

Intelligent Cloud Infrastructure Review with GenAI Assistance

(AWS Cloud Security & Optimization Project)

Name: SUMITHA A

Reg. No.: 727823TUCY048

1. Executive Summary

Modern cloud environments generate a massive volume of infrastructure telemetry, configuration data, and operational logs. Manually analyzing this data to identify security risks, inefficiencies, and misconfigurations is time-consuming and error-prone. This project demonstrates an Intelligent Cloud Infrastructure Review using Amazon Web Services (AWS) combined with Generative AI (GenAI) in an advisory, human-in-the-loop role.

The objective of this project is to design, deploy, monitor, and review a containerized cloud environment while leveraging GenAI to:

- Summarize infrastructure behavior
- Identify security risks and misconfigurations
- Suggest optimization and hardening strategies

The environment consists of an Amazon EC2 instance running Docker, hosting an Nginx web application, deployed inside a custom Amazon VPC. Amazon CloudWatch Logs is used for centralized logging of system and container logs, which are later reviewed using GenAI for insights.

Importantly, GenAI is used only for analysis and decision support. It has no direct access to the AWS environment, ensuring security and compliance. This approach aligns with enterprise best practices and the AWS Well-Architected Framework.

2. Problem Statement

2.1 Background

Traditional cloud monitoring and observability tools primarily focus on collecting metrics, events, and raw log data from cloud resources. While these tools are effective at providing visibility, they often lack the contextual intelligence required to interpret the data and derive meaningful insights. As cloud environments grow in complexity, manually correlating logs and configurations across multiple services becomes increasingly difficult.

IT leadership and cloud administrators require smarter, high-level insights into key areas such as:

- Infrastructure usage and resource efficiency
- Security exposure and access control weaknesses
- Operational inefficiencies affecting performance and cost
- Misconfigurations across compute, networking, storage, and containerized workloads

Although Artificial Intelligence can assist in analyzing this data, fully automated AI systems with direct write access to production infrastructure introduce significant security, compliance, and governance risks. Therefore, a controlled, advisory approach is necessary, where AI supports decision-making while critical actions remain under human oversight.

2.2 Problem Definition

The core challenge addressed in this project is the design of a GenAI-assisted cloud infrastructure review system that enhances visibility and decision-making without compromising security or control. The system must be capable of analyzing infrastructure logs and configuration data collected from cloud resources to extract meaningful operational and security insights.

Specifically, the solution should be able to:

- Analyze system and application logs along with configuration metadata
- Identify risky patterns, security gaps, and misconfigurations
- Suggest actionable improvements and optimization measures
- Maintain full human control by ensuring GenAI operates strictly in an advisory role

This approach balances the analytical strengths of Generative AI with the need for human oversight, making it suitable for real-world enterprise cloud environments.

2.3 Scope of the Project

The scope of this project is limited to the analysis and review of a controlled AWS-based cloud environment using GenAI as a decision-support tool. The project focuses on evaluating infrastructure behavior, security posture, and operational efficiency based on logs and configuration data.

The key areas covered within the scope include:

- Reviewing AWS infrastructure logs and configuration settings related to compute, networking, and containers
- Identifying potential security risks, misconfigurations, and operational inefficiencies
- Utilizing GenAI strictly for advisory analysis, insight generation, and recommendation purposes
- Emphasizing documentation, analysis, and reporting rather than automated remediation
- Ensuring that no real production workloads, sensitive information, or confidential data are used

By clearly defining these boundaries, the project maintains security, compliance, and ethical use of AI while effectively demonstrating GenAI-assisted cloud infrastructure review capabilities.

3. Objectives of the Project

The primary objective of this project is to demonstrate how Generative AI can be effectively integrated into cloud infrastructure reviews while maintaining security and human oversight. To achieve this, the project is guided by the following specific objectives:

1. To design and implement a secure AWS cloud architecture that follows industry best practices
2. To deploy a containerized web application using Docker on an Amazon EC2 instance
3. To enable centralized logging and monitoring using Amazon CloudWatch Logs
4. To review infrastructure behavior and log data using GenAI for analytical insights
5. To identify security misconfigurations and operational risks within the environment
6. To propose security hardening and optimization recommendations based on the review
7. To document the complete implementation, analysis, and findings with supporting evidence

These objectives ensure a structured approach to building, analyzing, and documenting a GenAI-assisted cloud infrastructure review.

4. High-Level Architecture

4.1 Architecture Overview

The cloud environment is designed using a modular and isolated architecture, ensuring clear separation between compute, networking, storage, and monitoring components. This approach improves security, scalability, and manageability by allowing each layer of the infrastructure to be independently monitored and controlled. The architecture follows AWS best practices and supports effective logging and analysis for GenAI-assisted reviews.

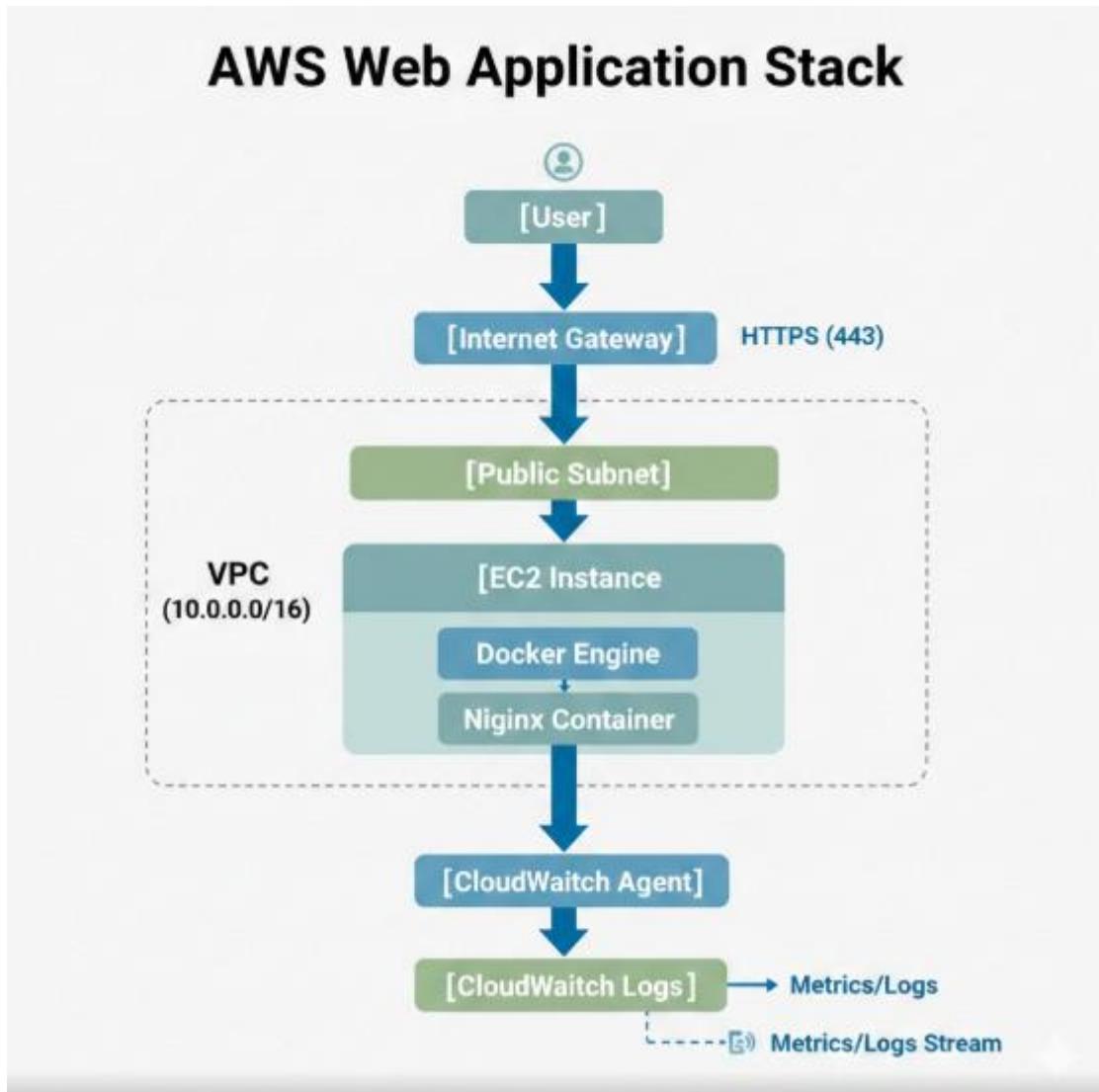
4.2 Architecture Components

The major components of the architecture are described below:

- **Amazon Virtual Private Cloud (VPC)**
 - CIDR Block: 10.0.0.0/16
 - Provides an isolated virtual network for hosting cloud resources
- **Public Subnet**
 - Hosts the EC2 instance running the containerized application
 - Enables controlled access to the application through defined security rules
- **Amazon EC2 (t2.micro)**
 - Acts as the primary compute resource
 - Runs the Docker Engine and hosts the Nginx container
- **Docker (Container Runtime)**
 - Provides a lightweight and portable environment for running the Nginx web server
 - Simplifies application deployment and management
- **Amazon CloudWatch Logs**
 - Collects and centralizes system and container logs

- Enables monitoring, troubleshooting, and analysis
- **Amazon S3 (Optional)**
 - Used for long-term archival of logs and audit data
 - Supports cost-effective storage and compliance requirements

4.3 Architecture Diagram

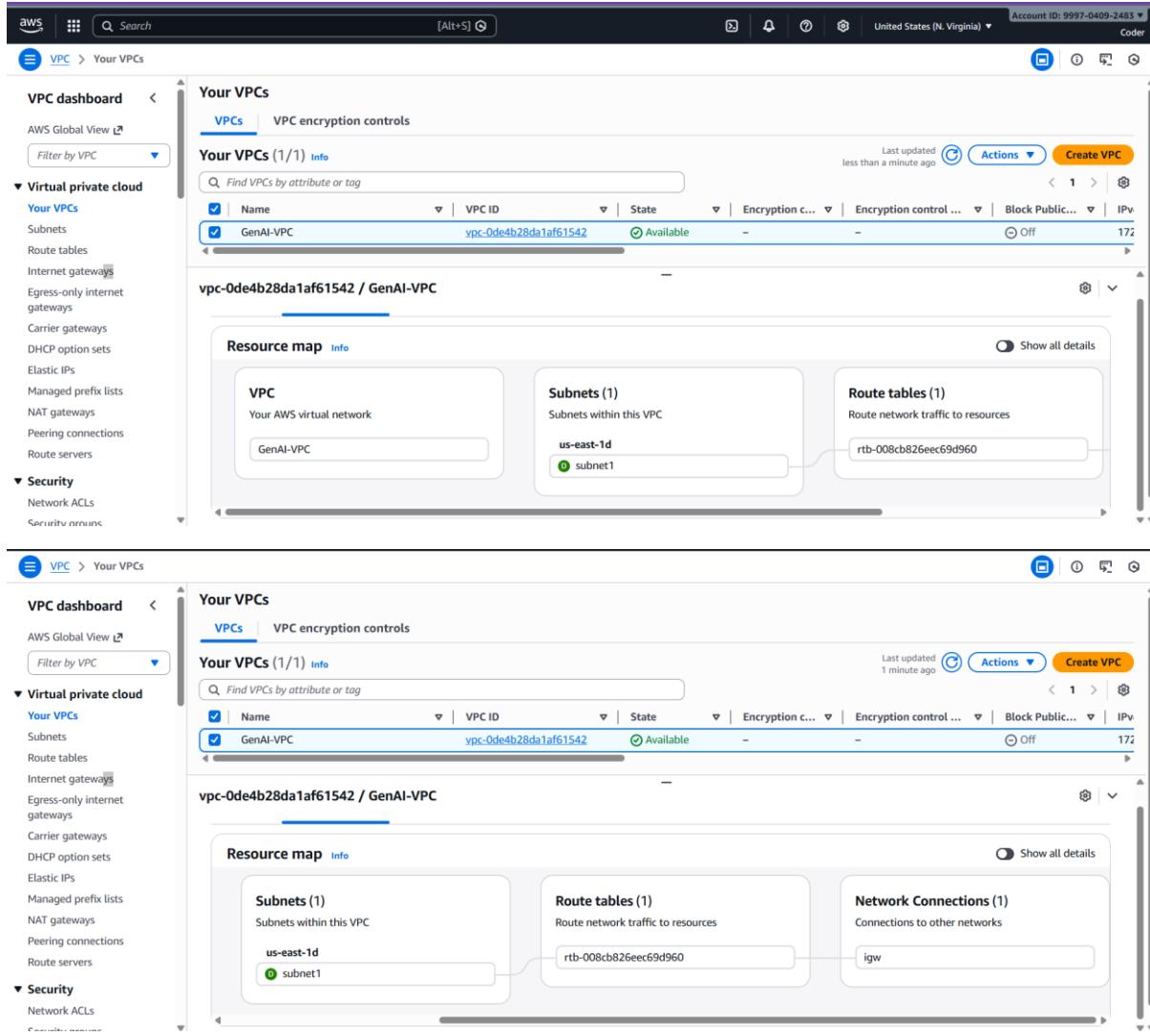


5. Infrastructure Implementation

5.1 VPC and Network Configuration

A custom Amazon VPC was created to provide network isolation for the application environment. Public subnets, route tables, and an Internet Gateway were configured to enable controlled external access. Security Groups acted as virtual firewalls, allowing only required inbound and outbound traffic.

VPC Resource Map:



5.2 EC2 Instance Setup

An Amazon EC2 instance running Amazon Linux 2023 was launched within the public subnet. Docker was installed using user-data scripts to ensure consistency and repeatability across deployments.

EC2 Instance Summary:

The screenshot shows the AWS EC2 Instances page. On the left, there's a sidebar with navigation links like Dashboard, EC2 Global View, Events, Instances (selected), Instance Types, Launch Templates, Spot Requests, Savings Plans, Reserved Instances, Dedicated Hosts, Capacity Reservations, Capacity Manager, Images (AMIs, AMI Catalog), and Elastic Block Store (Volumes). The main area displays 'Instances (1/1) info' for a single instance named 'genai-ec2'. The instance details include its ID (i-02454a89625ccb973), state (Running), type (t3.micro), and various metadata fields like Public IPv4 (54.145.129.244), Private IPv4 (172.31.11.170), and Public DNS (ec2-54-145-129-244.compute-1.amazonaws.com).

5.3 Docker and Application Deployment

Docker was configured on the EC2 instance to run an Nginx container. The container exposes port 80, enabling HTTP access to the web application and validating successful container deployment.

Docker Container Running:

```

Installed:
  container-selinux-4:2.242.0-1.amzn2023.noarch
  iptables-libc-1.8-8-3.amzn2023.0.2.x86_64
  libnftnl_comtrack-1.0-8-2.amzn2023.0.2.x86_64
  pigz-2.5-1.amzn2023.0.3.x86_64

containerd-2.1.5-1.amzn2023.0.4.x86_64
  iptables-nft-1.8-8-3.amzn2023.0.2.x86_64
  libnftnl-1.0.1-19.amzn2023.0.2.x86_64
  runc-1.3.4-1.amzn2023.0.1.x86_64

docker-25.0.14-1.amzn2023.0.1.x86_64
  libcgroup-3.0-1.amzn2023.0.1.x86_64
  libnftnl-1.2.2-2.amzn2023.0.2.x86_64

Complete!
[root@ip-172-31-11-170 ec2-user]# systemctl start docker
[root@ip-172-31-11-170 ec2-user]# systemctl enable docker
Created symlink /etc/systemd/system/multi-user.target.wants/docker.service → /usr/lib/systemd/system/docker.service.
[root@ip-172-31-11-170 ec2-user]# usermod -aG docker ec2-user
[root@ip-172-31-11-170 ec2-user]# docker run -d -p 80:80 nginx
Unable to find image 'nginx:latest' locally
latest: Pulling from library/nginx
0cd55a45c0d1: Pull complete
46f3a120c8e: Pull complete
414fe020542: Pull complete
7b6c8ccac7b: Pull complete
f73400a233fd: Pull complete
47cd406a84ef: Pull complete
bae5a1799a80: Pull complete
Digest: sha256:341bf30c5c5277d6002cf6e1fb0319fa4252add24ab6a0e262e0056d313208
Status: Downloaded newer image for nginx:latest
619d5502069cc220673067c97e789b2b36f00ca8ffabdel08ae56527237ab091
[root@ip-172-31-11-170 ec2-user]# [ ]

```

i-02454a89625ccb973 (genai-ec2)
PublicIPs: 54.145.129.244 PrivateIPs: 172.31.11.170



6. Logging and Monitoring Configuration

6.1 Need for Centralized Logging

In cloud-native environments, logs are generated across multiple layers such as operating systems, applications, containers, and network components. Centralized logging consolidates these logs into a single platform, making analysis and monitoring more efficient.

Centralized logging enables:

- Faster troubleshooting by correlating events across services
- Improved security auditing and forensic analysis
- Easier compliance reporting and log retention
- Effective AI-assisted analysis by providing structured and complete data

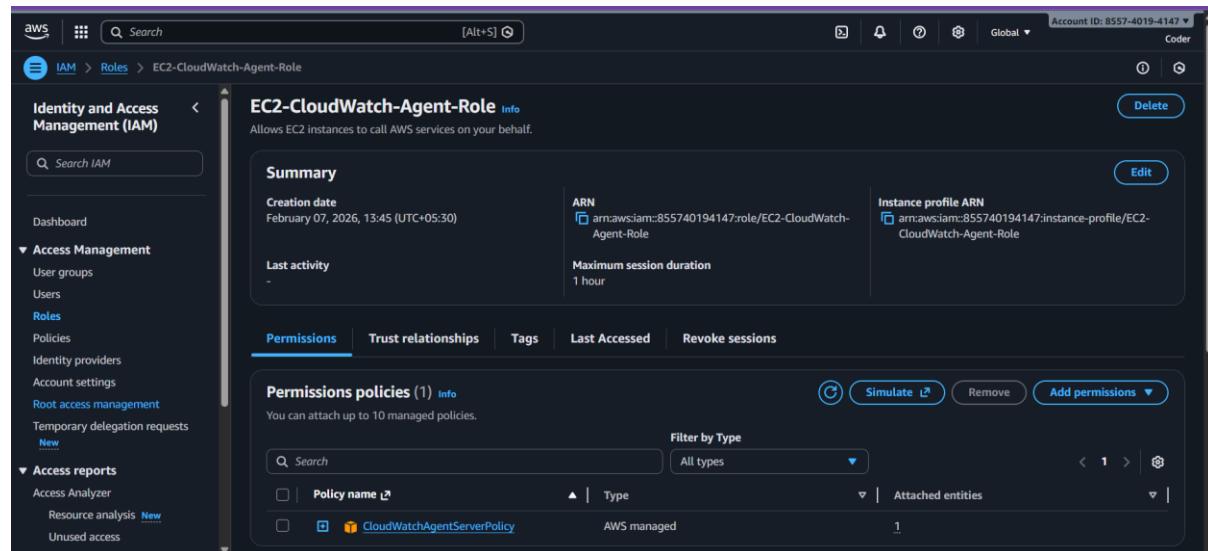
Amazon CloudWatch Logs was selected as it integrates seamlessly with AWS services and supports scalable log ingestion and analysis.

6.2 CloudWatch Agent Configuration

The Amazon CloudWatch Agent was installed and configured on the EC2 instance to collect system-level and application logs. Since Amazon Linux 2023 uses journald by default, rsyslog was enabled to forward logs into /var/log/messages, ensuring compatibility with the CloudWatch Agent configuration.

The agent configuration defined log file paths, log group names, and stream naming conventions based on the EC2 instance ID. This ensured logs were reliably collected and easily traceable to the source instance.

CloudWatch Agent Running:



Screenshot of the AWS EC2 Instances page showing a single instance named "genai-ec2".

Instances (1/1) Info

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IP
genai-ec2	i-02454a89625ccb973	Running	t3.micro	3/3 checks passed	View alarms +	us-east-1a	ec2-54-145-

i-02454a89625ccb973 (genai-ec2)

- Details
- Status and alarms
- Monitoring
- Security**
- Networking
- Storage
- Tags

Security details

IAM role EC2-CloudWatch-Agent-Role	Owner ID 855740194147	Launch time Sat Feb 07 2026 10:41:59 GMT+0530 (India Standard Time)
Security groups sg-05164ed1ee21bf30 (launch-wizard-1)		

Release notes:
<https://docs.aws.amazon.com/linux/al2023/release-notes/relnotes-2023.10.20260202.html>

```
Installed:
amazon-cloudwatch-agent-1.300062.1-1.amzn2023.x86_64

Complete!
[root@ip-172-31-11-170 ec2-user]# sudo nano /opt/aws/amazon-cloudwatch-agent/bin/config.json
[root@ip-172-31-11-170 ec2-user]# sudo /opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent-ctl \
-a fetch-config \
-m ec2 \
-c file:/opt/aws/amazon-cloudwatch-agent/bin/config.json \
-s
***** processing amazon-cloudwatch-agent *****
Starting config-downloader, this will map back to a call to amazon-cloudwatch-agent
Executing /opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent with arguments: [config-downloader -output-dir /opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.d -config /opt/aws/amazon-cloudwatch-agent/etc/common-config.toml -multi-config default -mode ec2 -download-source file:/opt/aws/amazon-cloudwatch-agent/bin/config.json]! Trying to detect region from ec2
D! No bound active network interface
I! imds retry client will retry 1 times
Start configuration validation...
2026/02/07 08:30:09 Reading json config file path: /opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.d/file_config.json.tmp ...
2026/02/07 08:30:09 I! Valid Json input schema.
2026/02/07 08:30:09 Configuration validation first phase succeeded
Starting config-translator, this will map back to a call to amazon-cloudwatch-agent
Executing /opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent with arguments: [config-translator -input /opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.d/file_config.json ...]
```

i-02454a89625ccb973 (genai-ec2)

PublicIPs: 54.145.129.244 PrivateIPs: 172.31.11.170

```
I! imds retry client will retry 1 times
/opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent -schematest -config /opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.toml
Configuration validation second phase succeeded
amazon-cloudwatch-agent has already been stopped
Created symlink /etc/systemd/system/multi-user.target.wants/amazon-cloudwatch-agent.service → /etc/systemd/system/amazon-cloudwatch-agent.service.
[root@ip-172-31-11-170 ec2-user]# sudo systemctl status amazon-cloudwatch-agent
● amazon-cloudwatch-agent.service - Amazon CloudWatch Agent
   Loaded: loaded (/etc/systemd/system/amazon-cloudwatch-agent.service; enabled; preset: disabled)
   Active: active (running) since Sat 2026-02-07 08:30:10 UTC; 12s ago
     Main PID: 33309 (amazon-cloudwatch)
      Tasks: 7 (limit: 1067)
        Memory: 53.3M
         CPU: 293ms
        CGroup: /system.slice/amazon-cloudwatch-agent.service
                  └─33309 /opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent -config /opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.toml -envconfig /etc/amazon-cloudwatch-agent-envconfig.json

Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33314]: Executing /opt/aws/amazon-cloudwatch-agent/bin/amazon-cloudwatch-agent with arguments: [config /opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.d/file_config.json ...]
Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33320]: D! [EC2] Found active network interface
Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33320]: I! imds retry client will retry 1 times! Detected the instance is EC2
Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33320]: 2026/02/07 08:30:10 Reading json config file path: /opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.json does not exist or is empty
Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33320]: 2026/02/07 08:30:10 Reading json config file path: /opt/aws/amazon-cloudwatch-agent/etc/amazon-cloudwatch-agent.d/file_config.json ...
Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33320]: 2026/02/07 08:30:10 I! Valid Json input schema.
Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33320]: I! Trying to detect region from ec2
Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33320]: 2026/02/07 08:30:10 Configuration validation first phase succeeded
Feb 07 08:30:10 ip-172-31-11-170.ec2.internal start-amazon-cloudwatch-agent[33309]: I! Detecting run_as_user...
```

i-02454a89625ccb973 (genai-ec2)

PublicIPs: 54.145.129.244 PrivateIPs: 172.31.11.170

6.4 Amazon S3 Log Archival Configuration

To support long-term log retention and cost-effective storage, an Amazon S3 bucket was created for archiving CloudWatch Logs. While CloudWatch is suitable for real-time monitoring and short-term analysis, S3 provides durable, scalable, and low-cost storage for historical data.

The S3 bucket was configured with:

- Private access enabled by default
- Server-side encryption (SSE-S3) for data at rest
- Block Public Access enabled to prevent accidental exposure

CloudWatch Logs can be exported periodically to the S3 bucket for compliance audits, forensic investigations, and offline analysis using analytics tools such as Amazon Athena.

Amazon S3 Bucket Configuration:

The screenshot shows the AWS S3 Bucket Properties page for the 'genai-log-archive-demo' bucket. The top navigation bar includes the AWS logo, search bar, and account information (Account ID: 8557-4019-4147). The main content area has tabs for Objects, Metadata, **Properties**, Permissions, Metrics, Management, and Access Points. The Properties tab is selected. Under the Bucket overview section, it shows the AWS Region as US East (N. Virginia) us-east-1, the Amazon Resource Name (ARN) as arn:aws:s3:::genai-log-archive-demo, and the Creation date as February 7, 2026, 14:47:49 (UTC+05:30). The Bucket Versioning section indicates it is disabled. The Multi-factor authentication (MFA) delete section also indicates it is disabled. The Bucket ABAC section notes that Attribute-based access control (ABAC) is an authorization strategy that defines permissions based on attributes. The page includes standard AWS navigation icons like back, forward, and refresh.

6.5 Benefits of Using Amazon S3 for Log Storage

Using Amazon S3 for log archival offers several advantages:

- Long-term retention of logs beyond CloudWatch limits
- Reduced operational cost for older log data
- High durability and availability
- Easy integration with analytics and AI-based review tools

This approach ensures that critical infrastructure and security logs remain accessible for future audits and GenAI-assisted analysis without impacting real-time system performance.

6.3 Log Groups Created

Two primary CloudWatch Log Groups were created for structured monitoring:

- **/ec2/system/logs** → Captures operating system and service logs
- **/ec2/docker/logs** → Captures Docker container logs

Log streams were automatically generated using the EC2 instance ID, enabling clear identification of log sources and simplifying multi-instance scalability.

7. Log Verification and Evidence

7.1 System Logs

System logs from /var/log/messages were successfully ingested into Amazon CloudWatch Logs. These logs include system boot events, service status updates, and security-related messages. The availability of these logs confirms correct agent configuration and network connectivity.

System Log Stream:

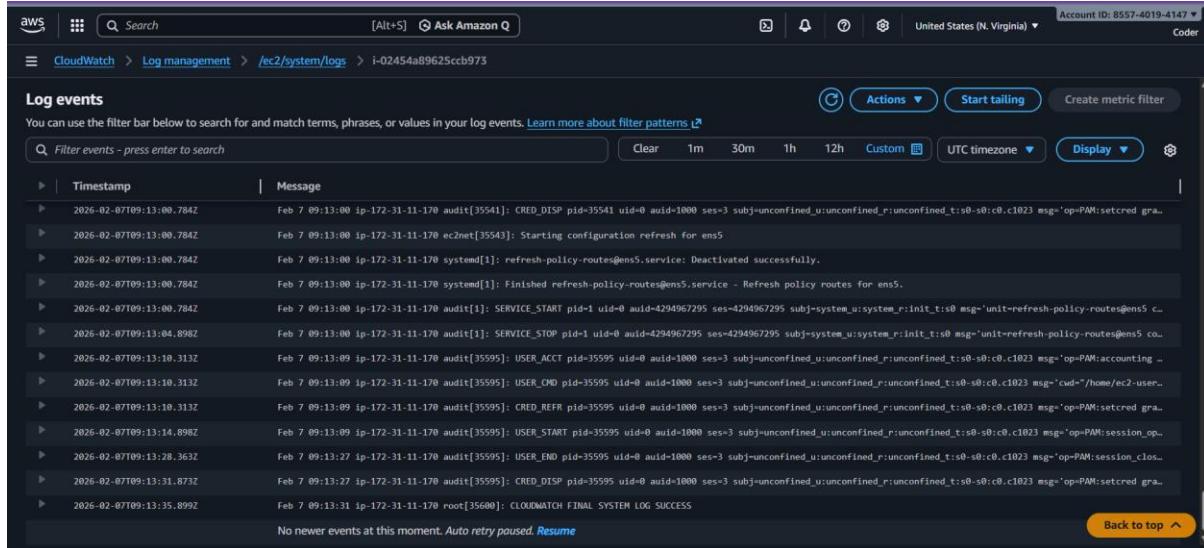
The screenshot shows the AWS CloudWatch Log Management interface. On the left, there's a sidebar with navigation links like 'CloudWatch', 'Logs', 'Log Management New', and 'Contributor Insights'. The main area is titled 'Log groups (2)' and shows two entries: '/ec2/system/logs' and '/ec2/docker/logs'. Both entries are listed under the 'Standard' log class, with 'Configure' options and 'Off' for deletion. A green banner at the top right says 'Log group "/ec2/docker/logs" has been created.' Below the log groups, there's a search bar, filter options, and buttons for 'Actions', 'View in Logs Insights', 'Start tailing', and 'Create log group'.

This screenshot shows the log stream for the '/ec2/docker/logs' group. The interface includes a header with 'Account ID: 8557-4019-4147' and various filtering and time selection tools. The main content area displays a list of log entries with columns for 'Timestamp' and 'Message'. The log entries are timestamped between February 2026 and March 2026, showing various system and application logs. At the bottom, it says 'No newer events at this moment. Auto retry paused. Resume'.

7.2 Docker Logs

Docker container logs were collected from `/var/lib/docker/containers/*/*.log`. These logs capture application-level events generated by the Nginx container, including startup messages and HTTP request handling. This ensures visibility into container behavior and application health.

Docker Log Stream:



The screenshot shows the AWS CloudWatch Log Stream interface. The top navigation bar includes the AWS logo, search bar, and account information (Account ID: 8557-4019-147). Below the navigation is a breadcrumb trail: CloudWatch > Log management > /ec2/system/logs > i-02454a89625ccb975. The main area is titled "Log events" and contains a table with two columns: "Timestamp" and "Message". The table lists numerous log entries from February 7, 2026, at 09:13:00 UTC. The log entries detail various system events such as configuration refreshes, audit logs for services like auditd and CRED_DISP, and system start/stop events. One entry indicates a successful audit log collection by CloudWatch. The interface includes various filters and time range controls at the top right.

8. GenAI-Assisted Security Review

8.1 Methodology

Infrastructure metadata, architectural details, and summarized log outputs were manually provided to GenAI (ChatGPT) for analysis. GenAI was used strictly in an advisory, human-in-the-loop role to:

- Identify risky configurations
- Highlight deviations from AWS best practices
- Suggest security and optimization improvements

GenAI had no direct access to the AWS environment, ensuring full human control and eliminating the risk of unintended automated changes.

8.2 Findings and Risk Analysis

Based on the review, the following key findings were identified:

Category	Finding	Risk Level	GenAI Recommendation
Network	SSH (port 22) open to 0.0.0.0/0	High	Restrict IP range or use SSM
IAM	Broad permissions assigned	Medium	Apply least-privilege policies
Logging	Logs manually reviewed	Low	Enable alerts and automation

Security Group Inbound Rules:

The screenshot shows the AWS EC2 Launch Wizard interface. On the left, under 'Inbound Security Group Rules', there is one rule: 'Security group rule 1 (TCP, 22, 0.0.0.0/0)'. The rule details are: Type: ssh, Protocol: TCP, Port range: 22, Source type: Anywhere. A note below the rule states: '⚠️ Rules with source of 0.0.0.0/0 allow all IP addresses to access your instance. We recommend setting security group rules to allow access from known IP addresses only.' On the right, the 'Summary' section shows: Number of instances: 1; Software Image (AMI): Amazon Linux 2023 AMI 2023.10...; Virtual server type (instance type): t3.micro; Firewall (security group): New security group; Storage (volumes): 1 volume(s) - 8 GiB. At the bottom right are 'Cancel' and 'Launch instance' buttons.

9. Security Hardening Recommendations

9.1 Identity and Access Management

To strengthen access control, the following IAM best practices are recommended:

- Enable Multi-Factor Authentication (MFA) for all IAM users
- Use IAM roles instead of long-term access keys
- Apply least-privilege permissions to reduce attack surface

9.2 Network Security

Network-level protections significantly reduce exposure to attacks:

- Restrict SSH access to specific IP addresses
- Replace SSH access with AWS Systems Manager Session Manager
- Enable VPC Flow Logs to monitor network traffic patterns

9.3 Container Security

Securing containers is essential in modern cloud deployments:

- Store images in Amazon ECR with vulnerability scanning enabled
- Avoid running containers with root privileges
- Regularly update base images and dependencies

9.4 Logging and Monitoring Enhancements

Further improvements to observability include:

- Creating CloudWatch Alarms for error patterns and anomalies
- Archiving logs to Amazon S3 with lifecycle policies
- Enabling Amazon GuardDuty for continuous threat detection

10. Benefits of GenAI Integration

Integrating GenAI into cloud reviews provides multiple benefits:

- Faster analysis of large and complex log datasets
- Reduced human error during manual reviews
- Improved decision-making through contextual insights
- Scalable advisory support for growing environments

GenAI acts as a security and operations sidekick, enhancing human expertise rather than replacing it.

11. Limitations of the Approach

Despite its benefits, the approach has certain limitations:

- GenAI analysis depends heavily on the quality of provided inputs
- No real-time or automated remediation is performed
- Requires skilled engineers to interpret and act on recommendations

12. Conclusion

This project successfully demonstrates an Intelligent Cloud Infrastructure Review using AWS services combined with GenAI assistance. By implementing centralized logging, secure architecture, and AI-driven analysis, the system delivers actionable security and operational insights while maintaining full human control.

The project aligns closely with enterprise-grade practices and the AWS Well-Architected Framework, proving that GenAI can significantly enhance cloud governance when used responsibly.

13. Future Enhancements

Future improvements to this project may include:

- Integrating AWS Config for continuous compliance monitoring
- Using Amazon Athena for advanced log analytics
- Adding automated alerting and incident workflows
- Extending the solution to multi-account AWS environments