☒ **Compute Savings Plans**  
One plan that automatically applies to all usage on EC2, Fargate, and Lambda. Up to 66% discount. [Learn more](#)

Reservation term  
☐ 1 year  
☒ 3 year

Payment Options  
☒ No upfront  
☐ Partial upfront  
☐ All upfront

Upfront: 0.00  
Monthly: 5.91/Month

☐ **EC2 Instance Savings Plans**  
Get deeper discount when you only need one instance family and region. Up to 72% discount. [Learn more](#)

Reservation term  
☐ 1 year  
☒ 3 year

Payment Options  
☒ No upfront  
☐ Partial upfront  
☐ All upfront

Upfront: 0.00  
Monthly: 5.18/Month

☐ **On-Demand**  
Maximize flexibility. [Learn more](#)

Expected utilization  
Enter the expected usage of Amazon EC2 instances

Usage  
100

Usage type  
Utilization percent per month

Instance: 0.0146/Hour  
Monthly: 10.66/Month

☐ **Spot Instances**  
Minimize cost by leveraging EC2's spare capacity. Recommended for fault tolerant and interruption tolerant applications. [Learn more](#)

The historical average discount for t2.micro is 63%

Assume percentage discount for my estimate  
-1

**Actual spot instance pricing varies**  
With spot instances, you pay the spot price that's in effect for the time period your instance is running

Instance: 0.0146/Hour  
Monthly: 10.76/Month

Figure 65: Bastion Host

Database costs peak at \$6498.76 per month for production. It reflects the utilization of resources tailored to meet the demands of a high-performance environment.

▼ Show calculations

1 instance(s) x 7.012 USD hourly x (24 / 24 hours in a day) x 730 hours in a month = 5118.7600 USD (RDS for SQL Server)  
**RDS for SQL server cost (monthly): 5,118.76 USD**

**RDS for SQL server cost (upfront): 0.00 USD**

Figure 66: DB-Instance-Production

▼ Show calculations

5,000 GB per month x 0.276 USD x 1 instances = 1,380.00 USD (Storage Cost)  
**Storage pricing (monthly): 1,380.00 USD**

Figure 67: DB-Storage-Production

In contrast, for the testing environment, cost optimization measures are implemented that results in a significantly reduced expenditure of \$144 per month. This cost reduction is achieved by selecting the db.t3.medium instance type and the Single-AZ deployment option.

▼ Show calculations

1 instance(s) x 0.104 USD hourly x (100 / 100 Utilized/Month) x 730 hours in a month = 75.9200 USD (RDS for SQL Server)  
**RDS for SQL server cost (monthly): 75.92 USD**

**RDS for SQL server cost (upfront): 0.00 USD**

Figure 68: DB-Instance-Testing

▼ Show calculations

500 GB per month x 0.138 USD x 1 instances = 69.00 USD (Storage Cost)  
**Storage cost (monthly): 69.00 USD**

Figure 69: DB-Storage-Testing

CloudWatch monitoring would incur a nominal fee of \$0.30 across both environments.

▼ Show calculations

Tiered price for: 1 metrics  
1 metrics x 0.30 USD = 0.30 USD  
Total tier cost = 0.30 USD (Metrics cost (includes custom metrics))  
**CloudWatch Metrics cost (monthly): 0.30 USD**

Figure 70: CloudWatch

In the production environment, it is assumed that the S3 bucket stores 5TB of data in Standard storage, along with 500,000 ongoing monthly requests for PUT, COPY, POST, or LIST, as well as 500,000 ongoing monthly requests for GET, SELECT, and other requests while the ongoing monthly volume of Standard Restore retrieval data is 2000 GB. This would incur a cost of \$132.57 per month.

▼ Show calculations

Unit conversions

S3 Standard storage: 5 TB per month x 1024 GB in a TB = 5120 GB per month

Pricing calculations

Tiered price for: 5,120 GB

5,120 GB x 0.025 USD = 128.00 USD

Total tier cost = 128.00 USD (S3 Standard storage cost)

500,000 PUT requests for S3 Standard Storage x 0.0000055 USD per request = 2.75 USD (S3 Standard PUT requests cost)

500,000 GET requests in a month x 0.0000044 USD per request = 0.22 USD (S3 Standard GET requests cost)

2,000 GB x 0.0008 USD = 1.60 USD (S3 select returned cost)

128 USD + 0.22 USD + 2.75 USD + 1.60 USD = 132.57 USD (Total S3 Standard Storage, data requests, S3 select cost)

**S3 Standard cost (monthly): 132.57 USD**

**S3 Standard cost (upfront): 0.00 USD**

Figure 71: S3-Standard-Production

After 3 months, it is assumed that 3TB of data would transition to S3 Glacier Instant Retrieval, accompanied by 125,000 ongoing monthly PUT, COPY, POST, or LIST requests and 125,000 ongoing monthly GET, SELECT, and other requests while the ongoing monthly volume of Standard Restore retrieval data is 300 GB. The associated cost amounts for this would be \$32.04 monthly.

▼ Show calculations

Unit conversions

S3 Glacier Instant Retrieval storage: 3 TB per month x 1024 GB in a TB = 3072 GB per month

S3 Glacier Instant Retrieval Average Object Size: 16 MB x 0.0009765625 GB in a MB = 0.015625 GB

Pricing calculations

3,072 GB per month / 0.015625 GB average item size = 196,608.00 unrounded number of objects

Round up by 1 (196608.0000) = 196608 number of objects

3,072 GB x 0.005 USD = 15.36 USD (S3 Glacier Instant Retrieval storage cost)

125,000 PUT requests for S3 Glacier Instant Retrieval Storage x 0.00002 USD per request = 2.50 USD (S3 Glacier Instant Retrieval PUT requests cost)

125,000 GET requests for S3 Glacier Instant Retrieval Storage x 0.00001 USD per request = 1.25 USD (S3 Glacier Instant Retrieval GET requests cost)

300 GB x 0.03 USD = 9.00 USD (S3 Glacier Instant Retrieval data retrievals cost)

15.36 USD + 2.50 USD + 1.25 USD + 9.00 USD = 28.11 USD (Total S3 Glacier Instant Retrieval Storage and other costs)

**S3 Glacier Instant Retrieval cost (monthly): 28.11 USD**

0.015625 S3 Glacier Instant Retrieval Average Object Size / 8 object size / 0.0009765625 GB = 2.00 parts (unrounded)

Round up by 1 (2.000000) = 2 parts

196,608 number of objects x 0.00002 USD = 3.93216 USD (Cost for PUT, COPY, POST requests for initial data)

**S3 Glacier Instant Retrieval cost (upfront): 3.93 USD**

Figure 72: S3-Glacier-Production

Consequently, the total cost for the production environment's S3 bucket reaches \$164.61 per month.

In contrast, for the testing environment, adopting a scaled-down approach with resources quartered, the total cost of the S3 bucket would amount to \$38.65 per month

▼ Show calculations

Tiered price for: 1,280 GB  
 $1,280 \text{ GB} \times 0.025 \text{ USD} = 32.00 \text{ USD}$   
 Total tier cost = 32.00 USD (S3 Standard storage cost)  
 $125,000 \text{ PUT requests for S3 Standard Storage} \times 0.0000055 \text{ USD per request} = 0.6875 \text{ USD}$  (S3 Standard PUT requests cost)  
 $125,000 \text{ GET requests in a month} \times 0.00000044 \text{ USD per request} = 0.055 \text{ USD}$  (S3 Standard GET requests cost)  
 $500 \text{ GB} \times 0.0008 \text{ USD} = 0.40 \text{ USD}$  (S3 select returned cost)  
 $32 \text{ USD} + 0.055 \text{ USD} + 0.6875 \text{ USD} + 0.40 \text{ USD} = 33.14 \text{ USD}$  (Total S3 Standard Storage, data requests, S3 select cost)  
**S3 Standard cost (monthly): 33.14 USD**

**S3 Standard cost (upfront): 0.00 USD**

Figure 73: S3-Standard-Testing

▼ Show calculations

Unit conversions

S3 Glacier Instant Retrieval Average Object Size:  $16 \text{ MB} \times 0.0009765625 \text{ GB in a MB} = 0.015625 \text{ GB}$

Pricing calculations

$320 \text{ GB per month} / 0.015625 \text{ GB average item size} = 20,480.00$  unrounded number of objects  
 Round up by 1 (20480.0000) = 20480 number of objects  
 $320 \text{ GB} \times 0.005 \text{ USD} = 1.60 \text{ USD}$  (S3 Glacier Instant Retrieval storage cost)  
 $41,667 \text{ PUT requests for S3 Glacier Instant Retrieval Storage} \times 0.00002 \text{ USD per request} = 0.8333 \text{ USD}$  (S3 Glacier Instant Retrieval PUT requests cost)  
 $41,667 \text{ GET requests for S3 Glacier Instant Retrieval Storage} \times 0.00001 \text{ USD per request} = 0.4167 \text{ USD}$  (S3 Glacier Instant Retrieval GET requests cost)  
 $75 \text{ GB} \times 0.03 \text{ USD} = 2.25 \text{ USD}$  (S3 Glacier Instant Retrieval data retrievals cost)  
 $1.60 \text{ USD} + 0.8333 \text{ USD} + 0.4167 \text{ USD} + 2.25 \text{ USD} = 5.10 \text{ USD}$  (Total S3 Glacier Instant Retrieval Storage and other costs)  
**S3 Glacier Instant Retrieval cost (monthly): 5.10 USD**

$0.015625 \text{ S3 Glacier Instant Retrieval Average Object Size} / 8 \text{ object size} / 0.0009765625 \text{ GB} = 2.00$  parts (unrounded)

Round up by 1 (2.000000) = 2 parts

$20,480 \text{ number of objects} \times 0.00002 \text{ USD} = 0.4096 \text{ USD}$  (Cost for PUT, COPY, POST requests for initial data)

**S3 Glacier Instant Retrieval cost (upfront): 0.41 USD**

Figure 74: S3-Glacier-Testing

Furthermore, assuming that 10 million web requests would be received across Web ACL with 3 rules, the implementation of Web Application Firewall would incur a cost of \$14 per month in the production environment.

▼ Show calculations

$1 \text{ Web ACLs per month} \times 5.00 \text{ USD} = 5.00 \text{ USD}$  (WAF Web ACLs cost)  
 $1 \text{ Web ACLs per month} \times 3.00 \text{ Billable Rules per web ACL per month} \times 1.00 \text{ USD} = 3.00 \text{ USD}$  (WAF Rules cost)  
 $10 \text{ requests per month} \times 1000000 \text{ multiplier for million} \times 0.0000006 \text{ USD} = 6.00 \text{ USD}$  (WAF Requests cost)  
 $5.00 \text{ USD} + 3.00 \text{ USD} + 6.00 \text{ USD} = 14.00 \text{ USD}$   
**WAF cost (monthly): 14.00 USD**

Figure 75: WAF

Lastly the disaster recovery plan which is critical for ensuring business continuity in unforeseen circumstances, would entail an estimated maximum cost of \$1000 monthly in production. No such plan is outlined for the testing environment.

The estimated monthly cost to run the application in the AWS production environment is approximately \$9352.36. This cost is driven by factors like NAT Gateway usage, data transfer out to the internet, database instance selection, and storage needs. The testing environment has a significantly lower estimated monthly cost of around \$462.91 due to reduced resource utilization.

## 5. Architectural Principles and Achievements

### 5.1 How The Architecture Achieves Reliability

Reliability within the architecture is guaranteed through the implementation of multiple key approaches. First, resources are deployed across two Availability Zones within the Sydney region. This redundancy safeguards against potential outages in any single zone. If a failure occurs in one AZ, resources deployed in the other AZ can continue operations, minimizing downtime and securing service continuity.

In addition, a backup strategy is implemented for the 3-tier AWS architecture using the AWS Backup service. An AWS Backup rule named *startup-backup* is created to automate backup processes. The backup rule is set to backup all resources associated with the application.

Furthermore, for the the web tier, an Application Load Balancer named *WebTier-ALB* is utilized as an internet-facing component operating within the public subnets. For the application tier, an internal-facing load balancer named *ApplicationTier-LoadBalancer* is strategically placed within two private subnets. These load balancers play a crucial role in distributing incoming traffic across multiple instances, enhancing the reliability of the architecture.

Lastly, a VPC Endpoint for Amazon S3 is integrated. This configuration reduces latency and minimizes the potential impact of internet outages on S3 access. Thus, the reliability of the architecture is improved even further.

Overall, these measures collectively contribute to the reliability and resilience of the architecture.

### 5.2 How The Architecture Achieves Elasticity

In the AWS architecture, elasticity is ensured by implementing various strategies. First of all, an Auto Scaling Group named *WebTier-ASG* is set up for the web tier and is deployed across two public subnets. This ASG is configured to scale from a minimum of 1 instance to a maximum of 5 instances based on an average CPU utilization of 50%.

This setup allows the infrastructure to dynamically adjust its capacity to handle traffic spikes efficiently. Additionally, an Application Load Balancer named *WebTier-ALB* is established as an internet-facing component. This ALB evenly distributes incoming traffic across instances within the WebTier-ASG, enhancing elasticity.

For the application tier, the setup is mirrored with an Auto Scaling Group named *ApplicationTier-ASG* deployed across two private subnets and configured with similar scaling policies based on CPU utilization. Besides, an internal-facing load balancer named *ApplicationTier-LoadBalancer* is also established. It efficiently distributes traffic among instances in the ApplicationTier-ASG.

Moreover, the deployment across multiple availability zones and the distribution of resources across public and private subnets not only ensures high availability but also contributes to elasticity by allowing the infrastructure to scale horizontally across different zones.

By implementing these configurations and components, elasticity is dynamically achieved to meet the workload demands of the business.

### 5.3 How The Architecture Achieves Security

The infrastructure is fortified by carefully designed security measures. To begin with, VPC endpoint is leveraged to access the S3 bucket exclusively from resources within the VPC. This eliminates potential exposure arising from public internet access.

Afterwards, the system is segmented into three tiers: web tier, application tier, and database tier, as well as Public and Private Subnets are implemented. This segregation protects application logic and database instances within private subnets from direct internet access, mitigating security risks.

Furthermore, each tier is bolstered by dedicated security groups. The *WebTier-SG* restricts inbound traffic to HTTP/HTTPS solely from the *ApplicationLoadBalancer-SG*, preventing direct access to web servers. Similarly, the *ApplicationTier-SG* allows SSH access only from the security group associated with the Bastion Host, ensuring administrative access originates from a secure entry point. The inclusion of a Bastion Host functions as a secure gateway, restricting SSH access to authorized devices. This multi-layered approach adds an additional security barrier before reaching application servers. Finally, the *DataBaseTier-SG* further strengthens security by restricting access to the database.

Moreover, to confront the looming threat of Distributed Denial of Service (DDoS) attacks,



the Amazon Web Application Firewall (WAF) service is deployed with a custom Web ACL named *DDoS-Protection-WebACL* to safeguard the architecture. This Web ACL utilizes Amazon’s threat intelligence and pre-configured rules to block malicious traffic originating from anonymous IPs, bots, and sources flagged as threats. This proactive defense strengthens the architecture’s resilience against DDoS attempts.

Through these measures such as network segmentation, access control via security groups, a bastion host, and WAF protection, a secure environment for the architecture is created.

## 5.4 How The Architecture Achieves Cost Optimization

Cost efficiency within the architecture has been a top priority. One key approach for that used in this architecture is right-sizing resources based on workload demands. Instance types are carefully selected to meet the specific requirements; for example, t3.medium instances are utilized for testing environment. This reduces unnecessary overhead while ensuring adequate performance levels.

Auto-scaling groups and on-demand provisioning are used to make it possible to change how resources are allocated based on changing demand patterns. This way, costs are kept to a minimum during times of low activity.

Moreover, cost-effective data transfer strategies are implemented, such as consolidating outbound traffic through NAT Gateways. For storage purposes, a lifecycle management policy is utilized that transitions infrequently accessed data to Glacier, a more cost-efficient storage tier that reduces long-term storage costs without compromising accessibility or durability.

Furthermore, the architecture incorporates monitoring and alerting mechanisms through CloudWatch, allowing proactive identification of instance failures and mitigating the cost of unhealthy instances. Additionally, Bastion Host is leveraged to safeguard against unauthorized usage and potential financial liabilities.

On top of that, a pragmatic approach to database management has been adopted. The architecture uses appropriate instance types and storage capacities based on workload characteristics for the database.

Overall, the cost optimization strategy is based on a holistic and proactive approach that includes resource optimization, data management, monitoring, security, and the strategic use of managed services.

## 6. Summary

In summary, this report has outlined the successful migration of the education software company's IT infrastructure to a secure, scalable, and cost-effective cloud-based architecture on AWS. The implemented three-tier architecture provides the necessary modularity and flexibility to accommodate future growth and fluctuating demands.

A cost analysis was conducted to safeguard that the migration aligns with financial goals. The report also demonstrated how the architecture adheres to key cloud principles of elasticity, security, reliability, and cost-optimization to avoid resource over-provisioning.

By adopting this cloud-based solution, the education software company is now well-positioned to:

- Effortlessly adapt to changing user demands and application growth.
- Respond quickly to new market opportunities and feature requests.
- Leverage the pay-as-you-go model and eliminate the need for upfront infrastructure investment.
- Benefit from AWS's robust security measures.
- Experience high availability with built-in redundancy and fault tolerance.

This successful migration paves the way for the education software company to focus on its core business – delivering a superior learning experience for its users.