CAPSTONE REPORT
MARCH 22 2024

Cyber Connexion
— Powered by —
The Fields Institute

From Books to Bytes:

Strengthening Security for the Toronto Public Library with Cloud Computing

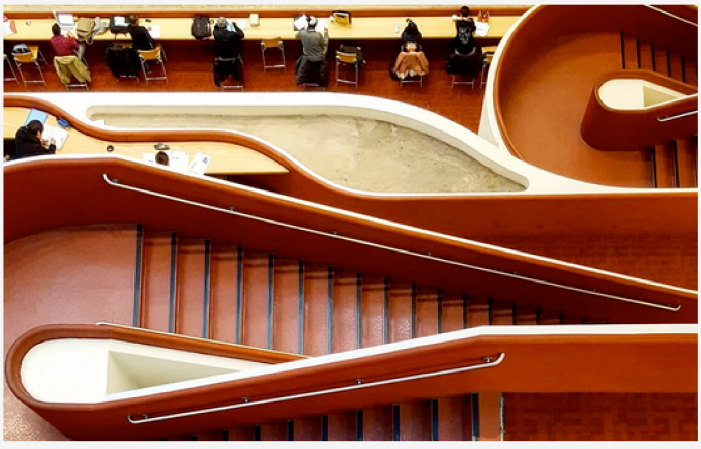
PREPARED BY
Andrew Colinet
Catherine Ducharme
Javaria Nauman
Galo Ginocchio

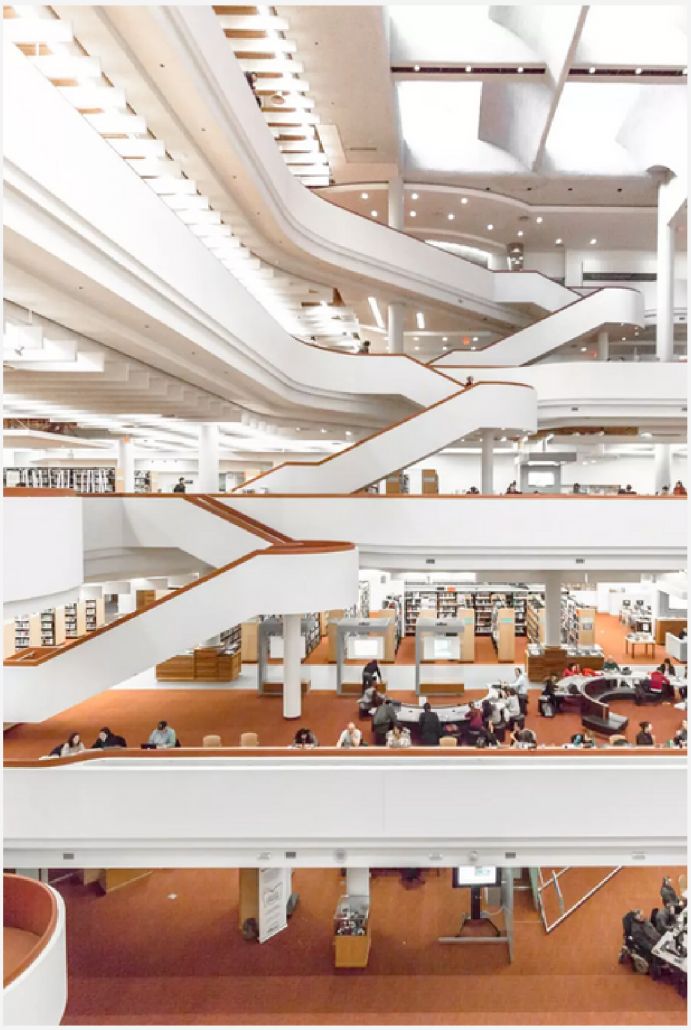
SUPERVISED BY
Aaron Crighton
Sonya Goulet
Mojdeh Akhavan
Charbel Akiki
Mukitul Khan
Osman Mohammed
Hendra Hendrawan

Toronto Public Library

1. Introduction

Following the ransomware attack targeting Toronto Public Library (TPL) on October 28th 2023, our team designed a secure network architecture to host a web application to be used by customers to manage their public library accounts, and browse the catalogue. The challenge this project met was protecting critical TPL infrastructure. In particular TPL's public-facing application and extensive catalogue of media from threat actors. We also ensured that these assets remain accessible at all times — a crucial business requirement for the library. In order to balance these needs and to enhance the application's security, accessibility, and scalability, we designed a hybrid cloud environment on Amazon Web Services (AWS), discussed in Section 2. The architecture that was deployed includes two virtual private clouds (VPC): one production environment to host the application and its database; and one development environment to test updates and patches before deployment. The rest of TPL's network infrastructure will remain in an on-premises data centre. The components of this architecture are detailed in Section 3, and a visual representation is included in the report's appendix. We developed this system with two foundational security principles in mind: defence in depth and zero trust. These principles are detailed in Section 4, along with the specific security measures used to implement them. We conducted regular vulnerability scans as part of our defence in depth strategy. The results from our first scans are discussed in Section 5. These scans include a CIS Benchmark scan to ensure compliance with the National Institute of Standards and Technology (NIST) Cybersecurity Framework. These frameworks guide organisations like TPL in managing cybersecurity risks to best protect their business interests. NIST framework compliance is discussed in Section 6.





2. Why a private cloud?

Cloud computing for a web application brings tangible business benefits to TPL:

- **Enhanced flexibility:** It can take significant time before additional servers can be up and running when demand for a service increases. In the cloud, it can be done in a matter of minutes.
- **Rapid elasticity:** When web traffic goes up in the morning, or down at night, the app automatically adjusts by creating or eliminating web servers.
- **Pay-as-you-go:** TPL should only pay for the services they use, when they use them. For an app like ours, which sees high traffic during the day and low at night, servers will not incur expenses when they are not used. This is in contrast to what would be needed if the app was hosted on-premises, and a key part of this architecture that will lead to substantial savings for TPL.

Nevertheless, we have elected to leave a portion of TPL's network infrastructure on-premises, specifically data and software that are more predictable in nature, and therefore do not require rapid scaling.

INFRASTRUCTURE AS CODE

We have chosen Terraform, an infrastructure as Code (IaC) tool, to facilitate deployment to the cloud. Terraform offers operational and security advantages, including:

- IaC facilitates automation, version control, and repeatability of deployments. This approach improves consistency, while reducing risk of security misconfigurations and human error.
- Terraform supports various cloud providers, which means portability if TPL ever needs to switch providers. Furthermore, it also supports deployment on-premises, allowing for consistent management of TPL's infrastructure.
- Terraform is scalable and suitable for managing infrastructure of any size, making it adaptable to TPL's evolving business needs.
- Terraform's modular approach promotes code reuse and simplifies management of complex environments like TPL's web application, further ensuring consistency.

Thus, using Terraform bolsters TPL's security posture while streamlining its cloud infrastructure management.

A three-tier web application in a production environment in the cloud

This VPC hosts a highly available, scalable and reliable platform to host the library web application and database. It is used by TPL's customers to browse the library catalogue and manage their loans. To ensure fault tolerance and redundancy, the application is spread over two availability zones (AZ), guaranteeing that in the event of a data centre outage, the app remains fully operational and available. All subnets are protected by Network Access Control Lists (NACL).

3. Network architecture

With the above cloud computing advantages in mind, and with Terraform as our tool, we designed a secure three-part architecture that efficiently balances TPL's operational and security needs.

01

PRESENTATION LAYER WITH BASTION HOST

The first, Internet-facing tier of the application is hosted on two public subnets. All traffic coming from the Internet is redirected through a round robin load balancer, configured with session stickiness. This tier includes a bastion host that provides an additional layer of security by acting as a gateway for TPL administrators to access and manage the production environment securely, without interfering with the functioning of the application. This layer acts as a Demilitarized Zone, or DMZ, to isolate the private networks that host the application and database from the resources that must remain connected to the Internet at all times.

02

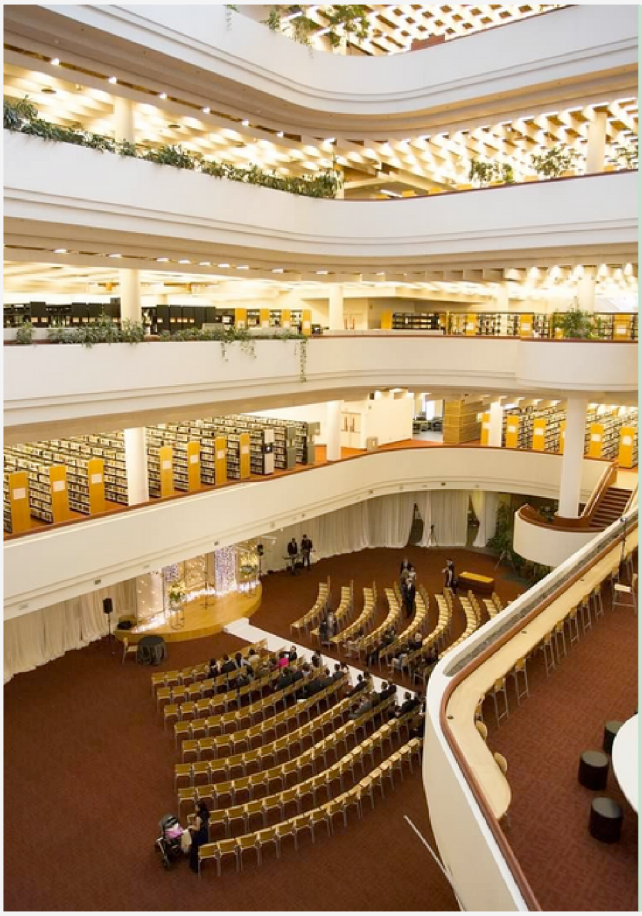
BUSINESS LOGIC LAYER

The second tier is hosted on private subnets and contains an auto-scalable group of Elastic Compute Cloud (EC2) machines acting as web servers for the application. Every EC2 instance is launched with an encrypted 5GB Elastic block Storage (EBS) volume. Being able to scale compute capacity as needed means that the web servers can be destroyed and recreated as needed, to adjust to the increases and decreases in web traffic to the application.

03

DATA STORAGE LAYER

The third tier, which also resides in private subnets, hosts a Relational Database Service (RDS) configured with phpmyadmin. The database hosted in the first AZ is used for main storage and interaction with the web app. The one hosted in the second AZ is a hot backup that will replace the main database in case of outage. Both databases are encrypted using server-side encryption with AWS Key Management Service (KMS). This ensures that the personal information contained in the online accounts of the library's customers stays secure and confidential.



A development environment

This second VPC hosts five EC2 Virtual Machines, each running a different operating system (MS Windows, Ubuntu, SUSE Linux, Amazon Linux and Debian). These machines will be used to facilitate development, testing, and deployment of new features, updates, and patches without impacting the live system in the production environment. This testing environment will also be cloud-based. This offers us the flexibility to scale the number of testing machines up and down, and to modify the operating systems and other configurations at will. As an added security measure, this non-production environment is not connected to the Internet. Like the bastion host in the production VPC, it can only be accessed through a transit gateway by TPL staff working in the on-premises environment.

A corporate, on-premises environment

This third component of our proposed architecture includes a public subnet with a customer gateway used by TPL administrators to connect to the bastion host with a site-to-site VPN that tunnels through the transit gateway. The private subnet hosts a virtual machine running Tenable, a SaaS vulnerability management system. Sensitive information that does not need to be shared with customers through their online TPL accounts, such as TPL staff payrolls and human resources data, will be kept here on a well-protected database in a private subnet, along with archival materials.

Protecting the business

By deliberately implementing safeguards guided by industry gold standards, Toronto Public Library business interests will be robustly protected as a result of this architecture implementation. This work is In response to a recent cyberattack on October 28th 2023 where threat actors stole confidential and sensitive staff information on a vulnerable server, ensuring protections for workers in this system is an urgently identified need. Protecting this sensitive information shields Toronto Public Library from future legal liability due to negligence, and failure to protect the sensitive information entrusted to them by their employees. Furthermore, the nature of the information stolen, including social insurance numbers, dates of birth, home addresses, and scans of government issue identification is a profound reputational risk to the institution, with ripple effects. These include a heightened risk of identity theft and fraud not only to past and present staff, but their families as well.

4. Foundational Security Principles and Security Measures

We adopted an approach blending defence-in-depth and zero trust principles to ensure TPL's assets are secured against threat actors that attempt to access their system, thereby ensuring their confidentiality, integrity and availability. This approach protects the revenue of the business by limiting, if not eliminating, risk of threat actors accessing staff records. In this section, we define these principles and detail the security measures that we implemented to enforce them.

DEFENCE IN DEPTH

This cybersecurity strategy involves deploying multiple layers of security controls and measures to protect against various types of threats and attacks, in order to create redundancy and overlap in security defences. This makes it challenging for threat actors to compromise our system and data: even if they breach one layer, additional security layers are in place to block attacks. Here is an overview of how we implemented defence in depth in our architecture:

- **Firewalls**: We deployed Network Access Control Lists that act as network-based firewalls at every subnet entrance point in Internet-connected environments. We configured security groups to act as an application-level firewall for every EC2 instance in our architecture. This allows us to filter and control incoming and outgoing traffic, protecting against common web-based attacks, and blocking access to threat actors.
- **Encryption**: Due to AWS account limitations, we have not been able to implement encryption and HTTPS at this stage. Both the content of our databases (primary and backup) and the EBS volumes used by the web application would be encrypted in order to protect data at rest. To protect data in transit, our firewalls would be configured to block unsecured HTTP connections on port 80 and use TLS-encrypted HTTPS exclusively (port 443). Encrypting data in transit is a crucial security measure that protects against common threats, such as spoofing or man-in-the-middle attacks, that might lead to compromised accounts and personally identifiable information (PII) leaks.
- **Network segmentation**: We segmented the production VPC into separate subnets in order to limit lateral movement and contain potential breaches. Each subnet has its own set of access controls and policies, enforced through NACLs and security groups, reducing the attack surface.
- The **bastion host** ("jump box"), hosted in the application's web tier, is not connected to the Internet, and can only be accessed through a VPN connection and from the TPL's on-premises data centre. This ensures that only trusted administrators can connect directly to the web application and modify its code and settings.
- **Multi-factor authentication**: Due to AWS account limitations, we have not been able to implement multi-factor authentication. AWS IAM would be configured to enforce multi-factor authentication by default for all accounts, whether user, employee, or administrator.
  - ***A note on MFA***: Multi-factor authentication would be enabled by default for all users, regardless of their roles, including both TPL staff and library customers. While this might be seen as a business impediment by some, MFA is one of the most efficient measures that can be taken to protect against a variety of cyber attacks. Implementing MFA by default enhances TPL's security posture by adding an additional layer of protection, significantly reducing the risk of unauthorised access and potential data breaches. Further, the use of MFA is endorsed by NIST, adopted by Canadian governments, as an integral component of the Zero Trust framework.
- **Testing environment**: Our non-production environment hosts a set of virtual machines that will be used to inspect and test all updates and patches before they are deployed to production. This helps TPL administrators ensure changes do not introduce any security risk into the TPL web app, and protects against supply chain attacks.
- **Logging and monitoring**: Due to AWS account limitations and budget constraints, we have not been able to implement Amazon Cloudwatch and Amazon Cloudtrail. Amazon Cloudwatch and CloudTrail would be used in tandem to ensure that all traffic in the cloud is being continuously monitored and that any irregular activity on the network will set off an alert for TPL's security team to investigate.
- **Session stickiness**: This feature enables load balancers to direct a user's requests to the same web server for the duration of their session. While this measure is being implemented primarily for operational reasons (maintain session state, improve application performance), it also offers protection against common attacks, such as session hijacking and replay attacks.

No entity, whether inside or outside TPL's network perimeter, should be trusted by default. This includes TPL employees and administrators. Several security controls that we implemented to ensure defence in depth, such as MFA, data encryption and network monitoring, help us enforce zero trust on our system. Here are a few examples of how we are applying the three fundamental rules of Zero Trust:

- **Least privilege**: This refers to granting users only the minimum level of access or permissions necessary to perform their tasks, reducing the risk of unauthorised actions or data breaches. For instance, using IAM, we would limit database access for the web application to read-only permissions for non-administrative users.
- **Always verify:** This refers to the validation of identities at every point of entry in TPL's system to prevent security vulnerabilities and ensure trustworthiness. This includes the necessary authentication of all users in order to use the web application.
- **Assume breach**: This means adopting the mindset that malicious actors may already be within the network, prompting proactive security measures and continuous monitoring to detect and respond to intrusions promptly. This includes limiting the possibility of lateral movement through network segmentation.

5. Vulnerability Management

The non-production environment will be regularly scanned for potential vulnerabilities before updates are deployed to production, giving developers insights into how any changes might affect security, and how developers can make the application more secure. Tenable Nessus, our selected Vulnerability Management Solution, outranks comparable products by InsightVM and Qualis on all metrics. Given this software's emphasis on asset discovery, vulnerability assessment, and compliance monitoring, it is the most versatile and appropriate tool to protect sensitive data. Therefore, the business interests of the Toronto Public Library are best served by this tool. Here are the results of the first scan we ran on this environment:

SCAN: NESSUS AGENT SCAN

The non-production environment will be regularly scanned for potential vulnerabilities before updates are deployed to production, giving developers insights into how any changes might affect security, and how developers can make the application more secure. Our selected Vulnerability Management Solution Tenable Nessus outranks comparable products on all metrics. Given this software's emphasis on asset discovery, vulnerability assessment, and compliance monitoring, it is the most versatile and appropriate tool to protect sensitive data. Nessus agents have been installed in the production environment to ensure a thorough vulnerability scanning capacity. Here are the results of the first internal, agent scan we ran on this environment in Appendix E:

As you can see, no serious vulnerabilities were detected by the agent scan on the production environment, due to our use of the latest, most secure operating system available, and of course the secure environment we created. However, it is also clear from the results that there are ways to make our application safer without impacting its functionality. For instance, it seems that having our web tier accessible by SSH is a potential security risk. This port is used by administrators to connect to the bastion host in the application's web tier. Therefore, we have learned from this report that we could reduce our attack surface by finding a different solution than a bastion host to enable administrator access.

SCAN: EXTERNAL SCAN

To ensure thorough testing of the environments' vulnerabilities, we ran external scans, in order to assess the actions we could take to further protect TPL's assets against external threat actors. For instance, the results for our instances running older versions of Debian and Amazon Linux have shown some vulnerabilities that will need to be addressed in Appendix F: The SSH server used by this instance is configured to support Cipher Block Chaining (CBC) encryption and is configured to allow key exchange algorithms which are considered weak. This may allow an attacker to recover the plaintext message from the ciphertext. Furthermore, the remote SSH server is vulnerable to a weakness known as Terrapin. This can allow a remote, man-in-the-middle attacker to bypass integrity checks and downgrade the connection's security. Those are clear weaknesses in the system that we are planning to fix.

We have also configured our system with internal Nessus scanners on our on-premises environment. These scanners connect to the development environment, which hosts the machines that are used for testing updates and patches before deployment. These scanners will be used to conduct compliance and ransomware ecosystem scans, among others, on both our non-production and production environment.

6. Conclusion and Future Improvements

Design constraints and costs have precluded a number of security measures that we believe would greatly improve the security posture of TPL's system.

- An **Intrusion detection system** solution was planned for the production environment but could not be included in the final product for cost reasons. We recommend configuring AWS Guard Duty in the production environment before it goes live. The combination of IDS and firewalls provides several layers of protection that will protect TPL's web assets against a wider variety of attacks. In particular, since most ransomware has been documented, an up-to-date signature-based IDS could detect it on the network before it did any damage.
- **AWS Firewalls** are a versatile solution that could be configured to complement NACLs and Security groups. These three services work at different layers of the OSI model and would contribute to our defence in depth strategy. For instance, at Layer 7, an AWS WAF (web access firewall) could protect against cross-site scripting, and SQL injection.
- **AWS Identity and Access Management** (IAM) must be set up to configure, among other important measures, role-based access policies and multi-factor authentication.
- Encryption of data at rest in TPL databases and EBS volumes, using server-side encryption, should absolutely be implemented. **AWS Key Management Service** (AWS KMS) would be an ideal choice to implement this.
- **AWS Systems Manager Session Manager** would have been a more effective option than a bastion host for the production environment. It leverages existing AWS infrastructure and would not require TPL to manage additional EC2 instances, reducing the potential for misconfigurations, as well as the size of the attack surface. Systems manager uses IAM roles and policies for access control instead of a virtual machine and seamlessly integrates with CloudTrail, potentially increasing visibility of the system.
- Since Systems Manager communicates over port 443, this would allow TPL to **close the SSH port** on the production environment's public subnets, in order to further reduce the attack surface. Furthermore, the web application and on-premises networks should only accept **HTTPS** traffic, rather than HTTP. HTTPS traffic is encrypted using the TLS protocol and ensures the confidentiality of data in transit.
- **AWS CloudTrail** and **AWS CloudWatch** should be configured to provide logging and alerts capabilities. A **SIEM solution** should be set up on the on-premises data centre in order to concentrate and manage all logs in one central dashboard.
- The **least connections load balancing algorithm**, which directs incoming requests to the instance with the fewest active connections, would have been a more efficient choice than round robin. This algorithm is better suited for a case like ours where customers perform different types of queries and all the web servers are of equal capacity. To avoid single points of failure, a **backup load balancer** that activates if the main one fails would also be configured.

We are aware that moving part of TPL's infrastructure in the cloud comes with risks. The scalable nature of the web application servers, for instance, could mean heightened financial repercussions in case of a Distributed Denial of Service (DDoS) attack. To accommodate the rapid increase in traffic, the auto-scaling nature of the web application would lead to a multiplication of running web servers, resulting in additional costs for TPL. Protecting against DDoS attacks through the AWS Shield service could be considered, although round-the-clock monitoring and a well-made incident response plan could be sufficient to efficiently mitigate this risk.

Since all TPL staff, including the security team, are new to the cloud, ongoing cybersecurity education and training will be necessary to ensure a smooth transition and avoid accidental security incidents. The security team, in particular, will have to adapt quickly to this new technology and approaches.

Fundamental elements of network security were outside the scope of this project, including corporate security policies, as well as user education and awareness. We stress that, in the end, application security depends not only on its architecture, but also on its users; our secure architecture, on its own, is not enough to ensure the system's safety.

APPENDIX A: FULL ARCHITECTURE

AWS Cloud

CloudWatch

Region

Prod VPC: 10.0.0.0/16

Availability Zone # 1

Data Storage Layer

Business Logic Layer

Presentation Layer

Route table
10.0.3.0/24: Local
10.0.2.0/24: Private subnet 1
10.0.5.0/24: Private subnet 3

Route table
10.0.2.0/24: Local
10.0.3.0/24: Private subnet 2
10.0.1.0/24: Public subnet 1
0.0.0.0/0: NAT Gateway

Route table
10.0.1.0/24: Local
10.0.2.0/24: Private subnet 1
0.0.0.0/0: Internet Gateway
192.168.0.0/24: Transit Gateway

Private subnet 2: 10.0.3.0/24

Primary Database

KMS Encryption

Security group

Amazon RDS

Private subnet 1: 10.0.2.0/24

Auto Scaling group

Security group

Public subnet 1: 10.0.1.0/24

NAT Gateway

Security group

Web Tier Jump Box

Internet

Internet Gateway

Load Balancer

Availability Zone # 2

Private subnet 4: 10.0.6.0/24

Secondary Database

KMS Encryption

Security group

Amazon RDS

Private subnet 3: 10.0.5.0/24

Security group

Public subnet 2: 10.0.4.0/24

Security group

Web Tier Jump Box

NAT Gateway

Route table
10.0.6.0/24: Local
10.0.5.0/24: Private subnet 3

Route table
10.0.5.0/24: Local
10.0.3.0/24: Private subnet 2
10.0.4.0/24: Public subnet 2
0.0.0.0/0: NAT Gateway

Route table
10.0.4.0/24: Local
10.0.5.0/24: Private subnet 3
0.0.0.0/0: Internet Gateway
192.168.0.0/24: Transit Gateway

Non Prod VPC:172.19.0.0/24

Availability Zone

Private subnet: 172.19.0.0/25

Security group

Route table
10.0.0.0/25: Local
192.168.0.0/24: Transit Gateway

Transit Gateway

Site-to-Site VPN

On-Prem Network: 192.168.0.0/24

Private subnet: 192.168.0.128/25

Security group

Public subnet: 192.168.0.0/25

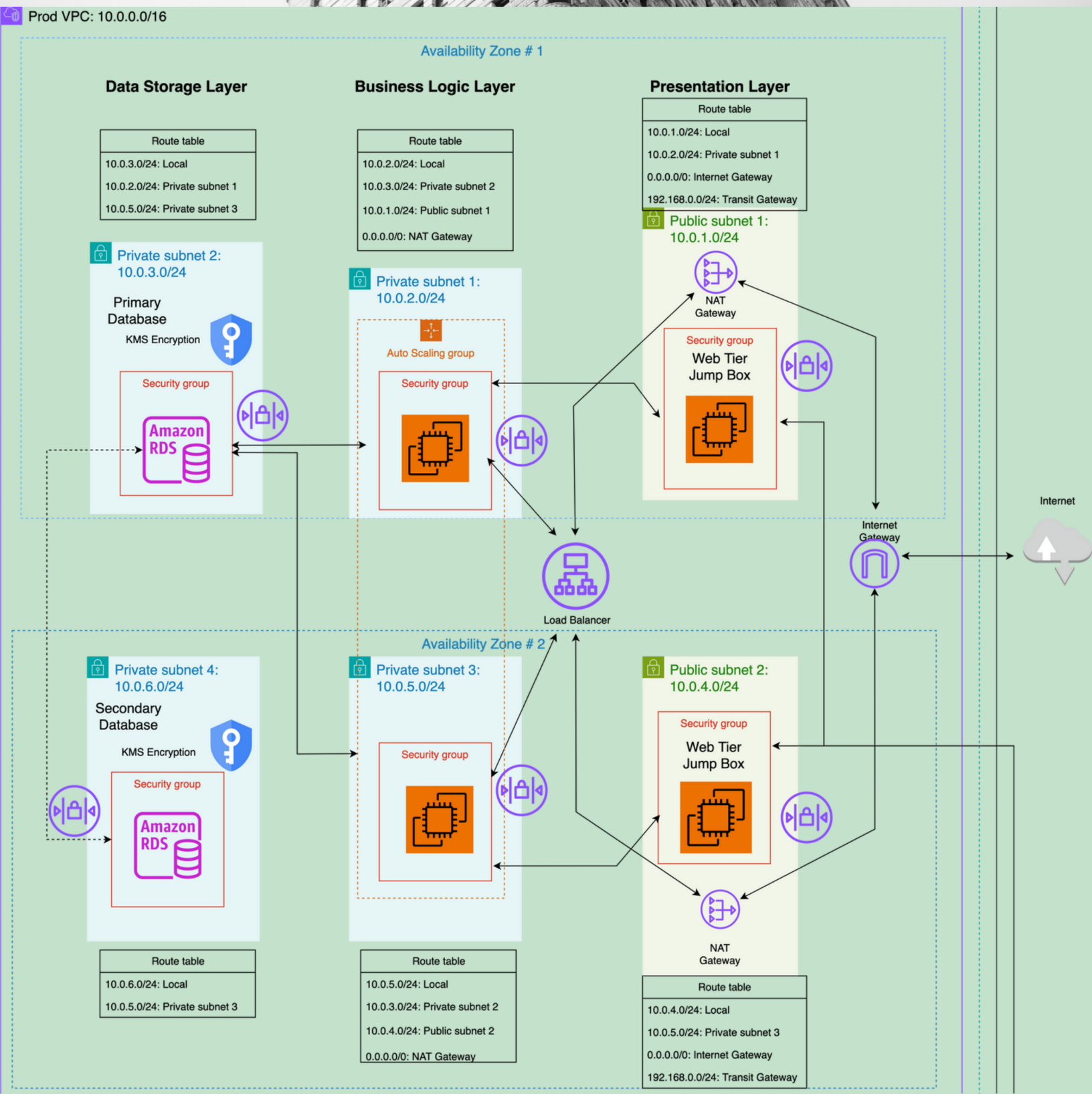
Route table
192.168.0.0/25: Local
0.0.0.0/0: Internet Gateway
10.0.0.0/16: Transit Gateway (to Production)
172.19.0.0/24: Transit Gateway (to Non Prod)

Customer Gateway

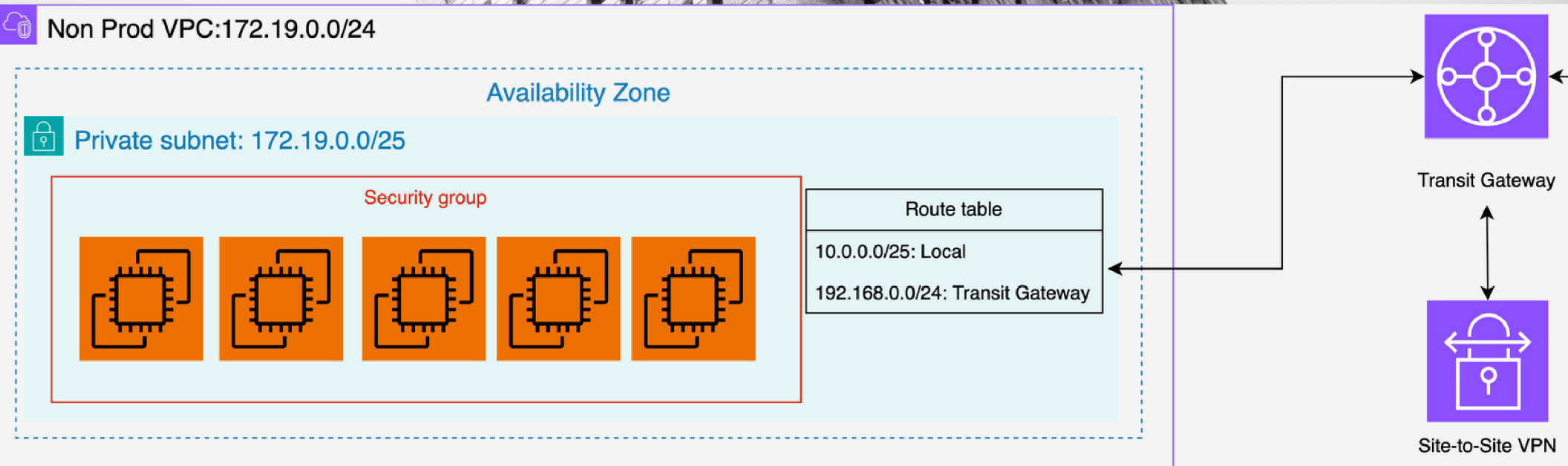
Internet Gateway

Internet

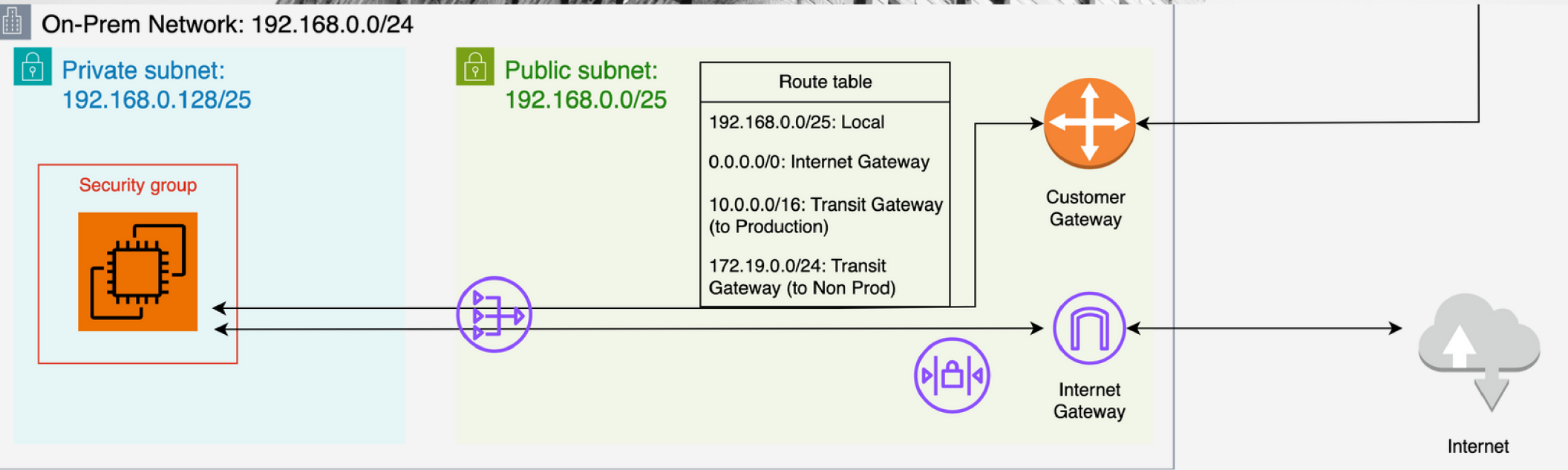
APPENDIX B:
PROD VPC



APPENDIX C:
NON PROD VPC



APPENDIX D:
ON PREM NETWORK



APPENDIX E:
NESSUS AGENT SCAN

NAME ↑	IPV4 ADDRESS	VULNERABILITIES	VULNERABILI...	CRITICAL	HIGH
agent-name	10.0.5.170		18	0	0
agent-name	10.0.2.181		18	0	0
agent-name	10.0.5.162		18	0	0
agent-name	10.0.2.79		18	0	0
agent-name	10.0.4.29		18	0	0
agent-name	10.0.1.240		18	0	0

SEVERITY	NAME	FAMILY	INSTANCES
Info	OS Identification	General	6
Info	Host Fully Qualified Domain Name (FQDN) Resolution	General	6
Info	Authenticated Check : OS Name and Installed Package Enumeration	Settings	6
Info	Netstat Portscanner (SSH)	Port scanners	6
Info	Nessus Scan Information	Settings	6
Info	Enumerate IPv6 Interfaces via SSH	General	6
Info	Enumerate IPv4 Interfaces via SSH	General	6
Info	Remote listeners enumeration (Linux / AIX)	Service detection	6
Info	BIOS Info (SSH)	General	6
Info	Common Platform Enumeration (CPE)	General	6
Info	Device Hostname	General	6
Info	Time of Last System Startup	General	6
Info	Netstat Connection Information	General	6
Info	Ethernet MAC Addresses	General	6
Info	OS Identification and Installed Software Enumeration over SSH v2 (...	Misc.	6
Info	OS Security Patch Assessment Not Available	Settings	6
Info	Unix Software Discovery Commands Available	Settings	6
Info	Enumerate the Network Interface configuration via SSH	General	6

0
CRITICAL VULNERABILITIES

0
HIGH VULNERABILITIES

0
MEDIUM VULNERABILITIES

0
LOW VULNERABILITIES

Scan Details

STATUS	Completed
START TIME	03/20/2024 at 4:46 PM
TEMPLATE	Basic Agent Scan

Agent Details

REPORTED
6 of 6

GROUPS
PriusAgents

APPENDIX F:
EXTERNAL SCAN

NAME ↑	IPV4 ADDRESS	VULNERABILITIES	VULNERABILI...	CRITICAL	HIGH
ec2-44-200-2...	44.200.225.1...		26	0	0

SEVERITY	NAME	FAMILY	INSTANCES
Info	RPC Services Enumeration	Service detection	2
Info	ICMP Timestamp Request Remote Date Disclosure	General	1
Info	RPC portmapper Service Detection	RPC	1
Info	SSH Server Type and Version Information	Service detection	1
Info	SSH Protocol Versions Supported	General	1
Info	Remote Desktop Protocol Service Detection	Service detection	1
Info	Nessus SYN scanner	Port scanners	1
Info	OS Identification	General	1
Info	Host Fully Qualified Domain Name (FQDN) Resolution	General	1
Info	Nessus Scan Information	Settings	1
Info	Service Detection	Service detection	1
Info	TCP/IP Timestamps Supported	General	1
Info	Backported Security Patch Detection (SSH)	General	1
Info	Common Platform Enumeration (CPE)	General	1
Info	RPC portmapper (TCP)	RPC	1
Info	Device Type	General	1
Info	Patch Report	General	1
Info	SSH Algorithms and Languages Supported	Misc.	1
Low	SSH Server CBC Mode Ciphers Enabled	Misc.	1
Info	Target Credential Status by Authentication Protocol - No Credentials Provided	Settings	1
Info	OS Security Patch Assessment Not Available	Settings	1
Info	SSH SHA-1 HMAC Algorithms Enabled	Misc.	1
Low	SSH Weak Key Exchange Algorithms Enabled	Misc.	1
Info	OpenSSH Detection	Misc.	1
Medium	SSH Terrapin Prefix Truncation Weakness (CVE-2023-48795)	Misc.	1

0
CRITICAL VULNERABILITIES

0
HIGH VULNERABILITIES

1
MEDIUM VULNERABILITIES

2
LOW VULNERABILITIES

Scan Details

STATUS	Completed
START TIME	03/22/2024 at 8:58 AM
TEMPLATE	Advanced Network Scan
SCANNER	US East Cloud Scanners

TARGETS
44.200.225.118

SCANNER GROUPS
US East Cloud Scanners