The State of AWS Security 2026: From IAM to Agentic AI

A practitioner's guide to securing cloud infrastructure in the age of autonomous AI agents

Author: Tarek Cheikh, Founder & AWS Cloud Architect, Toc Consulting
Published: February 2026
Website: tocconsulting.fr

About the Author

Tarek Cheikh is the founder of Toc Consulting, an AWS Security and Cloud Architecture consultancy based in Paris, serving clients worldwide. With deep expertise in IAM security audits, S3 hardening, compliance automation, and security first architecture, Tarek helps organizations secure their AWS infrastructure against modern threats. He is the creator of open-source AWS security tools including the S3 Security Scanner and IAM Activity Tracker, and the architect behind KloudSec, a Cloud Security Suite combining CSPM, Infrastructure as Code scanning, secrets detection, and more into a single platform for securing AWS environments at scale.

Executive Summary

Cloud security is not a checklist. It is a continuous, high-stakes discipline that determines whether organizations thrive or become the next headline.

AWS remains the leading cloud infrastructure provider in 2025. Its portfolio spans compute, storage, AI/ML, analytics, and beyond, and the attack surface available to adversaries has never been larger.

The numbers are stark. According to a 2022 Snyk "State of Cloud Security" survey, 80% of organizations reported at least one severe cloud security incident, and the proportion reporting significant cloud incidents surged 154% year-over-year (from 24% in 2023 to 61% in 2024) per the Check Point / Cybersecurity Insiders 2024 Cloud Security Report. Misconfiguration remains the dominant root cause: Gartner projected that through 2025, 99% of cloud security failures would be the customer's fault, a prediction that holds true in 2026. IBM's 2025 Cost of a Data Breach Report puts the global average cost at $4.44 million, with a 241-day average lifecycle to identify and contain, a nine-year low driven by security AI and automation.

Recent breaches underscore the stakes. The 2019 Capital One breach (an SSRF attack exploiting an overly permissive IAM role) exposed the personal data of over 100 million individuals. The 2021 Twitch leak, a server misconfiguration, exposed 125 GB of internal data including proprietary SDKs, internal AWS services, and creator payout records. And in January 2025, the Codefinger ransomware campaign turned AWS's own SSE-C encryption against its customers, making recovery without the attacker's key impossible (Chapter 1).

2026 brings a threat that none of these precedents fully prepared us for: **Agentic AI**.

Autonomous AI agents (systems that plan, decide, and execute actions without human intervention) are deploying on AWS at an accelerating pace. Amazon Bedrock Agents, custom LangChain deployments on Lambda, and enterprise-built agent frameworks grant AI systems direct access to AWS APIs, databases, and infrastructure.

Traditional IAM assumes human intent: a user requests access, and policies determine whether to grant it. An AI agent decides its own actions autonomously. How do you apply least privilege to a system that determines its own next step? How do you audit decisions made at machine speed? How do you prevent an agent hijacked through prompt injection from executing unauthorized API calls?

In December 2025, OWASP released the **Top 10 for Agentic Applications**, developed by over 100 security researchers, identifying critical risks including Agent Goal Hijack (ASI01), Tool Misuse and Exploitation (ASI02), and Rogue Agents (ASI10). A Dark Reading poll found 48% of cybersecurity professionals identify agentic AI as the number-one attack vector heading into 2026, outranking deepfake threats, board-level cyber recognition, and passwordless adoption.

Yet a Gartner Peer Community survey (April 2023, published September 2023) found only 34% of organizations were using or implementing AI application security tools, a figure unlikely to have closed the gap given the explosive growth of agentic AI since then.

This whitepaper bridges that gap.

Eight chapters cover the full spectrum of AWS security in 2026: foundational IAM and identity management, data protection, network security, monitoring, and the frontier challenge of securing autonomous AI agents on AWS. Every recommendation is grounded in real AWS services, real CLI commands, and real architecture patterns. No marketing fluff. No theoretical hand-waving. The security controls you need to implement, the commands to implement them, and the reasoning behind each decision.

Whether you are a CTO evaluating your cloud security posture, an architect designing a new AWS environment, or a security engineer hardening an existing one, this guide gives you the practical knowledge to act today.

This whitepaper is published by Toc Consulting, AWS Security & Cloud Architecture. For a free AWS security assessment, visit tocconsulting.fr/contact.

Table of Contents

Toc Consulting · AWS Security & Cloud Architecture

tocconsulting.fr | @TocConsulting

Chapter 1: The 2026 AWS Threat Landscape

The Evolution of Cloud Attacks

Cloud security transformed between 2020 and 2026. What began as perimeter defense and access control became a multi-front battle against adversaries who weaponize the cloud's own features, exploit supply chain trust, and, as of 2025, use AI to attack faster than any human team can respond.

The trajectory is clear. In the early 2020s, the dominant attack pattern was straightforward: find a misconfigured resource, exfiltrate data. The 2019 Capital One breach is the textbook example: an SSRF attack exploited an overly permissive IAM role on the EC2 instance running a misconfigured WAF, reaching the EC2 metadata service (IMDSv1) to obtain temporary credentials. The attacker exfiltrated the personal data of over 100 million individuals. The root cause was not a zero-day exploit. It was a misconfiguration that granted a single role far more access than it needed.

By 2021, the threat expanded to insider exposure and secrets sprawl. The Twitch source code leak showed the cascading impact of a single server misconfiguration: 125 GB of internal data exposed, including proprietary SDKs, internal AWS services, and creator payout records. GitGuardian researchers found approximately 6,600 secrets including 194 AWS access keys embedded in the leaked git repositories. Not a sophisticated exploit, but an internal configuration error that exposed an entire codebase.

By 2024, attackers targeted the human layer at scale. The Snowflake breach campaign (Mandiant identified approximately 165 potentially exposed customer organizations, including AT&T (call and text metadata for nearly 110 million customers), Ticketmaster (up to 560 million records per attacker claims; unconfirmed by Live Nation), and Santander Bank (customer and employee data across three countries; the attacker claimed 30 million records, unconfirmed by Santander)) used credentials stolen from infostealer malware infections dating as far back as November 2020. Every affected account Mandiant investigated lacked multi-factor authentication. The attack required no vulnerability exploitation: just a valid username and password.

And then 2025 changed the game entirely.

2025: The Year Cloud Attacks Became Cloud-Native

The Codefinger ransomware campaign, revealed by Halcyon researchers in January 2025, introduced something genuinely new: an attack that turned AWS's own infrastructure against its customers. Attackers used compromised AWS credentials with `s3:GetObject` and `s3:PutObject` permissions to overwrite S3 objects, re-encrypting them with SSE-C (Server-Side Encryption with Customer-Provided Keys) using attacker-controlled AES-256 keys, then marking files for automatic deletion within seven days via the S3 Object Lifecycle Management API. Ransom notes demanded Bitcoin payment.

Codefinger's recovery model set it apart from every previous ransomware variant. AWS does not retain SSE-C encryption keys. CloudTrail logs record only the HMAC of the key used, not the key itself. Decryption without the attacker's key was impossible. Organizations without independent backups (versioned buckets, cross-region replication, or off-site copies) had no path to recovery.

A separate discovery compounded the risk: security researchers found a publicly accessible server containing over 158 million AWS secret key records, pointing to 1,229 unique credentials, many already rotated, but some still active. A related SSE-C ransomware operation demanded 0.3 BTC per victim. AWS responded: beginning April 6, 2026, SSE-C will be disabled by default for all new S3 general-purpose buckets and for all existing buckets in AWS accounts that do not have any SSE-C encrypted data.

Codefinger was only the beginning of what 2025 would bring.

AI-Powered Attacks Arrive

On November 28, 2025, the Sysdig Threat Research Team documented the clearest example yet of AI-assisted cloud intrusion: an attacker achieved administrative access to an AWS environment in under 10 minutes.

The entry point: valid AWS credentials in publicly accessible S3 buckets used to store RAG (Retrieval-Augmented Generation) data for AI models, a new class of credential exposure tied directly to the AI workload boom. From initial access, the attacker escalated privileges through Lambda function code injection, using `UpdateFunctionCode` and `UpdateFunctionConfiguration` permissions to modify existing Lambda functions. Within minutes, the attacker accessed 19 unique AWS principals (6 IAM roles assumed across 14 separate sessions, plus 5 IAM users) and created a backdoor admin user with `AdministratorAccess`.

The attacker's objectives went beyond data theft. They abused Amazon Bedrock access for LLMjacking, invoking nine foundation models including Claude, DeepSeek R1, Llama 4 Scout, Amazon Nova Premier, Amazon Titan Image Generator, and Cohere Embed v3, all at the victim's expense, and spun up a p4d.24xlarge GPU instance (approximately $23,600 per month per the Sysdig report) with a JupyterLab backdoor for persistent compute access.

The evidence of AI assistance was unmistakable: LLM-generated code contained Serbian-language comments, hallucinated AWS account IDs, and references to non-existent GitHub repositories, all telltale signs of large language model output. AWS confirmed its services operated as designed throughout the incident. The vulnerability was entirely in the customer's configuration.

The Supply Chain Threat Multiplier

Supply chain attacks against AWS environments escalated sharply in 2025. The Verizon 2025 Data Breach Investigations Report found that third-party involvement in breaches doubled from 15% to 30% year-over-year.

Three incidents from the past year illustrate the pattern:

Abandoned S3 Bucket Supply Chain Attack (February 2025). WatchTowr Labs researchers identified approximately 150 abandoned S3 buckets still referenced by applications and websites for software updates. These buckets had previously belonged to governments, Fortune 500 companies, technology and cybersecurity firms, and major open-source projects. The researchers spent $400+ to re-register them and received more than 8 million HTTP requests

from government agencies, military networks, Fortune 500 companies, payment card networks, banks, and universities. WatchTowr assessed that exploiting this at scale could dwarf the SolarWinds supply chain attack.

CodeBreach (disclosed January 2026, initially fixed August 2025). Wiz Research discovered a misconfigured AWS CodeBuild webhook that allowed bypass of actor ID verification checks. The root cause: missing `^` and `$` anchors in a regex filter for the `ACTOR_ID` check. Any GitHub user whose ID was a superstring of an approved ID could trigger the build. Affected repositories included the AWS JavaScript SDK (which powers the AWS Management Console), AWS Libcrypto, Amazon Corretto Crypto Provider, and the Registry of Open Data on AWS. AWS was notified on August 25, 2025, applied an initial fix on August 27, 2025 (anchoring regex filters and revoking the compromised token), deployed additional hardening in September 2025, and confirmed no evidence of exploitation in the wild.

JavaGhost: Phishing via AWS SES and WorkMail (February 2025). Palo Alto Networks Unit 42 reported threat actors exploiting exposed AWS access keys to hijack Amazon SES and WorkMail for phishing campaigns. The attackers created new SES and WorkMail users, set up SMTP credentials, and sent phishing emails that bypassed standard protections, because they came from a legitimate, trusted entity. They also created new IAM roles with trust policies granting cross-account access to attacker-controlled AWS accounts.

The Numbers Behind the Threat

The aggregate data from 2025 is stark.

IBM Cost of a Data Breach Report 2025 found the global average cost at $4.44 million, a 9% decline from $4.88 million the prior year, the first decrease in five years. One major factor: organizations that deployed security AI and automation extensively cut their breach lifecycle by 80 days and saved approximately $1.9 million on average.

The details are worse:

- US organizations faced an average cost of **$10.22 million** per breach, a record high
- Cloud-specific breaches in multi-environment setups cost **$5.05 million** and took 276 days to contain, the costliest and slowest configuration
- Breaches involving stolen credentials as the initial vector cost **$4.67 million** each
- Breaches contained within 200 days cost approximately $3.87 million versus $5.01 million for those that took longer
- **13% of organizations** reported breaches of AI models or applications; 97% of these lacked proper AI access controls

- Shadow AI (unsanctioned AI use) was a factor in 20% of breaches, adding $670,000 to average costs

The **Verizon 2025 DBIR** examined 22,052 cyber incidents and 12,195 confirmed data breaches:

- Compromised credentials were the initial access vector in **22%** of breaches
- Vulnerability exploitation reached **20%**, a 34% increase from the prior year
- Edge devices and VPNs as targets grew from 3% to **22%**, nearly an 8x increase
- **88%** of Basic Web Application attacks involved stolen credentials
- AI-generated phishing emails **doubled over two years** (from ~5% to ~10%), per email security partner data cited in the DBIR

The Shared Responsibility Model in 2026

AWS secures the infrastructure of the cloud. Customers secure their workloads in the cloud. That split has not changed since day one. What changed in 2026 is the complexity AI services introduce.

When you deploy Amazon Bedrock to run foundation models, the responsibility split differs from traditional compute. AWS secures the Bedrock service infrastructure and tenant isolation. You are responsible for:

- **What the model can access:** IAM roles, VPC endpoints, and data sources attached to agents
- **What the model generates:** prompt injection defenses, content filtering, output validation
- **How the model acts:** guardrail configuration, tool permission boundaries, agent session scoping
- **Who the model represents:** identity propagation, audit trails linking agent actions to human initiators

When an AI agent invokes an AWS API, the Shared Responsibility Model still applies, but the "customer" side now includes controlling autonomous systems that decide on their own which APIs to call.

New Threat Vectors: What Changed in 2025-2026

Prompt Injection Against AI Agents

Prompt injection attacks against production AI agents moved from theoretical research to documented incidents in 2025. EchoLeak (CVE-2025-32711), a zero-click prompt injection vulnerability in Microsoft 365 Copilot, allowed remote data exfiltration through crafted

documents (emails, Word files, PowerPoint presentations) with no user interaction. GitHub Copilot (CVE-2025-53773) and Cursor IDE (CVE-2025-59944) both had remote code execution vulnerabilities exploited through prompt injection. Researchers found fundamental design weaknesses in Devin AI, an autonomous coding agent, that allowed attackers to expose ports, leak access tokens, and connect to a command-and-control server.

AI Agent Abuse and Credential Theft

The sub-10-minute AWS intrusion documented by Sysdig is the template: credentials exposed through AI workloads (RAG data stores on S3) feeding rapid, AI-assisted privilege escalation. IBM's 2025 report found that 13% of organizations experienced breaches of AI models or applications, and 97% of those had no proper AI access controls in place.

Supply Chain Attacks on AI/ML Infrastructure

Supply chain risk and AI infrastructure are colliding. Abandoned S3 buckets previously used for model training data, MCP tool servers with no integrity verification, malicious Docker images in container registries (like the crypto mining image with 100,000+ pulls discovered in the November 2025 campaign), all vectors where AI workloads inherit untrusted dependencies.

Five Patterns Defining the 2026 Threat Picture

Five patterns define the AWS threat picture heading into 2026:

1. Credential compromise is still the number one way in. Compromised credentials drove 22% of breaches in the Verizon DBIR 2025, powered attacks from Codefinger ransomware to the Snowflake campaign to the November 2025 crypto mining operation. Long-lived access keys, secrets in public repositories, and credentials stored in AI data pipelines keep providing the easiest path in.

2. Cloud-native features are being weaponized. Codefinger abused SSE-C encryption. Attackers exploited Lambda for privilege escalation. JavaGhost turned SES and WorkMail into phishing infrastructure. None of these were AWS vulnerabilities; they were legitimate features accessed through compromised credentials.

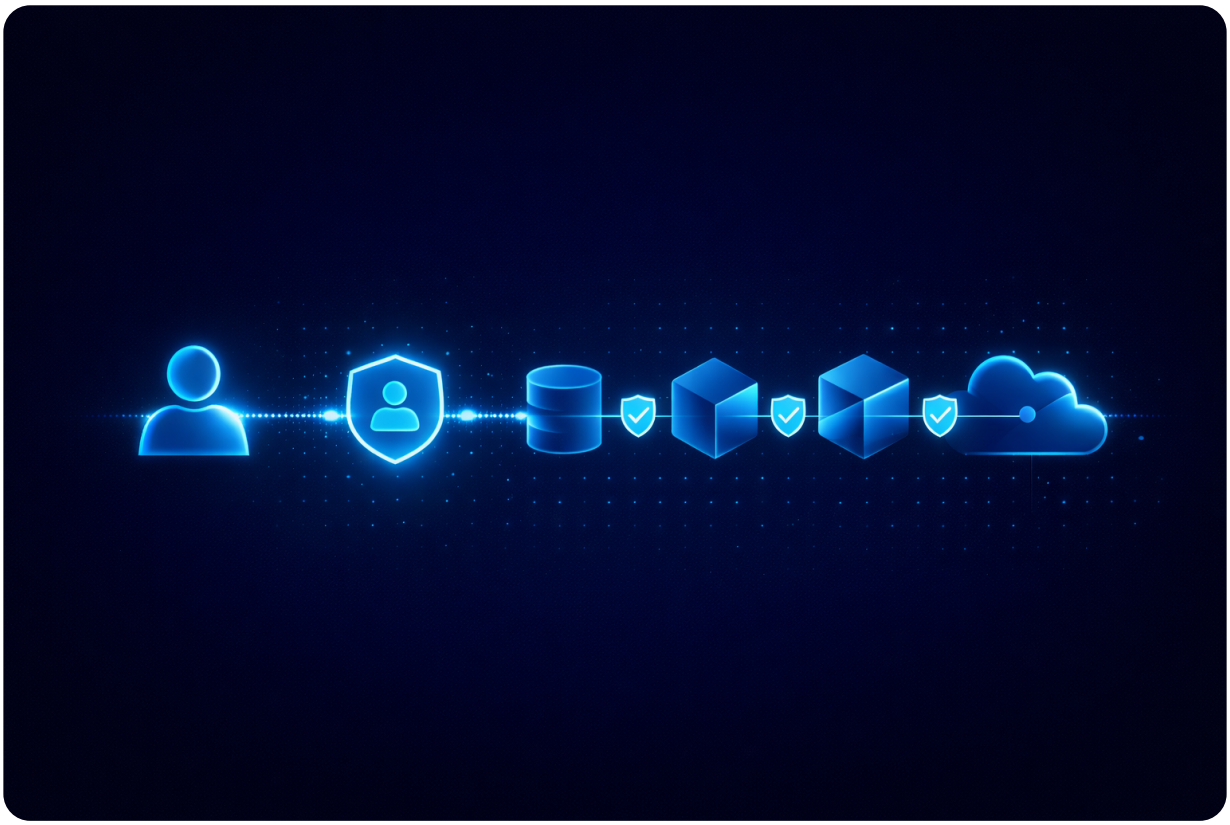
3. Supply chain risk is exploding. Third-party involvement in breaches doubled to 30%. Abandoned S3 buckets received millions of requests from government and enterprise networks. A missing regex anchor in CodeBuild nearly exposed the AWS Console's JavaScript SDK. A single supply chain compromise now reaches across organizations, industries, and borders.

4. AI accelerates both attack and defense. Attackers used AI to achieve admin access in under 10 minutes. AI-generated phishing emails doubled over two years to ~10% of phishing volume. Defenders who deployed AI extensively saved $1.9 million per breach and cut breach lifecycle by 80 days. If you are not adopting security AI, you are already behind the attackers who have.

5. Misconfiguration is still the root cause of nearly every incident. Missing regex anchors. Exposed access keys. Public S3 buckets. No MFA. Overly permissive IAM roles. Every major incident in this chapter traces back to a configuration error, not a zero-day exploit. The most sophisticated attack tooling in the world still walks through open doors.

The following chapters provide the controls, architectures, and implementation details to address each of these patterns, starting with the foundation: IAM and identity.

Chapter 2: IAM & Identity: From Basics to Trusted Identity Propagation



The Foundation of Everything

Every API call, every console action, every service-to-service interaction in AWS begins with an identity making a request and a policy deciding whether to allow it. Get IAM wrong, and no amount of encryption, network isolation, or monitoring will save you.

In 2025 and 2026, AWS shipped major IAM changes: Resource Control Policies joined Service Control Policies as organizational guardrails, SCPs gained full IAM policy language support, IAM Access Analyzer expanded to internal access findings, root user MFA became mandatory across all account types, and Trusted Identity Propagation matured into a production-ready capability for passing real user identities through AWS analytics service chains. Attackers, meanwhile, kept exploiting the same misconfigurations (overly permissive roles, long-lived access keys, and absent MFA) that have defined cloud breaches since 2019.

What follows: how policy evaluation actually works, what goes wrong in practice, and the new capabilities that should change how you approach identity in 2026.

How IAM Policy Evaluation Works

You cannot secure IAM without understanding how AWS evaluates policies. When a principal makes a request, the evaluation follows a specific order, and an explicit Deny in any policy type always wins. (AWS evaluates policies simultaneously; the sequential model below is a conceptual simplification.)

The evaluation flow for a single-account request:

1. **Default: Implicit Deny.** All requests start as denied. The AWS account root user is the exception; it has full access by default.
2. **Explicit Deny check.** The enforcement code evaluates all applicable policies: SCPs, RCPs, resource-based, identity-based, permissions boundaries, and session policies. If any policy contains an explicit Deny that matches the request, the final decision is Deny. No Allow can override this.
3. **Resource Control Policies (RCPs).** If the account is in an AWS Organization with RCPs attached, the code checks for an applicable Allow. If no Allow is found, the result is an implicit deny.
4. **Service Control Policies (SCPs).** Same logic as RCPs: if no applicable Allow is found, the result is an implicit deny.
5. **Resource-based policies.** For IAM users, an Allow in a resource-based policy grants access directly (union logic). For role session principals, the behavior depends on how the principal is specified in the resource-based policy:
6. If the resource-based policy specifies the **role session ARN** (e.g., `arn:aws:sts::123456789012:assumed-role/RoleName/SessionName`), permissions are granted directly to the session and are **not** limited by permissions boundaries or session policies.
7. If the resource-based policy specifies the **role ARN** (e.g., `arn:aws:iam::123456789012:role/RoleName`), permissions boundaries and session policies **do** apply, and the resource-based policy alone does not grant access without identity-based policy alignment.

Note: IAM role trust policies and KMS key policies have special evaluation behavior: they can independently grant access without requiring identity-based policy alignment.
6. **Identity-based policies.** The user's or role's attached policies are evaluated. If no Allow matches, the result is Deny.
7. **Permissions boundaries.** If a permissions boundary exists, the action must be allowed by both the identity-based policy and the boundary (intersection logic).
8. **Session policies.** For role sessions and federated users, the session policy must also allow the action.

The relationships that matter:
- Identity-based + Resource-based = **Union** (either can grant access for same-account IAM users)
- Identity-based + Permissions boundary = **Intersection** (both must allow)
- Identity-based + SCPs = **Intersection** (both must allow)
- Identity-based + RCPs = **Intersection** (both must allow)

Common IAM Misconfigurations in 2025

Every major breach discussed in Chapter 1 traces back to an IAM misconfiguration. These patterns keep recurring.

Long-Lived Access Keys

Static access keys are the most common credential exposure vector. They never expire. They end up in plaintext everywhere: environment variables, `.env` files, git repositories, CI/CD configurations. Anyone who gets them can use them immediately. The Snowflake campaign used credentials stolen from infostealer malware dating back to November 2020. The Codefinger ransomware attack used compromised keys with just `s3:GetObject` and `s3:PutObject` permissions to encrypt entire S3 buckets.

The fix: Use IAM Identity Center with temporary credentials for human access. Use IAM roles for workloads. When access keys are truly unavoidable, enforce rotation policies and monitor with IAM Access Analyzer's unused access findings.

Overly Permissive Default Roles

Aqua Security's "Shadow Roles" research showed that several AWS services auto-create IAM roles with overly broad permissions when users accept defaults:

Service	Default Role	Overly Broad Permission
Amazon SageMaker	`AmazonSageMaker-ExecutionRole-<timestamp>`	Custom policy equivalent to `AmazonS3FullAccess`
AWS Glue	`AWSGlueServiceRole`	`AmazonS3FullAccess`
Amazon EMR	`AmazonEMRStudio_RuntimeRole_<epoch>`	`AmazonS3FullAccess`

The research also identified shadow role vectors in Amazon Lightsail and the Ray framework on AWS. An attacker who gains access to any of these default roles can search the account for S3 buckets used by other services, modify assets like CloudFormation templates, and move laterally across services within the same account. AWS responded by tightening default role policies. The AWS CDK now restricts asset uploads to buckets in the user's account.

Mis-Scoped AWS Managed Policies

On July 10, 2025, Cymulate disclosed that the `AmazonGuardDutyFullAccess` managed policy included `organizations:RegisterDelegatedAdministrator` with `Resource: "*"` and no condition restricting which service could be delegated. Any principal in the management account holding this policy could register a delegated administrator for any Organizations-integrated service, not just GuardDuty. Chain that to IAM Identity Center or CloudFormation StackSets and an attacker goes from a single compromised identity to full control of every account in the organization.

AWS released `AmazonGuardDutyFullAccess_v2`, which restricts the action with a condition:

```
"Condition": {
  "StringEquals": {
    "organizations:ServicePrincipal": [
      "guardduty.amazonaws.com",
      "malware-protection.guardduty.amazonaws.com"
    ]
  }
}
```

Starting August 26, 2025, new attachments of the v1 policy were blocked. Existing v1 attachments remain functional and must be manually updated.

Root Users Without MFA

AWS closed this gap methodically over 2024-2025:

Date	Milestone
May 2024	MFA enforcement began for Organizations management account root users
June 11, 2024	FIDO2 passkey support added as MFA method for root and IAM users
June 2024	MFA enforcement extended to standalone account root users
November 15, 2024	Centralized root access management launched; programmatic removal of root credentials from member accounts via `AssumeRoot` for short-term, task-scoped root sessions

Date	Milestone
June 17, 2025	MFA enforced for member account root users, achieving 100% root MFA enforcement across all account types

Enforcement uses a 35-day grace period. Users must register MFA within 35 days of their first sign-in attempt. AWS supports up to 8 MFA devices per root user.

What Goes Wrong in Practice

- Teams default to `AdministratorAccess` during development and never scope it down before production. Build least-privilege from the start; retrofitting permissions on a running production workload is exponentially harder.
- "Shadow roles" created by AWS services silently accumulate. No one owns them, no one reviews them, and they persist with broad permissions long after the original workload is decommissioned. Inventory service-created roles quarterly.
- Access key rotation policies exist on paper but break in practice because applications hardcode keys. Before enforcing rotation, audit every key's consumer. If a key feeds a CI/CD pipeline or third-party integration, migrate to IAM roles first.
- Credential reports and console last-sign-in dates only tell you about IAM user activity. For role usage, you need CloudTrail and IAM Access Analyzer's unused access findings. Do not rely on a single data source.

Organizational Guardrails: SCPs and RCPs

Service Control Policies (SCPs)

SCPs cap the maximum permissions for principals in member accounts. They do not grant permissions; they restrict what identity-based policies can allow. SCPs do not affect the management account or service-linked roles.

September 2025: Full IAM Policy Language. Before September 19, 2025, SCPs had real syntax limitations: conditions only worked in Deny statements, resource ARNs only accepted wildcards, and `NotAction` with Allow was unavailable. The September update removed all of these restrictions:

Capability	Before September 19, 2025	After September 19, 2025
Conditions in Allow statements	Not supported	Supported

Capability	Before September 19, 2025	After September 19, 2025
Individual resource ARNs in Allow	Wildcards only	Individual ARNs supported
`NotAction` with Allow	Not supported	Supported
`NotResource` in Allow and Deny	Not supported	Supported
Wildcards anywhere in Action strings	End only (`s3:Get*`)	Anywhere (`*:PutBucket*`)

In practice, SCPs can now enforce region restrictions via conditions in Allow statements, allowing actions only when `aws:RequestedRegion` matches approved regions. Previously, this required Deny-based approaches that were harder to maintain.

Resource Control Policies (RCPs): New in November 2024

RCPs complement SCPs by restricting access to resources rather than what principals can do. SCPs answer "what can this principal do?" RCPs answer "who can access this resource?"

RCPs are built for data perimeters: they prevent S3 buckets, KMS keys, SQS queues, and other resources from being accessed by external principals (any principal not belonging to your AWS Organization, including third-party accounts, federated users from outside the organization, and anonymous access), regardless of what permissions those principals hold.

Supported services (as of February 2026):

Service	Added
Amazon S3	November 2024 (launch)
AWS STS	November 2024 (launch)
AWS KMS	November 2024 (launch)
Amazon SQS	November 2024 (launch)
AWS Secrets Manager	November 2024 (launch)
Amazon ECR	June 2025
Amazon OpenSearch Serverless	June 2025
Amazon Cognito	January 2026

Service	Added
Amazon CloudWatch Logs	January 2026
Amazon DynamoDB	February 2026

Key differences from SCPs:
- SCPs restrict principals; RCPs restrict resources
- SCPs support both Allow and Deny effects; RCPs support only Deny (they define a permissions guardrail via `RCPFullAWSAccess` default policy)
- Neither SCPs nor RCPs affect the management account. SCPs do not affect service-linked roles. RCPs do not affect service-linked roles
- AWS-managed KMS keys cannot be restricted by RCPs

RCPs expanded to AWS GovCloud (US) regions on May 2, 2025.

Multi-Account IAM Architecture with SCPs, RCPs, and IAM Identity Center

What Goes Wrong in Practice

- Teams apply restrictive SCPs to the root OU without testing in a sandbox OU first, then spend the next four hours on an incident call figuring out why deployments, SSO, and automation broke simultaneously. Always test SCPs against a dedicated sandbox OU with representative workloads before promoting to production OUs.
- SCPs do not affect the management account. This is the single most misunderstood fact about SCPs. If your security strategy depends on SCPs restricting actions in the management account, it does not work. Minimize what runs in the management account.
- RCPs are Deny-only. Teams try to write Allow-effect RCPs and wonder why validation fails. RCPs define what access to deny on resources; the `RCPFullAWSAccess` default policy provides the implicit Allow baseline.
- Forgetting that neither SCPs nor RCPs affect service-linked roles leads to false confidence. A service-linked role can still perform its actions even if an SCP or RCP would block an equivalent user-created role.
- When building data perimeters with RCPs, start with S3 and KMS. These two services cover the vast majority of data exfiltration vectors. Add STS next to prevent external role assumption. Expand from there.

IAM Access Analyzer: Finding What's Wrong Before Attackers Do

IAM Access Analyzer uses automated reasoning (mathematical proofs, not pattern matching) to analyze policies and identify access risks. It operates across three dimensions:

External Access Findings

Finds resources shared with external principals: an IAM role with a trust policy open to another account, an S3 bucket policy granting public access, a KMS key policy allowing cross-account usage. This is the original Access Analyzer capability, covering the broadest set of resource types: IAM roles, S3 buckets, KMS keys, Lambda functions, SQS queues, Secrets Manager secrets, SNS topics, EBS volume snapshots, ECR repositories, EFS file systems, and more. Available at no additional charge.

Internal Access Findings: New in June 2025

Announced June 17, 2025, internal access findings reveal which IAM roles and users within your own organization can reach your resources. External access analysis answers "who outside can get in?" Internal access findings answer: "Which principals inside my organization can reach this S3 bucket, this DynamoDB table, this RDS snapshot?"

Supported resource types for internal access analysis: Amazon S3 buckets, S3 directory buckets, Amazon RDS DB snapshots, Amazon RDS DB cluster snapshots, Amazon DynamoDB streams, and Amazon DynamoDB tables.

The analyzer evaluates SCPs, RCPs, identity-based policies, resource-based policies, permissions boundaries, and declarative policies simultaneously using automated reasoning to determine effective access. Findings update daily and are aggregated in a unified dashboard with EventBridge integration for automated notifications.

Unused Access Findings

Flags IAM roles not used within a specified period, untouched access keys, service actions granted but never exercised, and unused passwords. This is the operational backbone of least-privilege enforcement. It tells you exactly what permissions to remove.

Custom Policy Checks

Validates policies in your CI/CD pipeline using automated reasoning. Before deploying a new policy, check whether it grants access to resources or actions you want to prohibit. Policy generation creates fine-grained policies from CloudTrail activity across AWS services using up to a 90-day activity window.

What Goes Wrong in Practice

- Unused access findings generate thousands of results in a mature AWS environment. Start with external access findings (publicly exposed resources), which are always urgent. For unused access, filter by last-accessed date greater than 90 days and prioritize roles attached to production workloads first. Do not try to remediate all findings at once. You will burn out your team.
- Teams enable Access Analyzer, get overwhelmed by the volume of findings, and then ignore it entirely. Treat it like a backlog: triage weekly, fix the highest-severity findings first, and track remediation rate over time.
- Policy generation from CloudTrail activity is powerful but only captures actions that were actually invoked during the observation window. If your application has seasonal or monthly batch jobs, a 90-day window may still miss them. Cross-reference generated policies against application documentation before replacing existing permissions.
- Custom policy checks in CI/CD pipelines catch problems before deployment, but only if you define the right checks. At minimum, block policies that grant `*:*`, `iam:PassRole` with wildcard resources, or `sts:AssumeRole` without conditions.
- To quickly list unused access findings sorted by severity:

```bash
aws accessanalyzer list-findings-v2 \
  --analyzer-arn <arn> \
  --filter '{ "findingType": {"eq": ["UnusedAccess"]}, "status": {"eq": ["ACTIVE"]} }' \
  --sort-criteria '{ "attributeName": "severity", "orderBy": "DESC" }'
```

Trusted Identity Propagation: The End of the Service Account Anti-Pattern

Trusted Identity Propagation (TIP) is the most consequential IAM capability AWS has shipped in years. Announced at re:Invent 2023 and matured through 2024-2025, it solves a persistent problem: when a user accesses data through a chain of AWS analytics services, their individual identity is lost.

The Problem

A data analyst opens Amazon QuickSight to query Amazon Redshift. Without TIP, the connection from QuickSight to Redshift uses a shared IAM role. Every analyst gets the same access. CloudTrail logs show the IAM role accessed Redshift, but not which human made the query. You cannot enforce per-user access controls, and you cannot attribute actions to individuals.

This is the service account anti-pattern: shared roles masking individual identity, granting uniform access regardless of the user, and making per-user auditing impossible.

How TIP Works

TIP uses an OAuth 2.0 token exchange to carry the user's corporate identity through the entire service chain:

Step 1: External Authentication. The user authenticates with the corporate identity provider (Okta, Microsoft Entra ID, or another OAuth 2.0 authorization server) and receives a signed JWT token.

Step 2: Token Exchange with IAM Identity Center. The application exchanges that JWT with IAM Identity Center using the `sso-oidc:CreateTokenWithIAM` API call with the `urn:ietf:params:oauth:grant-type:jwt-bearer` grant type. Identity Center validates the JWT, maps the external user to an Identity Center user, and returns an `idToken` containing the `sts:identity_context` claim.

Step 3: Identity-Enhanced IAM Role Session. The application calls `sts:AssumeRole` with the `ProvidedContexts` parameter, passing the `sts:identity_context` value and the context provider ARN `arn:aws:iam::aws:contextProvider/IdentityCenter`. The result is an identity-enhanced IAM role session that carries both the IAM role identity and the corporate user identity.

Step 4: Downstream Authorization. When the identity-enhanced session calls a downstream service (Redshift, Athena, S3 Access Grants, Lake Formation), that service checks the user's identity against its authorization rules. CloudTrail logs the user's IAM Identity Center `userId` in the `onBehalfOf` element:

```
"onBehalfOf": {
    "userId": "1111-1111-1111-1111-1111",
    "identityStoreArn":
      "arn:aws:identitystore::111111111111:identitystore/d-1111"
}
```

Supported Services (as of February 2026)

TIP supports these AWS analytics and data services:

Service	Notes
Amazon Redshift	Direct query and Data API
Amazon Athena	Direct query
Amazon EMR / EMR Serverless	Interactive workloads (via Apache Livy)
AWS Lake Formation	Fine-grained data catalog permissions
Amazon S3 (via S3 Access Grants)	Per-user object-level access

Service	Notes
AWS Glue	Interactive sessions (Glue 5.0+)
Amazon OpenSearch Service	Search and analytics
Amazon QuickSight	Propagates identity to Redshift and Athena
Amazon Q Business	AI-powered business intelligence
AWS Transfer Family	Web application file transfers
Amazon SageMaker Studio	GA August 13, 2025
Amazon SageMaker Unified Studio	SQL analytics from September 30, 2025

TIP with S3 Access Grants

S3 Access Grants paired with TIP eliminates per-user bucket policies entirely. Administrators define grants that bind IAM Identity Center users or groups to specific S3 locations with specific permissions (Read, Write, or both). When a user requests access, S3 Access Grants evaluates the grants, assumes a backing IAM role, and vends temporary credentials scoped to that user's specific S3 prefix.

Authorization lives centrally in S3 Access Grants. Each user gets credentials scoped down to their specific grants.

New IAM Capabilities from re:Invent 2025

IAM Policy Autopilot

An open-source MCP server that analyzes application code to generate baseline least-privilege IAM policies. It supports Python, Go, and TypeScript, and integrates with AI coding assistants including Kiro, Amazon Q Developer, Cursor, Cline, and Claude Code.

Instead of starting with `AdministratorAccess` and hoping to scope it down later, developers generate a least-privilege policy from their code before the first deployment. The tool performs static analysis: it reads the application code locally, identifies which AWS API calls the code makes, and generates an IAM policy granting only those permissions.

Outbound Identity Federation

AWS workloads now authenticate with external services (third-party cloud providers, SaaS applications, self-hosted services) using short-lived JWTs from AWS STS, eliminating long-term credentials. Workloads call the `sts:GetWebIdentityToken` API to request cryptographically signed, publicly verifiable tokens.

Tokens support RS256 and ES384 signing algorithms, with durations from 60 to 3,600 seconds (default 300 seconds). AWS generates a unique issuer URL per account with OIDC discovery endpoints for token verification. OIF requires regional STS endpoints, not the global endpoint. Available in all commercial and GovCloud regions at no additional cost. AWS documentation also lists China regions as supported.

`aws login` CLI Command

A new CLI command that opens the browser for authentication and generates temporary credentials cached locally. Credentials auto-refresh every 15 minutes and sessions remain valid for up to 12 hours. Requires AWS CLI v2.32.0 or later. Developers authenticate once; the CLI handles credential rotation from there. No more manually configuring access keys.

Network Perimeter Condition Keys

Three new VPC endpoint condition keys (August 2025) -- `aws:VpceAccount`, `aws:VpceOrgPaths`, and `aws:VpceOrgID` -- eliminate the need to enumerate individual VPC endpoint IDs in policies. As endpoints change, these keys scale automatically. They work in SCPs, RCPs, session policies, permissions boundaries, identity-based policies, and resource-based policies.

The `aws:SourceVpcArn` condition key (November 2025) returns the ARN of the VPC where the endpoint is attached. Because VPC ARNs include the region, this opens up region-based access controls, essential for data residency requirements.

IAM Identity Center: The Human Access Layer

IAM Identity Center is AWS's answer to human access management. It replaces IAM users with long-lived credentials.

AWS is explicit: "As a best practice, AWS recommends that you require human users to assume an IAM role to access AWS so that they're using temporary credentials."

Multi-Region Replication: February 2026

On February 3, 2026, AWS launched IAM Identity Center multi-region replication at general availability. It replicates workforce identities, permission sets, assignments, and sessions from the primary region to additional regions, providing active AWS access portal endpoints in each one. If the primary region goes down, users access AWS accounts through the access portal in the additional regions.

Available in 17 enabled-by-default commercial regions. Requirements include an organization instance, an external identity provider (Okta, Microsoft Entra ID, PingFederate, PingOne, or JumpCloud), and a multi-region customer-managed KMS key. Available at no additional cost beyond standard KMS charges.

CloudTrail Event Security Changes: 2025

AWS changed how CloudTrail events work for IAM Identity Center. Changes began taking effect July 14, 2025, with full deployment completing later in 2025:

- `userName` **and** `principalId` **removed** from the `userIdentity` element, replaced with `userId` (unique, immutable) and `identityStoreArn`
- **New identity type** `IdentityCenterUser` replaces the generic `Unknown` type for authenticated users
- `credentialId` **field added** for correlating API calls to specific authentication credentials
- **Group** `displayName` **masked** as `HIDDEN_DUE_TO_SECURITY_REASONS` in `CreateGroup` and `UpdateGroup` events, preventing accidental recording of sensitive information in CloudTrail logs

The net effect: immutable identifiers resistant to username changes and no sensitive data leaking into audit logs.

What Goes Wrong in Practice

- Running IAM Identity Center directly from the management account is a common mistake. Set up a delegated administrator account for Identity Center administration. The management account should have minimal direct usage.
- Permission sets start narrow and grow over time as teams request "just one more action." Review permission sets quarterly. Compare them against IAM Access Analyzer unused access findings to identify bloated grants.
- Teams configure SAML/OIDC federation with their identity provider and assume the job is done. Test failover scenarios: what happens when the IdP is unreachable? Multi-region replication (February 2026) addresses the AWS-side resilience gap, but you still need IdP-side redundancy.
- Attribute-based access control (ABAC) through Identity Center is underused. Instead of creating dozens of permission sets for every team-and-environment combination, pass user attributes (department, team, cost center) as session tags and write policies that reference those tags. This scales far better than per-team permission sets.

The IAM Audit: A Practitioner's Approach

A systematic IAM audit covers five areas:

1. Credential Inventory. List all IAM users, access keys, and their last-used dates. Flag access keys older than 90 days, keys never used, and users without MFA. Use IAM Access Analyzer unused access findings to spot dormant permissions.

2. Policy Review. Hunt for policies with `"Action": "*"` or `"Resource": "*"`. Check for `iam:PassRole` with wildcard resources. Review all trust policies on IAM roles for overly broad principals. Confirm that AWS managed policies in use are current versions (not deprecated v1 policies like `AmazonGuardDutyFullAccess`).

3. Organizational Guardrails. Confirm SCPs are attached at appropriate OU levels. Confirm RCPs enforce data perimeters on S3, KMS, STS, SQS, and Secrets Manager. Verify the management account has minimal direct access and uses delegated administration.

4. Root User Security. Check that MFA is on for all root users. Use centralized root access management to remove root credentials from member accounts. Verify no active root access keys exist.

5. Identity Center Configuration. Confirm all human users access AWS through Identity Center, not IAM users. Review permission sets for least privilege. Test external identity provider integration end-to-end. Turn on multi-region replication for resilience.

IAM is the foundation, but identity without data protection is incomplete. Next: securing the data itself, starting with S3.

Chapter 3: Securing Data: S3, Encryption & Data Classification

The Data Protection Imperative

Every breach in Chapter 1 targeted data. Codefinger encrypted it for ransom. The Snowflake campaign (Chapter 1) exfiltrated hundreds of millions of records. The AI-assisted intrusion documented by Sysdig created backdoors to reach it. In AWS, most customer data at rest lives in Amazon S3, and S3 remains one of the most targeted services in cloud attacks.

This chapter covers the controls that protect data at rest and in transit: encryption, access control, classification, monitoring, and resilience.

Encryption at Rest: The Four Options

Since January 5, 2023, Amazon S3 encrypts every new object with SSE-S3 (AES-256) at no additional cost and with no performance impact. You cannot disable encryption. SSE-S3 is the mandatory baseline for every bucket.

SSE-S3 is only the starting point. AWS offers four server-side encryption options, each with a different key management model:

SSE-S3 (Server-Side Encryption with Amazon S3 Managed Keys)

S3 encrypts each object with a unique data key. That data key is itself encrypted by a key that AWS manages and regularly rotates. You never see or control the keys.

Use when: You need encryption at rest with zero operational overhead. Suitable for general workloads without regulatory key management requirements.

SSE-KMS (Server-Side Encryption with AWS KMS Keys)

Ties AWS Key Management Service into S3. You choose between an AWS managed key (created automatically per service) or a customer managed key (created, owned, and controlled by you). SSE-KMS gives you what SSE-S3 does not: visible key policies you control, CloudTrail logging of key usage, configurable rotation, and cross-account encrypted data sharing (customer managed keys only).

Use when: You need audit trails of key usage, granular access control via key policies, or cross-account encrypted data sharing.

DSSE-KMS (Dual-Layer Server-Side Encryption with AWS KMS Keys)

Applies two independent layers of AES-256 encryption: the first using an AWS KMS data encryption key, the second using a separate S3-managed key.

Use when: Regulatory requirements mandate two layers of encryption. S3 Bucket Keys are not supported with DSSE-KMS, resulting in higher per-request KMS costs.

SSE-C (Server-Side Encryption with Customer-Provided Keys)

You provide the encryption key with every request. S3 performs the encryption and decryption but never stores the key. If you lose the key, your data is permanently unrecoverable. AWS cannot help.

Use when: You should not. SSE-C is disabled by default starting April 6, 2026 (see below). Migrate to SSE-KMS with customer managed keys.

The SSE-C Deprecation: Lessons from Codefinger

The Codefinger ransomware campaign (Chapter 1) demonstrated why SSE-C is dangerous: S3 never retains the customer-provided key, and CloudTrail logs only an HMAC of the key used, not the key itself. Recovery without the attacker's key is technically impossible.

AWS responded on multiple fronts. In November 2025, S3 added a bucket-level `BlockedEncryptionTypes` setting via the `PutBucketEncryption` API. Then came the decisive change: **starting April 6, 2026, SSE-C is disabled by default for all new S3 general purpose buckets and for all existing buckets in accounts that have no SSE-C encrypted data.**

Scope is per-account: if any bucket in your account contains SSE-C encrypted objects, all existing bucket configurations remain unchanged. Applications that still require SSE-C must explicitly re-enable it via `PutBucketEncryption`.

When SSE-C is blocked on a bucket, S3 rejects the following operations with HTTP 403 if they include SSE-C headers: `PutObject`, `CopyObject`, `PostObject`, multipart upload requests, and replication requests. Existing SSE-C encrypted objects remain readable when you provide the correct key on `GetObject` or `HeadObject`.

Immediate action: Audit your S3 buckets for SSE-C usage. If you do not need SSE-C, block it now using the `BlockedEncryptionTypes` setting. If you do use SSE-C, plan your migration to SSE-KMS with customer managed keys before April 2026.

KMS: Key Management That Matters

Customer Managed Keys vs. AWS Managed Keys

The choice between these two key types affects audit, rotation, and cross-account access directly:

Capability	Customer Managed Keys	AWS Managed Keys
Key policy control	Full customer control	View only (cannot modify)
CloudTrail audit	Full audit trail	Full audit trail
Cross-account sharing	Yes	No
Custom rotation period	Yes (90–2,560 days)	Fixed (~365 days)
On-demand rotation	Yes	No
Tags and aliases	Yes	No

Capability	Customer Managed Keys	AWS Managed Keys
Schedule for deletion	Yes	No
Direct API usage	Yes	No (service use only)

AWS managed keys are a legacy key type. Since 2021, new AWS services default to AWS owned keys instead. For any workload requiring key control, use customer managed keys.

Key Rotation

In April 2024, AWS expanded KMS automatic key rotation with three changes:

1. **Configurable rotation period:** 90 to 2,560 days (previously fixed at 365 days)
2. **On-demand rotation:** rotate immediately without waiting for the schedule (maximum 25 on-demand rotations per key)
3. **Rotation history:** view all previous rotations via the `ListKeyRotations` API

In June 2025, on-demand rotation was extended to symmetric encryption keys with imported key material.

Key rotation changes only the current key material; it does not re-encrypt previously protected data or rotate data keys generated by the KMS key. Decryption is transparent: AWS KMS selects the correct key material version automatically.

S3 Bucket Keys: Reducing KMS Costs

Without S3 Bucket Keys, S3 makes a call to AWS KMS for every request against a KMS-encrypted object. With Bucket Keys enabled, KMS generates a short-lived bucket-level key that S3 uses to create data keys locally, reducing KMS API calls by up to 99%.

The tradeoff: CloudTrail events log the bucket ARN instead of the object ARN in the encryption context. If your IAM or KMS key policies reference object ARN as encryption context, update them to use bucket ARN. S3 Bucket Keys are not supported with DSSE-KMS.

What Goes Wrong in Practice

Three KMS mistakes cause the most operational pain. First, teams trigger on-demand rotations without tracking how close they are to the 25-rotation limit per key. Once you hit 25, you cannot rotate that key again -- you must create a new key, re-encrypt data, and update every policy and grant that references the old key. Second, teams default to AWS managed keys because they are simpler, then discover months later they cannot share encrypted snapshots or S3 objects cross-account. AWS managed keys do not support cross-account access. Migrating to customer managed keys after the fact means re-encrypting everything. Third, teams disable or schedule deletion of a KMS key without realizing the blast radius: every object, snapshot, volume, and database encrypted with that key becomes immediately unreadable when the key is disabled, and permanently unrecoverable after the deletion waiting period expires. Before disabling any KMS key, audit its usage across all services and accounts first.

Post-Quantum Cryptography: Preparing for Tomorrow

The "harvest now, decrypt later" threat -- adversaries collecting encrypted data today to decrypt once quantum computers become viable -- makes post-quantum cryptography relevant now.

AWS is deploying post-quantum protection across its services:

ML-KEM (FIPS 203) for key exchange. AWS turned on hybrid TLS key establishment combining classical ECDH with ML-KEM (Module-Lattice-based Key-Encapsulation Mechanism) across AWS KMS, ACM, and Secrets Manager. AWS-LC (AWS's open-source cryptographic library) was the first open-source cryptographic library to include ML-KEM in its FIPS 140-3 validation. CRYSTALS-Kyber, the predecessor algorithm, is being phased out in favor of ML-KEM.

ML-DSA (FIPS 204) for digital signatures. Launched in June 2025, AWS KMS now supports post-quantum digital signatures with three key specs (ML_DSA_44, ML_DSA_65, and ML_DSA_87), all operating within FIPS 140-3 Security Level 3 validated HSMs. Combined with AWS Private CA, organizations can create quantum-resistant roots of trust and code signing certificates.

S3 post-quantum TLS. Amazon S3 endpoints support post-quantum TLS key exchange as of November 2025. AWS plans to deploy ML-KEM-based hybrid post-quantum key agreement to all AWS services with HTTPS endpoints over the coming years.

Services currently supporting post-quantum TLS: AWS KMS, ACM, Secrets Manager, Amazon S3, Amazon CloudFront, AWS Payments Cryptography, and AWS Transfer Family (ML-KEM for SFTP, launched May 2025).

Encryption in Transit: TLS Enforcement

Since February 27, 2024, TLS 1.2 is the enforced minimum across all AWS API endpoints -- TLS 1.0 and 1.1 are no longer supported. Defense-in-depth still requires explicit enforcement at the bucket level.

Denying HTTP (Non-TLS) Requests

Use `aws:SecureTransport` in a bucket policy Deny statement. When AWS services make calls to S3 on your behalf, network-specific context including `aws:SecureTransport` is redacted, so exclude service principals to avoid blocking legitimate AWS operations:

```
{
  "Version": "2012-10-17",
  "Statement": [{
    "Effect": "Deny",
    "Principal": "*",
    "Action": "s3:*",
    "Resource": [
      "arn:aws:s3:::DOC-EXAMPLE-BUCKET",
      "arn:aws:s3:::DOC-EXAMPLE-BUCKET/*"
    ],
    "Condition": {
      "Bool": {
        "aws:SecureTransport": "false",
        "aws:PrincipalIsAWSService": "false"
      }
    }
  }]
}
```

Enforcing Minimum TLS Version

The `s3:TlsVersion` condition key restricts access based on the TLS version negotiated. AWS already enforces TLS 1.2 at the endpoint level, but the condition key remains useful as a defense-in-depth measure and for enforcing TLS 1.3:

```
{
  "Condition": {
    "NumericLessThan": {
      "s3:TlsVersion": 1.2
    },
    "Bool": {
      "aws:PrincipalIsAWSService": "false"
    }
  }
}
```

S3 Access Control: Three Layers of Defense

Layer 1: S3 Block Public Access

S3 Block Public Access has four settings that override any bucket policy or ACL granting public access:

1. **BlockPublicAcls** rejects PUT requests that include public ACLs.
2. **IgnorePublicAcls** ignores all public ACLs on a bucket and its objects.
3. **BlockPublicPolicy** rejects `PutBucketPolicy` calls if the policy allows public access.
4. **RestrictPublicBuckets** restricts public bucket access to AWS service principals and authorized account users only.

These settings operate at three enforcement levels. The most restrictive combination wins:

Bucket-level. Applied to individual buckets (enabled by default since April 2023).

Account-level. Applied to all buckets in an AWS account.

Organization-level (November 2025). Set via Amazon S3 policies in AWS Organizations at the root or OU level. Propagates automatically to all sub-accounts and new members. At this level, Block Public Access applies all four BPA settings together (all-or-nothing), but the policy itself can be scoped to specific OUs or accounts. When applied, it overrides account-level and bucket-level configurations.

Layer 2: ACLs Disabled by Default

Since April 2023, all new S3 buckets use the **Bucket owner enforced** Object Ownership setting, disabling ACLs entirely. The bucket owner owns and controls every object, regardless of who uploads it. PUT requests that specify an ACL (other than `bucket-owner-full-control`) fail with HTTP 400.

For existing buckets still using ACLs, AWS recommends migrating:

1. Review existing bucket and object ACLs to identify permissions granted via ACLs
2. Use S3 server access logs or CloudTrail to identify ACL-dependent requests
3. Replace ACL-based permissions with equivalent bucket policies or IAM policies
4. Set Object Ownership to Bucket owner enforced

Layer 3: Data Perimeters with RCPs

Resource Control Policies (Chapter 2) are the strongest access control for S3 at the organizational level. AWS defines three data perimeter dimensions:

Only trusted identities. Use `aws:PrincipalOrgID` in an RCP to ensure S3 buckets across your organization can only be accessed by principals within your organization, regardless of individual bucket policies. Exempt AWS service principals with `aws:PrincipalIsAWSService`.

Only trusted resources. Use `aws:ResourceOrgID` in SCPs to prevent your identities from accessing resources outside your organization.

Only expected networks. Use `aws:SourceVpc` or `aws:SourceVpce` in RCPs to restrict S3 access to requests originating from your VPCs or specific VPC endpoints.

All three dimensions are documented in the AWS whitepaper "Building a Data Perimeter on AWS," with template policies available in the `aws-samples/data-perimeter-policy-examples` GitHub repository.

S3 Access Grants: User-Level Object Access

S3 Access Grants (GA November 2023) map permissions directly to S3 data locations -- prefix, bucket, or object -- for IAM principals or corporate directory users and groups. Each grant assigns one of three permission levels: read-only, write-only, or read-write.

Paired with IAM Identity Center and Trusted Identity Propagation (Chapter 2), Access Grants let applications request S3 data on behalf of an authenticated corporate directory user without mapping to an IAM principal first. The end-user identity propagates all the way to S3, and CloudTrail data events reference the actual end user -- not generic IAM sessions.

This solves two scaling problems at once:

1. **Policy size limits.** S3 bucket policies cap out at 20 KB, and IAM policies have per-entity size limits (6,144 characters for managed policies, 10,240 characters aggregate for role inline policies). With hundreds of users needing different object-level access, you hit the ceiling fast. Access Grants have no such limit.
2. **The service account anti-pattern.** Without identity propagation, organizations fall back to shared IAM roles that multiple users assume, making it impossible to attribute data access to individual users. Access Grants with TIP preserve the actual user identity through the entire chain.

Ransomware Protection: Versioning and Object Lock

S3 Versioning

Versioning preserves every version of every object, enabling recovery from unintended user actions and application failures. When an object is "deleted," S3 places a delete marker on top of the current version -- previous versions remain.

MFA Delete adds a second layer: multi-factor authentication is required to change the versioning state of a bucket or permanently delete an object version. Only the root user (bucket owner) can enable MFA Delete, and only through the CLI or API -- not the console.

Cost consideration: Every version is charged at standard S3 rates. Without lifecycle management, costs accumulate rapidly. Use lifecycle rules to transition noncurrent versions to cheaper storage classes (S3 Glacier Flexible Retrieval, S3 Glacier Deep Archive) and expire them after a defined retention period.

S3 Object Lock

Object Lock enforces Write-Once-Read-Many (WORM) storage, preventing deletion or overwriting for a defined period or indefinitely. It requires S3 Versioning (turned on automatically when Object Lock is activated).

Compliance mode. No one, including the root user, can delete or overwrite a protected object before the retention period expires. Retention cannot be shortened and the mode cannot be changed. Use for regulatory compliance (SEC Rule 17a-4(f), CFTC Regulation 1.31, FINRA Rule 4511, validated by Cohasset Associates).

Governance mode. Users with the `s3:BypassGovernanceRetention` permission can override protections. The `x-amz-bypass-governance-retention:true` header is required. Use for flexible protection with administrative override capability.

Legal hold. Same protection as retention periods but with no expiration date. Remains in effect until explicitly removed. Independent from retention periods; both can coexist on the same object. Use when the protection duration is uncertain (ongoing audits, investigations).

Against ransomware, Object Lock in compliance mode is the strongest control: even if an attacker compromises credentials, they cannot delete or overwrite protected object versions before the retention period expires. Pair it with cross-region replication and Object Lock in the destination bucket for multi-region resilience.

What Goes Wrong in Practice

The biggest operational mistake with S3 Object Lock is enabling compliance mode without carefully planning retention periods. Teams lock data they cannot delete, storage costs spiral, and there is no override -- not even root can remove compliance-mode-locked objects before retention expires. Always start with governance mode, test retention periods against your data lifecycle, then migrate critical compliance data to compliance mode.

AWS Backup for S3

AWS Backup adds cross-region and cross-account backup to S3. It requires S3 Versioning and backs up all objects (including all versions), delete markers, tags, ACLs, and user-defined metadata.

Two backup types:

- **Continuous backups:** point-in-time restore within a maximum retention of 35 days
- **Periodic snapshots:** configurable frequency (1 hour to 1 month) with retention up to 99 years

Both are incremental at the object level after the first full backup. Cross-account and cross-region copies work, though copies of continuous backups lose point-in-time restore capabilities.

Supported storage classes: S3 Standard, S3 Standard-IA, S3 One Zone-IA, S3 Glacier Instant Retrieval, and S3 Intelligent-Tiering. **Not supported:** S3 Glacier Flexible Retrieval, S3 Glacier Deep Archive, and SSE-C encrypted objects.

AWS Backup Vault Lock adds WORM protection to the backup vaults themselves. In compliance mode, after a mandatory minimum 72-hour cooling-off period, the vault and its contents become immutable -- no user, including root, can delete or alter them until the retention period expires.

Data Classification: Amazon Macie

Amazon Macie scans S3 for sensitive data using machine learning and pattern matching. It detects credentials (private keys, AWS secret access keys), financial information (credit card numbers, bank account numbers), and personal information -- both PII (driver's licenses, passport numbers) and PHI (health insurance numbers, medical IDs).

Two Discovery Approaches

Automated sensitive data discovery. Macie continually samples your S3 bucket inventory, selecting representative objects from each bucket. It produces an interactive heat map showing sensitive data concentration across your buckets. Includes a 30-day free trial.

Sensitive data discovery jobs. On-demand, targeted analysis on specific buckets. Configurable as one-time, daily, weekly, or monthly. Use for deeper analysis when automated discovery identifies an area of concern.

Both approaches support custom data identifiers using regular expressions (up to 512 characters, PCRE syntax subset) with optional keyword sequences and proximity rules.

Integration and Automation

Macie publishes findings to Amazon EventBridge for near real-time response. Common automations: Lambda functions that apply server-side encryption to unencrypted objects, tighten bucket policies, or route findings to a SIEM.

Macie feeds into AWS Security Hub, publishing policy findings automatically (with optional sensitive data finding publication). In multi-account environments, a delegated administrator account manages Macie across all member accounts and reviews S3 buckets and findings from one place.

S3 Security Monitoring

CloudTrail Data Events

CloudTrail logs S3 object-level API operations -- `GetObject`, `PutObject`, `DeleteObject`, and more -- as data events. Trails do not log data events by default; you must enable them, and they incur additional charges.

Data events are essential for detecting unauthorized access patterns, tracking who touched what data, and investigating incidents. In the Codefinger attack, `requestParameters.x-amz-server-side-encryption-customer-algorithm` in data events would reveal SSE-C usage.

GuardDuty S3 Protection

GuardDuty S3 Protection monitors CloudTrail S3 data events for data exfiltration, data destruction, and unauthorized access. It activates by default when you first enable GuardDuty on an account.

Key finding types include:

- **Exfiltration:S3/AnomalousBehavior** (High): unusual data transfer patterns

- **Impact:S3/AnomalousBehavior.Delete** (High): anomalous deletion activity
- **Impact:S3/AnomalousBehavior.Permission** (High): anomalous permission changes
- **Policy:S3/BucketAnonymousAccessGranted** (High): bucket made publicly accessible
- **Policy:S3/BucketPublicAccessGranted** (High): bucket policy grants public access

GuardDuty Malware Protection for S3 (launched June 2024) automatically scans newly uploaded objects for malware, tagging results as `NO_THREATS_FOUND`, `THREATS_FOUND`, `UNSUPPORTED`, `ACCESS_DENIED`, or `FAILED`. An 85% price reduction in February 2025 ($0.60 to $0.09 per GB in us-east-1) makes this practical for high-volume buckets.

The Monitoring Stack

For full S3 security monitoring, deploy all four layers:

1. **CloudTrail management events** (default): bucket creation, policy changes, encryption configuration
2. **CloudTrail data events** (must enable): object-level access, encryption method used
3. **GuardDuty S3 Protection** (default with GuardDuty): anomaly detection, exfiltration alerts
4. **Macie** (must enable): sensitive data discovery, bucket security assessment

> **What Goes Wrong in Practice**
>
> Teams skip CloudTrail data events because of cost, then discover during an incident that they have no record of who accessed which objects. Management events tell you a bucket policy changed; data events tell you someone downloaded 50,000 objects at 3 AM from an IP address in a country where you have no employees. If you cannot afford data events on every bucket, enable them on buckets containing sensitive data first, and use S3 server access logs (free) as a lower-fidelity fallback on everything else.

Recent S3 Security Enhancements (2025–2026)

S3 Block Public Access -- organization-level enforcement (November 2025). Set through Amazon S3 policies in AWS Organizations, propagating automatically to all sub-accounts.

SSE-C disabled by default (April 2026). All new buckets and existing buckets in accounts without SSE-C data will have SSE-C blocked.

S3 post-quantum TLS (November 2025). S3 endpoints now support post-quantum TLS key exchange, protecting against harvest-now-decrypt-later attacks.

S3 Conditional Writes (August 2024, expanded November 2024). The `if-none-match` and `if-match` headers prevent accidental overwrites and detect concurrent modifications. Bucket policies can enforce conditional writes using `s3:if-none-match` and `s3:if-match` condition keys.

S3 Metadata (GA January 2025). Automatically captures object metadata -- encryption status, requester information, source IP -- in a queryable Apache Iceberg table. Updates land within minutes, making security queries at scale practical.

S3 Tables (December 2024). Managed Apache Iceberg tables with AWS KMS encryption using customer managed keys, IAM resource policies for table-level access control, and integration with AWS Lake Formation through Glue Data Catalog federation. Cross-region and cross-account replication of Apache Iceberg tables added in December 2025.

The Data Protection Audit Checklist

Use this checklist to assess your S3 and data protection controls:

Encryption

- [ ] SSE-KMS with customer managed keys is the default encryption for all sensitive buckets
- [ ] SSE-C is blocked on all buckets that do not require it (`BlockedEncryptionTypes`)
- [ ] S3 Bucket Keys are enabled to reduce KMS costs
- [ ] KMS key rotation is enabled with an appropriate period (90–2,560 days)
- [ ] Post-quantum TLS is evaluated for data subject to long-term confidentiality

Access Control

- [ ] S3 Block Public Access is enabled at the organization level
- [ ] ACLs are disabled (Bucket owner enforced) on all buckets
- [ ] Bucket policies use `aws:PrincipalOrgID` to restrict access to organization principals
- [ ] RCPs enforce data perimeter: trusted identities, trusted resources, expected networks
- [ ] Cross-account access uses explicit conditions (no wildcard principals)
- [ ] S3 Access Grants with Identity Center replace shared IAM roles for user-level access

Resilience

- [ ] S3 Versioning is enabled on all critical buckets
- [ ] S3 Object Lock (compliance mode) protects immutable data
- [ ] Cross-region replication with Object Lock in destination is configured
- [ ] AWS Backup with Vault Lock provides independent backup copies
- [ ] MFA Delete is enabled on buckets containing critical data
- [ ] TLS enforcement (`aws:SecureTransport`) is in all bucket policies

Classification and Monitoring

- [ ] Amazon Macie is enabled with automated sensitive data discovery
- [ ] Custom data identifiers are configured for organization-specific data types
- [ ] CloudTrail data events are enabled for S3
- [ ] GuardDuty S3 Protection and Malware Protection for S3 are enabled
- [ ] EventBridge rules route high-severity S3 findings to automated remediation
- [ ] S3 Storage Lens provides organization-wide visibility into encryption and versioning status

---

Encryption and access controls hold only if an attacker cannot reach the data through the network. The next chapter covers network security: VPC architecture, zero trust, and the controls that keep traffic where it belongs.

Chapter 4: Network Security: VPC, Zero Trust & WAF

Rethinking the Perimeter

The traditional network security model -- build a perimeter, protect what's inside -- was already under pressure when organizations began migrating to the cloud. In 2026, it is obsolete. Workloads span multiple VPCs, accounts, and regions. Services communicate through APIs, not firewall ports. AI agents call AWS services autonomously. And the Verizon 2025 DBIR confirmed the trend: vulnerability exploitation in edge devices and VPNs -- the physical manifestation of perimeter security -- grew from 3% to 22% of CVE-related breaches, nearly an 8x increase.

AWS network security in 2026 is built on layers, not walls. VPCs provide isolation. Security Groups and NACLs enforce traffic rules. VPC Endpoints and PrivateLink keep traffic off the public internet. Verified Access and VPC Lattice apply zero-trust principles at the application and service layers. WAF, Shield, and Network Firewall defend against external threats. Route 53 DNS Firewall blocks malicious domains before a connection is ever established.

What follows covers each layer -- from VPC architecture through zero trust to edge protection -- with the practical detail you need to build a defense-in-depth network stack.

VPC Architecture: The Foundation Layer

Multi-Tier Design

A well-architected VPC separates resources into tiers by exposure and function:

- **Public subnets** -- only resources that must accept inbound internet traffic belong here: Application Load Balancers, NAT Gateways, and bastion hosts (if still used). No application servers or databases should ever sit in a public subnet.
- **Private subnets** -- application workloads (EC2, ECS, Lambda via VPC), internal ALBs, and service endpoints. Outbound internet access goes through NAT Gateways only.
- **Isolated subnets** -- databases (RDS, ElastiCache, DynamoDB Accelerator) and other resources with zero internet connectivity in either direction. No route to an internet gateway or NAT Gateway.

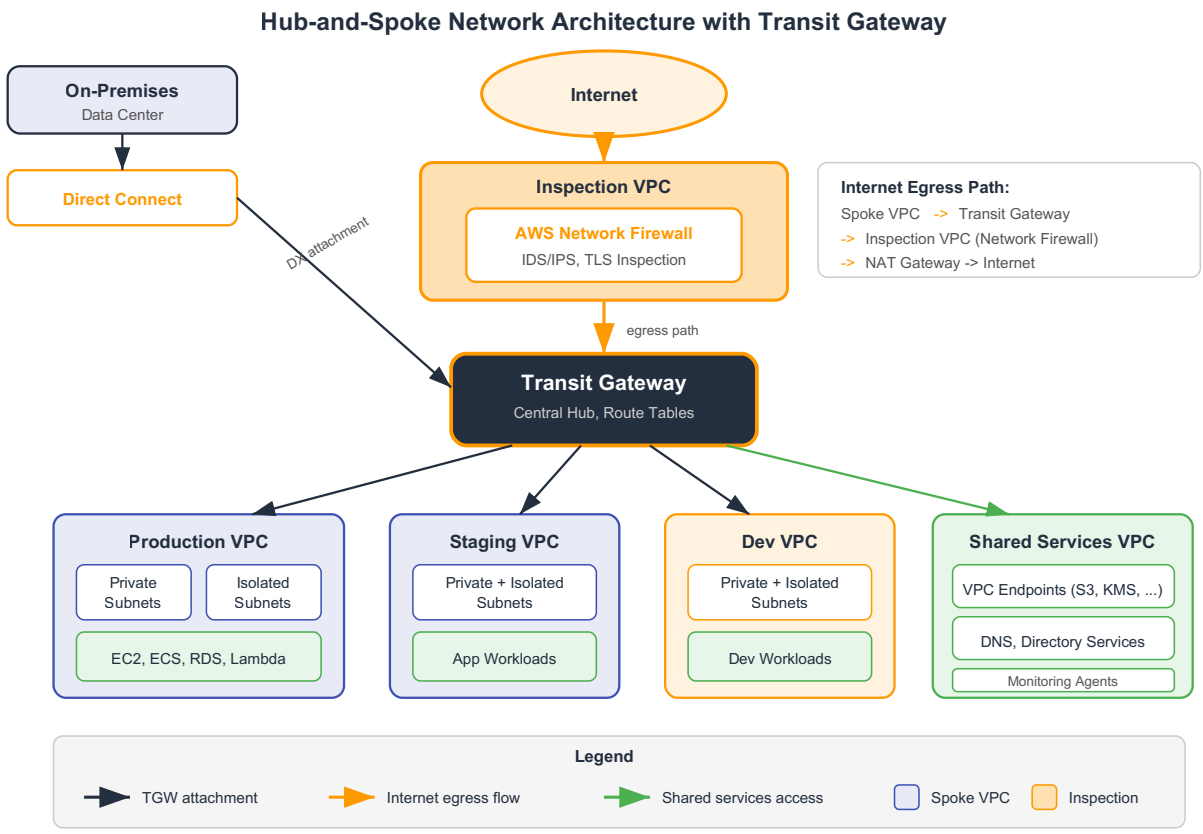
Span each tier across at least two Availability Zones for resilience. Give each tier its own route table to enforce segmentation: the public route table has a route to the internet gateway, the private route table has a route to a NAT Gateway, and the isolated route table has no external routes.

Multi-Account VPC Design

In an AWS Organizations environment, network architecture follows a hub-and-spoke model:

- **Network account** -- hosts the Transit Gateway, shared services VPC (DNS, directory services, monitoring), and centralized egress through AWS Network Firewall or NAT Gateways.
- **Workload accounts** -- each account has one or more VPCs connected to the Transit Gateway via attachments. Transit Gateway route tables control which VPCs can talk to each other.
- **Inspection VPC** -- a dedicated VPC with Network Firewall appliances in the path between spokes and the internet (or between spokes, for east-west inspection).

AWS Transit Gateway supports up to 5,000 attachments per gateway, with inter-region peering for multi-region architectures. In June 2025, AWS launched native Transit Gateway integration for Network Firewall, eliminating the complex routing configurations that centralized inspection VPCs previously required.



Toc Consulting - AWS Security Whitepaper

Security Groups and NACLs

Security Groups: Stateful, Instance-Level

Security Groups operate at the elastic network interface (ENI) level and are **stateful** -- return traffic for an allowed inbound rule is automatically permitted. Key properties:

- **Allow-only rules** -- no deny rules exist. You control access by including only the rules you want; everything else is implicitly denied.
- **Evaluated collectively** -- when multiple Security Groups are attached to an ENI, all rules from all groups are aggregated before evaluation.
- **Self-referencing** -- a Security Group can reference itself, letting resources within the group communicate freely. Standard pattern for application clusters.
- **Cross-group references** -- rules can reference other Security Groups by ID rather than IP ranges. This keeps security intact as instances scale: the web tier's Security Group references the ALB's Security Group, and the database tier references the application tier's Security Group.

- **Default limits:** 60 inbound and 60 outbound rules per Security Group (adjustable via quota increase, subject to the constraint that Security Groups per ENI multiplied by rules per Security Group cannot exceed 1,000 per network interface), and 5 Security Groups per ENI (adjustable to 16).

Security Group Sharing (October 2024). AWS added cross-account Security Group association using VPC associations shared through AWS Resource Access Manager (RAM). In a shared VPC model, the VPC owner shares Security Groups with participant accounts -- centralizing network policy management without giving up account-level workload isolation.

NACLs: Stateless, Subnet-Level

Network ACLs operate at the subnet level and are **stateless** -- you must explicitly allow both inbound and outbound traffic, including return traffic on ephemeral ports. Key differences from Security Groups:

- **Allow and deny rules** -- this is their primary advantage. You can write explicit deny rules that Security Groups cannot express.
- **Processed in rule number order** -- evaluated from lowest number to highest; the first matching rule wins. Fundamentally different from Security Groups, where all rules are evaluated collectively.
- **One NACL per subnet** -- each subnet is associated with exactly one NACL. The default NACL allows all traffic; custom NACLs deny all traffic by default.

When to use NACLs. Security Groups handle the vast majority of traffic control. NACLs add value in two scenarios:

1. **Explicit deny** -- blocking a specific IP range or CIDR that should never reach the subnet, regardless of Security Group rules. Belt and suspenders.
2. **Subnet-wide baseline** -- enforcing a floor of restrictions across all resources in a subnet, independent of individual Security Group configurations.

NACLs are not a replacement for Security Groups, and Security Groups are not a replacement for NACLs. They operate at different layers and complement each other.

What Goes Wrong in Practice

Three mistakes show up repeatedly in Security Group and NACL audits. First, security groups with `0.0.0.0/0` on non-public ports -- SSH (22), RDP (3389), database ports -- left over from debugging sessions that were never cleaned up. These rules are invisible until an attacker finds them. Second, NACLs with overly permissive allow rules because teams do not understand the difference between stateless (NACLs) and stateful (Security Groups) evaluation. A NACL that allows inbound on port 443 also needs an explicit outbound rule for ephemeral ports, or return traffic gets dropped -- so teams open everything outbound to "fix" connectivity, defeating the purpose. Third, not using VPC Flow Logs to verify that traffic actually flows through the expected path. Security Groups and NACLs can look correct in the console and still not behave as intended when route tables send traffic through an unexpected path. Flow Logs are the only way to confirm what is actually happening on the wire.

VPC Endpoints: Keeping Traffic Private

VPC Endpoints connect resources in your VPC to AWS services without traversing the public internet, a NAT Gateway, or a VPN connection. Traffic stays on the AWS private network.

Gateway Endpoints

Gateway Endpoints exist for **S3 and DynamoDB only**. They are route table entries that direct traffic destined for the service to the endpoint. Key characteristics:

- **No cost** -- no hourly charge or per-GB data processing fee.
- **Route table-based** -- you associate the endpoint with specific route tables, and a prefix list entry is automatically added.
- **Endpoint policies** -- attach a policy to the gateway endpoint to control which S3 buckets or DynamoDB tables are reachable through it. One of the strongest controls for data exfiltration prevention.
- **Regional** -- gateway endpoints are regional and cannot be accessed from another region or from on-premises networks.

Interface Endpoints (AWS PrivateLink)

Interface Endpoints create elastic network interfaces (ENIs) with private IP addresses in your subnets. They cover most AWS services and many third-party services in the AWS Marketplace.

- **Per-hour + per-GB pricing** -- hourly charge (varies by region, typically $0.01/hour per AZ) plus a data processing charge.

- **Security Group controlled** -- the endpoint ENI is protected by a Security Group, giving you precise control over which resources can reach the service.
- **Private DNS** -- when turned on, the endpoint overrides the default public DNS name for the service (e.g., `ec2.us-east-1.amazonaws.com` resolves to the private IP of the endpoint).
- **Endpoint policies** -- like gateway endpoints, interface endpoints support policies that restrict which API actions and resources are accessible through the endpoint.

VPC Endpoint Condition Keys (2025)

In August 2025, AWS added three new condition keys for VPC endpoint policies and IAM policies:

- `aws:VpceAccount`: the account that owns the VPC endpoint
- `aws:VpceOrgPaths`: the organizational unit path of the VPC endpoint owner's account
- `aws:VpceOrgID`: the organization ID of the VPC endpoint owner's account

In November 2025, AWS added `aws:SourceVpcArn`, which returns the full ARN of the VPC (including the region) from which the request originates. Particularly useful for region-based access controls -- for example, restricting access to resources in a specific region to VPCs in that same region.

These condition keys unlock data perimeter controls that were previously impossible to express: verify that traffic not only comes through a VPC endpoint, but that the endpoint belongs to a specific account, OU, or organization.

Cross-Region PrivateLink (November 2025)

Before November 2025, interface endpoints were strictly regional. Cross-region PrivateLink removed that restriction: interface endpoints now connect to services in other regions over the AWS private network.

The initial launch supported S3, IAM, ECR, KMS, ECS, Lambda, Amazon Data Firehose, managed Apache Flink, and Route 53 -- eliminating the need for Transit Gateway peering or VPN connections for cross-region service access.

Console Private Access. AWS Management Console Private Access uses VPC endpoints (via AWS PrivateLink), endpoint policies, and SCPs to restrict which AWS accounts can be reached through the Management Console from your network. Browser-based console access is locked to authorized accounts and must flow through your private network -- a strong control for regulated environments.

What Goes Wrong in Practice

The most common VPC endpoint deployment error is forgetting to enable Private DNS. Without it, applications still resolve service endpoints to public IPs and traffic exits through the NAT Gateway (or worse, the internet gateway) instead of staying on the AWS backbone. Always enable Private DNS for interface endpoints, and verify with `nslookup` from within the VPC that the service endpoint resolves to a private IP. If the response comes back with a public IP, your traffic is not using the endpoint.

Zero Trust on AWS

Zero trust is not a product -- it is an architecture principle. The core tenet: **never trust, always verify**. Every request must be authenticated and authorized, regardless of where it originates. AWS delivers zero trust through two services: Verified Access for user-to-application access, and VPC Lattice for service-to-service access.

AWS Verified Access: VPN-Less Application Access

AWS Verified Access grants users access to corporate applications **without a VPN**. It evaluates every request against a policy that checks identity context (from an identity provider) and device security state (from endpoint management solutions), then grants or denies access per-request.

How it works:

1. A user attempts to access an internal application through a Verified Access endpoint.
2. Verified Access evaluates the request against the configured access policy, checking the user's identity via a trust provider (IAM Identity Center or any OIDC-compatible identity provider) and the device's security state (if configured).
3. If the policy conditions are met, traffic is forwarded to the application. If not, the request is denied.
4. Verified Access continuously monitors the connection and revokes access if security requirements change -- for example, if the device's security posture degrades during the session.

Non-HTTP protocol support (GA February 2025). Verified Access originally supported only HTTP/HTTPS. In February 2025, non-HTTP protocol support reached general availability, extending zero-trust access to TCP, SSH, and RDP workloads. This makes full VPN replacement viable for both web applications and remote server access.

Trust providers and device security. Verified Access integrates with:
- **Identity providers** -- IAM Identity Center and any OIDC-compatible provider
- **Device security providers** -- CrowdStrike, Jamf, and JumpCloud for endpoint verification

Access policies are written in Cedar, an open-source policy language developed by AWS. You can write conditions like "allow access if the user is in the engineering group AND the device has disk encryption enabled AND the device's security agent is running."

Full logging. All access attempts -- successful and denied -- are logged with full context (identity, device state, policy evaluation result), giving you an audit trail for every access decision.

VPC Lattice: Service-to-Service Zero Trust

VPC Lattice provides **Layer 7 service-to-service connectivity** with built-in authentication, authorization, load balancing, and observability. Instead of managing VPC peering, Transit Gateway routes, and Security Group rules for inter-service communication, you create a logical "service network" that connects services regardless of VPC, account, or compute platform.

Core concepts:

- **Service network** -- a logical grouping that connects services and applies shared access policies. Services register with the network; clients discover them through it.
- **Service** -- an application (running on Lambda, ECS, EKS, or EC2) published to the service network.
- **Target group** -- the backing compute resources for a service (instances, containers, Lambda functions, or ALBs).
- **Auth policy** -- IAM-based policies that control which principals can invoke which services. These policies use AWS Signature Version 4 (SigV4) or Signature Version 4A (SigV4A) for request authentication.

Protocol support. VPC Lattice supports HTTP, HTTPS, gRPC, TLS passthrough, and TCP. TCP support -- through **Resource Gateway** -- opens access to non-HTTP resources like RDS databases, custom DNS endpoints, and arbitrary TCP endpoints through the service network.

Cross-account and cross-VPC. VPC Lattice supports cross-account service discovery and communication through AWS RAM sharing out of the box. A service in Account A can be shared with Account B through the service network, with auth policies controlling exactly which principals in Account B can invoke which actions.

Regional by default. VPC Lattice is natively regional. Cross-region access works via Resource Gateway and Service Network Endpoints (introduced late 2025), but for broader cross-region service-to-service communication, Transit Gateway, cross-region VPC peering, Direct Connect, or Cloud WAN remain the primary options.

When to use VPC Lattice vs. traditional approaches:

Requirement	Traditional	VPC Lattice
Service-to-service auth	Security Groups + app-level auth	IAM auth policies (SigV4)
Cross-account access	VPC peering + complex SG rules	RAM sharing + auth policies
Load balancing	Deploy and manage ALB/NLB per service	Built-in per-service
Observability	Custom access logs per service	Unified access logs across network
Non-HTTP access	Direct network + SG rules	Resource Gateway

WAF, Shield & DDoS Protection

AWS WAF

AWS WAF filters HTTP/HTTPS requests based on configurable rules. It operates at the application layer (Layer 7) and integrates with CloudFront, ALB, API Gateway, AppSync, Cognito user pools, App Runner, Verified Access, and Amplify.

Automatic Layer 7 DDoS Protection (June 2025). The most significant WAF change in 2025: automatic application-layer DDoS mitigation directly within WAF via the `AWSManagedRulesAntiDDoSRuleSet`. The managed rule group costs $1/month per WebACL (standard AWS Managed Rules subscription fee; additional request-based charges apply) and consumes 50 WCU. WAF builds a machine learning baseline of normal traffic patterns within 15 minutes and responds to volumetric L7 attacks within seconds. A far lighter-weight alternative to Shield Advanced ($3,000/month), though Shield Advanced retains capabilities the WAF rule group does not offer: historical baselining over 24 hours to 30 days, 24/7 Shield Response Team (SRT) support, and DDoS cost protection credits.

WAF Classic End of Life. AWS deprecated WAF Classic (WebACL v1) in 2025: no new WebACL v1 resources could be created after May 1, 2025, and the service reached full end of support on September 30, 2025. Organizations that had not migrated to WAFv2 lost the ability to manage their Classic rules.

New Console Experience (re:Inforce 2025). AWS redesigned the WAF console to cut configuration steps by up to 80% through pre-configured protection packs. Instead of manually assembling rule groups, you select a protection profile and the console generates the appropriate rule configuration.

Web Bot Auth (November 2025). A cryptographic authentication mechanism that verifies automated bot identity using signatures in HTTP messages, based on emerging IETF standards. Verified bots are automatically allowed. Currently available for CloudFront-protected resources.

Key WAF best practices:

1. **Start with AWS Managed Rules** -- the Core Rule Set (CRS), Known Bad Inputs, and SQL Injection rule groups cover the most common attack patterns.
2. **Turn on logging** -- send WAF logs to S3, CloudWatch Logs, or Firehose for analysis. Use the Sampled Requests feature for real-time debugging.
3. **Set rate-based rules** -- limit request rates per IP to block brute force and credential stuffing. Consider separate limits for login endpoints.
4. **Deploy Bot Control for high-value endpoints** -- login pages, APIs, and any endpoint that handles financial transactions.

AWS Shield

AWS Shield offers DDoS protection at two tiers:

Shield Standard -- included with every AWS account at no additional cost. Provides baseline Layer 3 (network) and Layer 4 (transport) DDoS protection for all AWS resources. Edge services -- CloudFront, Route 53, and Global Accelerator -- get enhanced protection, including deeper mitigation against SYN floods, UDP reflection attacks, and DNS amplification attacks.

Shield Advanced -- a paid service ($3,000/month per payer account, 1-year commitment required, plus data transfer fees; covers the payer account and all linked accounts) that adds:

- Automatic application-layer DDoS mitigation (now partially available through WAF's auto L7 DDoS rule group, but Shield Advanced adds historical baseline analysis built over 24 hours to 30 days)
- DDoS cost protection -- AWS credits charges that result from DDoS-related scaling
- 24/7 access to the Shield Response Team (SRT) during active attacks
- Near real-time visibility into attacks with advanced monitoring

Shield Network Security Director (Preview June 2025). A centralized dashboard that shows VPC-level resources, flags missing or misconfigured network security controls, and integrates with Amazon Q Developer for remediation recommendations. In December 2025, multi-account analysis was added, expanding coverage across AWS Organizations.

When Shield Advanced Is Worth the Cost

Shield Advanced makes economic sense when:
- Your application generates revenue that exceeds $3,000/month and cannot tolerate downtime
- You operate in a high-risk industry (finance, gaming, media) where DDoS attacks are frequent
- You need the SRT for incident response and cannot build equivalent expertise in-house
- DDoS-related auto-scaling costs would exceed the Shield Advanced subscription

For most organizations, Shield Standard (free) plus WAF with automatic L7 DDoS protection is enough.

AWS Network Firewall

AWS Network Firewall is a managed, stateful network inspection service running in your VPC. It uses the Suricata-compatible rules engine for deep packet inspection, supporting both IDS (detect) and IPS (detect and block) modes.

Architecture

Network Firewall deploys as firewall endpoints in dedicated subnets within your VPC. Traffic reaches these endpoints through VPC route tables. The standard deployment patterns:

- **Centralized egress inspection:** all outbound internet traffic from spoke VPCs routes through the Network Firewall in a centralized inspection VPC via Transit Gateway.
- **Centralized east-west inspection:** inter-VPC traffic routes through the inspection VPC for lateral movement detection.
- **Distributed inspection:** Network Firewall deployed directly in each workload VPC, typically in front of the NAT Gateway.

Native Transit Gateway Integration (June-July 2025). AWS added native integration between Network Firewall and Transit Gateway, eliminating the need for a dedicated inspection VPC with complex routing configurations (return routes, appliance mode on TGW attachments). This dramatically simplifies centralized inspection architectures and cuts the operational burden of managing network firewalls at scale.

TLS Inspection

Network Firewall decrypts, inspects, and re-encrypts TLS traffic using certificates managed through AWS Certificate Manager (ACM). For outbound TLS inspection, deploy a subordinate CA certificate that the firewall uses to generate certificates for destination domains.

Session Holding. For TLS inspection, Network Firewall holds the initial establishment packets until it can evaluate the Server Name Indication (SNI) from the TLS ClientHello message, then applies domain-based rules before allowing or blocking the connection. Even the first packets of a TLS session are inspected.

Split-Packet Handling (September 2025). AWS improved default rule actions for handling split-packet TLS and HTTP traffic, catching evasion techniques that fragment packets to bypass inspection.

Active Threat Defense (re:Inforce June 2025)

Active Threat Defense feeds Amazon's MadPot threat intelligence directly into Network Firewall managed rule groups. MadPot is AWS's globally distributed network of honeypot threat sensors that detect and catalog emerging threats.

When MadPot identifies new indicators -- malware hosting domains, botnet command-and-control servers, crypto mining pools -- it pushes them to Network Firewall's managed threat intelligence rule groups **within 30 minutes**. You get protection against newly identified threats without writing a single rule.

Network Firewall Proxy (Preview November 2025)

AWS launched Network Firewall Proxy in preview -- a managed egress proxy integrated with NAT Gateway. Instead of transparent deep packet inspection, clients configure proxy settings (e.g., via HTTP proxy environment variables). This delivers:

- URL-level filtering (not just domain-level)
- Full visibility into HTTP headers
- Integration with existing enterprise proxy policies
- Simplified deployment for organizations already using proxy-based architectures

Price Reductions (February 2026)

AWS reduced Network Firewall pricing in February 2026: NAT Gateway data processing discounts were extended to Network Firewall secondary endpoints, and TLS inspection data processing charges were dropped in 13 AWS regions. Encrypted traffic inspection at scale got significantly cheaper.

What Goes Wrong in Practice

Two issues dominate Network Firewall deployments. First, Suricata rule ordering. Suricata evaluates rules in order, and a common mistake is writing a broad allow rule (e.g., `pass tls any any → any any`) that matches before more specific deny rules further down. The allow-all rule fires, the deny rule never executes, and the firewall passes traffic it was supposed to block. Always structure rules so that specific deny rules evaluate before general allow rules, and test with actual traffic to confirm the evaluation order matches your intent. Second, cost surprises. Network Firewall charges per GB of data processed, and in high-throughput environments -- particularly those with TLS inspection enabled -- the bill grows fast. Before enabling Network Firewall on a busy VPC, estimate your data processing volume and model the monthly cost. The February 2026 price reductions help, but the per-GB model still catches teams off guard when traffic spikes.

Route 53 Resolver DNS Firewall

DNS Firewall operates at the DNS resolution layer, blocking queries to known malicious domains before a network connection is ever established. It hooks into Route 53 Resolver -- the DNS resolver used by all resources in your VPC.

How It Works

DNS Firewall evaluates DNS queries against domain lists. You define rule groups with ordered rules, each rule matching queries against a domain list and taking an action (ALLOW, BLOCK, or ALERT). Rule groups are associated with VPCs, and all DNS queries from resources in those VPCs are evaluated.

AWS Managed Domain Lists. AWS provides four managed domain lists that are continuously updated:

1. **AWSManagedDomainsMalwareDomainList**: domains associated with malware distribution
2. **AWSManagedDomainsBotnetCommandandControl**: domains used by botnets for C2 communication
3. **AWSManagedDomainsAggregateThreatList**: a superset list covering domains associated with multiple threat categories including malware, ransomware, botnet, spyware, and DNS tunneling
4. **AWSManagedDomainsAmazonGuardDutyThreatList**: domains associated with GuardDuty DNS security findings

DNS Firewall Advanced

DNS Firewall Advanced goes beyond simple domain list matching:

- **Domain Generation Algorithm (DGA) detection** -- spots DNS queries to algorithmically generated domains used by malware for C2 communication
- **DNS tunneling detection** -- catches attempts to exfiltrate data by encoding it in DNS queries
- **Dictionary DGA detection (November 2025)** -- detects DGA domains composed of concatenated dictionary words, designed to look like legitimate domain names and evade traditional DGA detection

Route 53 Global Resolver (Preview November 2025)

AWS launched the Route 53 Global Resolver in preview -- an internet-reachable, anycast DNS resolver with integrated DNS Firewall. DNS Firewall protection now extends beyond VPCs to any device that can point at a custom DNS resolver, including on-premises networks and remote endpoints.

Security Hub Integration (January 2025)

Since January 2025, DNS Firewall publishes findings to AWS Security Hub, putting DNS-based threats in the same dashboard as GuardDuty, IAM Access Analyzer, and other AWS security service findings.

Network Monitoring and Visibility

Network security without visibility is guesswork. These tools show you what your traffic is actually doing.

VPC Flow Logs

VPC Flow Logs record IP traffic going to and from network interfaces in your VPC. Enable them at the VPC, subnet, or individual ENI level, and publish to CloudWatch Logs, S3, or Amazon Data Firehose.

Key considerations:

- **GuardDuty flow log independence** -- GuardDuty uses its own stream of VPC flow data, not your configured Flow Logs. Enabling or disabling VPC Flow Logs has no effect on GuardDuty's network threat detection.
- **Not packet capture** -- Flow Logs record metadata (source/dest IPs, ports, protocol, bytes, action), not payload content. For deep packet analysis, use Traffic Mirroring.
- **Cost optimization** -- use custom log formats to record only the fields you need, and publish to S3 (Parquet format) for cost-effective long-term storage and Athena queries.

Traffic Mirroring

Traffic Mirroring copies network traffic from ENIs and sends it to monitoring appliances for deep packet inspection. It captures actual packet content -- not just metadata like Flow Logs.

Use Traffic Mirroring for:
- Deep packet inspection with third-party IDS/IPS
- Network forensics and incident response
- Content-based threat detection that requires payload analysis

Network Access Analyzer

Network Access Analyzer identifies network paths that allow unintended access to your resources. Define access scopes (your intended network access patterns), and the analyzer flags paths that fall outside those scopes -- an internet gateway path to an RDS instance, or an unexpected cross-VPC route.

Reachability Analyzer

Reachability Analyzer tests connectivity between two resources in your VPC. It walks through your network configuration (route tables, Security Groups, NACLs, VPC peering connections) and produces a hop-by-hop path analysis showing where connectivity succeeds or fails.

Use it to debug connectivity issues without generating test traffic, and to confirm that network changes have not accidentally opened paths that should not exist.

Recent Enhancements (2025-2026 Summary)

Date	Enhancement	Impact
February 2025	Verified Access non-HTTP protocol support GA (TCP, SSH, RDP)	Full VPN replacement for all access patterns
June 2025	WAF automatic Layer 7 DDoS protection	L7 DDoS defense at WAF pricing, no Shield Advanced required
June 2025	Shield Network Security Director (preview)	Centralized visibility into VPC-level network security controls
June 2025	Network Firewall Active Threat Defense (MadPot)	New threat indicators in managed rules within 30 minutes
June-July 2025	Network Firewall native Transit Gateway integration	Simplified centralized inspection architecture

Date	Enhancement	Impact
August 2025	VPC endpoint condition keys (aws:VpceAccount, aws:VpceOrgPaths, aws:VpceOrgID)	Fine-grained data perimeter controls for VPC endpoints
September 2025	Network Firewall split-packet handling enhancements	Improved evasion detection for TLS and HTTP traffic
September 2025	WAF Classic full end of support	All organizations must be on WAFv2
November 2025	PrivateLink cross-region connectivity	Cross-region service access over AWS private network
November 2025	aws:SourceVpcArn condition key	Region-based access controls using VPC ARN
November 2025	Route 53 Global Resolver (preview)	DNS Firewall protection for any network, not just VPCs
November 2025	DNS Firewall Dictionary DGA detection	Detection of human-readable algorithmically generated domains
November 2025	Network Firewall Proxy (preview)	Managed egress proxy with URL-level filtering
November 2025	WAF Web Bot Auth	Cryptographic bot identity verification for CloudFront
December 2025	Shield Network Security Director multi-account analysis	Organization-wide network security visibility
February 2026	Network Firewall price reductions (TLS inspection data processing charges removed in 13 regions)	Reduced cost for encrypted traffic inspection at scale
February 2026	Bedrock PrivateLink for OpenAI API endpoints	Private connectivity for third-party model APIs

Chapter 4 Audit Checklist

Use this checklist to evaluate your AWS network security controls:

VPC Architecture

- [ ] VPCs use multi-tier subnet design (public, private, isolated)
- [ ] Each tier has its own route table with appropriate routes
- [ ] No application servers or databases reside in public subnets
- [ ] Transit Gateway connects workload VPCs in a hub-and-spoke model
- [ ] Centralized egress inspection is in place for internet-bound traffic

Security Groups & NACLs

- [ ] Security Groups reference other Security Groups (not IP ranges) for inter-tier traffic
- [ ] Default Security Groups allow no inbound or outbound traffic
- [ ] NACLs enforce explicit deny rules for known-malicious IP ranges
- [ ] Security Group rules are reviewed and pruned regularly for unused rules

VPC Endpoints

- [ ] Gateway endpoints deployed for S3 and DynamoDB in every VPC
- [ ] Interface endpoints deployed for all AWS services used in private subnets
- [ ] Endpoint policies restrict access to specific resources (not `"*"`)
- [ ] Private DNS enabled for interface endpoints
- [ ] Cross-region PrivateLink evaluated for multi-region architectures

Zero Trust

- [ ] Verified Access deployed for user-to-application access (replacing VPN)
- [ ] VPC Lattice evaluated for service-to-service communication
- [ ] Lattice auth policies enforce SigV4 authentication for all inter-service calls
- [ ] Verified Access policies include device security checks

WAF & Shield

- [ ] WAF deployed on all internet-facing ALBs, CloudFront distributions, and API Gateways
- [ ] AWS Managed Rules (Core Rule Set, Known Bad Inputs, SQL Injection) enabled
- [ ] Automatic L7 DDoS protection is active
- [ ] Rate-based rules configured for login and sensitive endpoints
- [ ] WAF Classic fully migrated to WAFv2
- [ ] Shield Advanced evaluated for high-value, high-risk applications

Network Firewall

- [ ] Network Firewall deployed for centralized egress inspection

- [ ] Active Threat Defense (MadPot) managed rules enabled
- [ ] TLS inspection configured for outbound traffic to detect encrypted threats
- [ ] Suricata IPS rules tuned for your environment
- [ ] Native Transit Gateway integration adopted (if applicable)

DNS Firewall

- [ ] DNS Firewall rule groups associated with all VPCs
- [ ] All four AWS Managed Domain Lists (Malware, Botnet C2, Aggregate Threat, GuardDuty Threat) enabled
- [ ] DNS Firewall Advanced enabled for DGA and DNS tunneling detection
- [ ] Security Hub integration active for centralized DNS threat visibility

Monitoring

- [ ] VPC Flow Logs enabled for all VPCs (published to S3 in Parquet format)
- [ ] Network Access Analyzer scans run regularly to detect unintended paths
- [ ] Reachability Analyzer used to validate connectivity after network changes

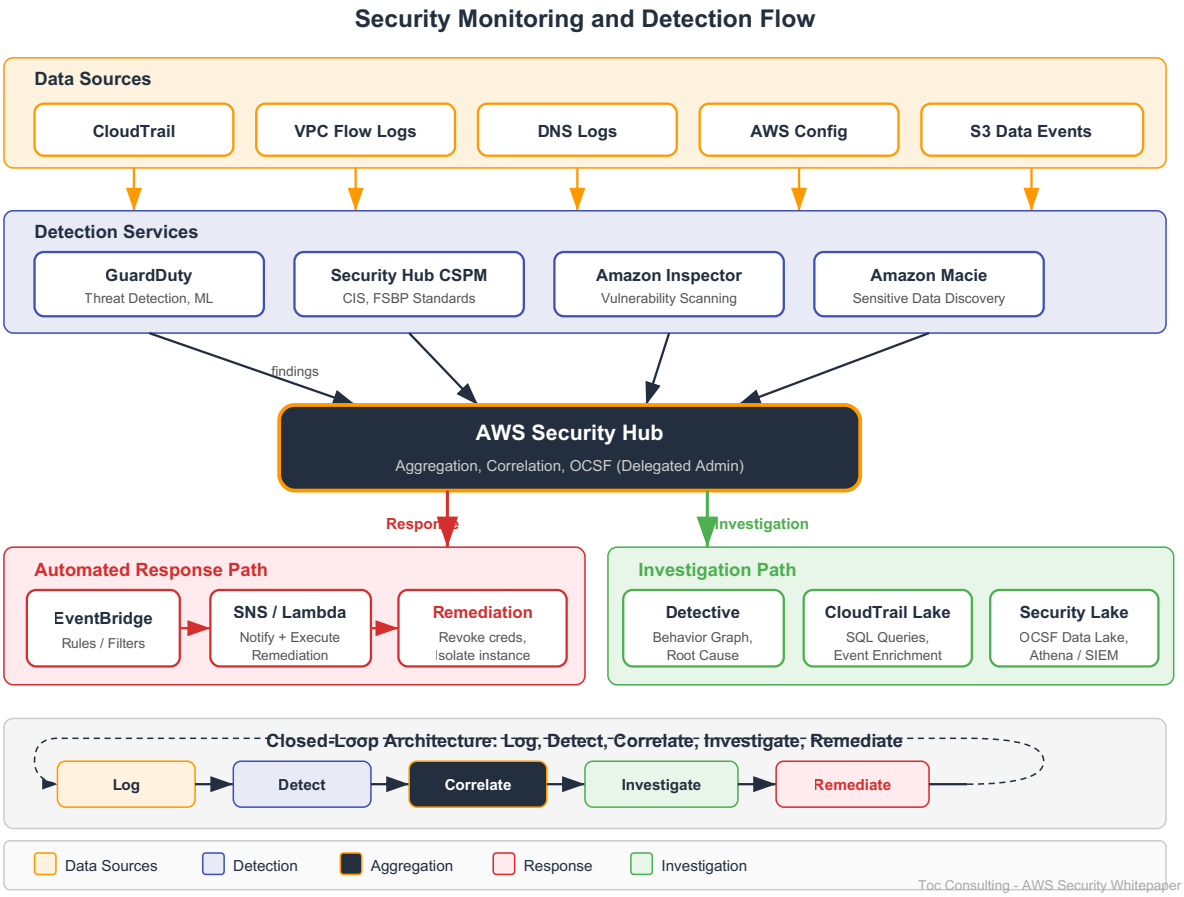
Chapter 5: Monitoring & Detection: See Everything, Miss Nothing

Why Detection Is Not Optional

Preventive controls reduce the attack surface, but no preventive layer is complete. Credentials leak. Misconfigurations slip through code review. Zero-day vulnerabilities exist by definition before patches do. The difference between a security incident and a full breach is how fast you detect and respond.

AWS provides a detection stack that spans API-level audit logging (CloudTrail), continuous configuration assessment (AWS Config), behavioral threat detection (GuardDuty), correlated risk prioritization (Security Hub), deep investigation (Amazon Detective), unified data management (CloudWatch), centralized security data lake (Security Lake), and automated remediation (EventBridge + Lambda). Each service handles a specific detection layer. Together, they form a closed loop: log, detect, correlate, investigate, remediate.

What follows covers each layer with the practical detail you need to operationalize detection -- not just turn it on.



AWS CloudTrail: The Audit Foundation

Organization Trail

Every AWS API call generates an event. CloudTrail records these events and delivers them to S3, CloudWatch Logs, or both. An organization trail, created from the management account, captures management events across every account in the organization automatically -- no per-account configuration required.

Key configuration decisions:

- **Multi-region trail.** A single trail captures events from all regions. Always enable this. Without it, API calls in regions you do not actively use go unrecorded -- exactly where an attacker operates.

- **Management events.** Recorded by default. These cover control plane operations: creating or deleting resources, modifying IAM policies, changing security group rules. Capture management events in both read and write mode -- no exceptions.
- **Data events.** Track data plane operations: S3 object access, Lambda invocations, DynamoDB item operations. They generate high volume and incur additional cost, so enable them selectively for high-value resources (S3 buckets containing sensitive data, production Lambda functions).
- **CloudTrail Insights.** Detects unusual patterns in API call volume and error rates. Originally limited to management events, Insights was extended to data events in November 2025, expanding anomaly detection to cover S3, Lambda, and DynamoDB access patterns. In the same month, AWS launched data event aggregation, which consolidates high-volume data events into 5-minute summaries to reduce noise while preserving investigative value.

Integrity and Tamper Protection

CloudTrail log file integrity validation generates a SHA-256 hash for every log file delivered to S3. A digest file (delivered separately and signed using SHA-256 with RSA) contains the hashes. Use this chain to detect whether log files were modified or deleted after delivery. Combined with S3 Object Lock (compliance mode), it creates a tamper-evident audit trail that meets regulatory requirements for immutable logging.

CloudTrail Lake: SQL-Based Investigation

CloudTrail Lake is a managed, immutable event data store for querying events using SQL (Trino SELECT syntax). Unlike searching raw log files in S3, Lake aggregates and indexes events for fast retrieval across accounts and regions.

Key capabilities:

- **Event enrichment** (May 2025): appends resource tags and global condition keys to events at ingestion time. Query "show all API calls against resources tagged Environment=Production" without correlating separate data sources.
- **Expanded event size.** Events up to 1 MB (up from 256 KB), supporting richer event payloads from services that generate detailed metadata.
- **AI natural language query** (GA in 7 regions): converts plain English questions to SQL queries. Ask "Which IAM users created access keys in the last 7 days?" and Lake generates and executes the corresponding SQL.
- **AI query result summarization** (preview in 3 regions): summarizes query results in natural language, reducing the time to interpret large result sets.
- **SQL JOIN across event data stores.** Correlate events from multiple accounts, organization trails, and non-AWS event sources in a single query.

- **Storage format.** Apache ORC columnar format, optimized for analytical queries.

CloudTrail Lake replaces the manual workflow of downloading log files from S3 and querying them with Athena. For organizations that need investigative capability beyond CloudTrail's default 90-day event history, Lake provides a managed alternative with configurable retention (up to 10 years with one-year extendable pricing, where the first year of retention is included in the ingestion price, or 7-year retention pricing, where all seven years of retention are included in the ingestion price).

> **What Goes Wrong in Practice**
>
> The most common CloudTrail mistake is not enabling data events -- S3 object-level access, Lambda invocations, DynamoDB item operations -- because of cost. Teams skip them to save money, then discover during an incident that they have no visibility into who accessed which objects or when. If you cannot enable data events everywhere, enable them selectively on your highest-value S3 buckets and production Lambda functions. The cost of missing forensic data during a breach dwarfs the cost of logging. A second common mistake: reaching for CloudTrail Lake as the default query tool. Lake is powerful and convenient, but expensive at scale. For cost-effective long-term analysis, deliver CloudTrail logs to S3 in your centralized logging account and query with Athena. Reserve Lake for interactive investigations where speed matters more than cost.

Amazon GuardDuty: Behavioral Threat Detection

GuardDuty analyzes CloudTrail management events, an independent stream of VPC flow data, DNS query logs, and optional data sources from protection plans. It identifies threats using machine learning, anomaly detection, and integrated threat intelligence.

Protection Plans

GuardDuty organizes its detection capabilities into protection plans. Each plan monitors a specific data source or workload type:

Plan	What It Monitors	Enabled by Default
S3 Protection	CloudTrail S3 data events for suspicious access patterns	Yes
EKS Protection	EKS audit logs for Kubernetes-layer threats	Yes
RDS Protection	Login activity on Aurora MySQL, Aurora PostgreSQL, RDS PostgreSQL, and Aurora Limitless databases	Yes

Plan	What It Monitors	Enabled by Default
Lambda Protection	Lambda network activity logs for threats	Yes
Malware Protection for EC2	EBS volume scanning when triggered by suspicious findings	Yes
Malware Protection for S3	Automatic scanning of new S3 object uploads	No
Runtime Monitoring	OS-level process, network, and file activity using eBPF-based security agents on EC2, EKS, and ECS Fargate (deployed as sidecar containers on Fargate)	No

A delegated administrator account can manage GuardDuty for up to 50,000 member accounts.

Extended Threat Detection

Launched December 1, 2024, Extended Threat Detection correlates signals across multiple data sources and time periods to surface multi-stage attacks that individual findings would miss. Each extended threat detection finding carries critical severity and includes a summary, events timeline, MITRE ATT&CK mapping, and remediation steps.

Five attack sequence finding types ship across three launches:

- **IAM and S3** (December 2024): `AttackSequence:IAM/CompromisedCredentials` and `AttackSequence:S3/CompromisedData`. Correlates credential anomalies, API call patterns, and data access to detect credential compromise leading to data exfiltration.
- **EKS** (June 2025): `AttackSequence:EKS/CompromisedCluster`. Correlates Kubernetes audit events, runtime activity, and network behavior to detect cluster compromise.
- **EC2 and ECS** (December 2025): `AttackSequence:EC2/CompromisedInstanceGroup` and `AttackSequence:ECS/CompromisedCluster`. Correlates runtime activity, malware detections, VPC flow data, DNS queries, and CloudTrail events to detect coordinated compromise.

Extended Threat Detection analyzes data GuardDuty already collects -- no additional data sources required. Available in all commercial regions since December 2024, with GovCloud and China Regions added in 2025.

Malware Protection for AWS Backup

Launched November 2025, this scans EBS snapshots, EC2 AMIs, and S3 recovery points for malware. It runs asynchronously and uses incremental scanning -- only changed data since the last scan is analyzed. Enable it independently of other GuardDuty protection plans; organizations that do not use GuardDuty for real-time detection can still scan their backups.

Pricing Note

GuardDuty Malware Protection for S3 received an 85% price reduction in February 2025 (from $0.60 to $0.09 per GB scanned in us-east-1), making automated scanning of S3 uploads economically viable for high-volume workloads.

> **What Goes Wrong in Practice**
>
> These are the GuardDuty finding types that appear most often in real AWS environments: `UnauthorizedAccess:IAMUser/InstanceCredentialExfiltration` (credentials used from outside EC2 -- almost always a real incident), `Recon:EC2/PortProbeUnprotectedPort` (noisy but important if the port is actually open), and `CryptoCurrency:EC2/BitcoinTool.B` (cryptomining -- always a real compromise). Prioritize these three and you will catch the majority of actual incidents. Too many teams treat all GuardDuty findings equally, which leads to alert fatigue and slow response to the findings that matter. To filter for the highest-priority finding type programmatically:
>
> ```
> bash aws guardduty list-findings \ --detector-id <id> \ --finding-criteria '{ "Criterion": { "type": { "Eq": [ "UnauthorizedAccess:IAMUser/InstanceCredentialExfiltration.OutsideAWS" ] } } }'
> ```

AWS Security Hub: Correlated Risk Prioritization

Security Hub aggregates findings from AWS services (GuardDuty, Inspector, Config, Macie, IAM Access Analyzer, Firewall Manager, and more) and third-party integrations into a single pane. In December 2025, AWS split Security Hub into two services: the unified Security Hub for findings aggregation, and Security Hub CSPM for compliance standards and security checks. This section covers both. Aggregation alone is not the value -- correlation is.

Near Real-Time Analytics and Risk Prioritization (GA, December 2025)

Generally available since December 2025, near real-time analytics automatically correlates findings across GuardDuty, Inspector, Macie, and Cloud Security Posture Management (CSPM). When correlated findings reveal exploitable paths, Security Hub generates exposure findings

showing how individual findings combine into attack paths. The potential attack path graph -- an interactive visualization -- lets you trace from an internet-facing resource through misconfigurations to sensitive data.

Key components:

- **Exposure findings.** Auto-generated when correlated findings create an exploitable path.
- **Summary dashboard.** Exposure summaries and threat widgets tracking how your risk profile changes over time.
- **Unified enablement.** Single-click activation across all regions and accounts in an organization.

Security Standards and Controls

Security Hub evaluates your environment against security standards through automated checks:

- **CIS AWS Foundations Benchmark.** Supports v5.0.0 (added October 2025, with 40 automated controls), v3.0.0, v1.4.0, and v1.2.0. AWS recommends v5.0.0 to stay current with security best practices.
- **AWS Foundational Security Best Practices (FSBP).** AWS's own standard, spanning hundreds of controls across AWS services.
- **PCI DSS.** Automated checks for Payment Card Industry compliance.
- **NIST SP 800-53 Rev. 5.** Automated checks mapped to SP 800-53 control families.

In December 2025, AWS added 176 new Security Hub controls to the Control Tower Control Catalog, letting governance teams enforce security standards across landing zones.

Automation Rules

Security Hub supports up to 100 automation rules per delegated administrator account. These rules fire before findings reach EventBridge, so you can suppress noise (auto-archive known false positives), update finding severity, or add notes -- all without writing Lambda functions. For automation that requires external action (remediation, notification), EventBridge rules triggered by Security Hub findings are the integration point.

Partner Integrations via OCSF

Security Hub uses OCSF (Open Cybersecurity Schema Framework) as its data format. OCSF-based partner integrations (including Cribl, CrowdStrike, Datadog, Dynatrace, Expel, SentinelOne, Splunk, and others) send and receive findings in a normalized schema, eliminating custom format transformation.

What Goes Wrong in Practice

A 95% Security Hub compliance score feels reassuring. It should not. Security Hub checks technical controls -- encryption enabled, logging turned on, public access blocked -- but it cannot see configuration drift between deployments, gaps in human processes, or business logic flaws. An environment can score 95% and still have IAM roles with excessive permissions, stale access keys that have not been rotated, or incident response runbooks that no one has tested. Do not let a green dashboard create false confidence. Treat Security Hub as one input to your security posture, not the final word.

Amazon Detective: Investigation and Root Cause Analysis

GuardDuty or Security Hub fires a finding. The next question: what happened? Amazon Detective answers by building a behavior graph from CloudTrail logs, VPC flow data, GuardDuty findings, EKS audit logs, and (via Security Lake integration) additional log sources.

Key Investigation Capabilities

- **Finding group summaries.** GenAI-powered natural language summaries describing what happened, which resources were affected, and the likely impact. Generated using Amazon Bedrock for finding groups (clusters of related findings).
- **IAM investigations.** Analyze IAM users and roles for indicators of compromise. Detective maps observed behavior to MITRE ATT&CK tactics and techniques, showing which API calls match known attack patterns.
- **EKS investigations.** Trace suspicious activity to specific pods, container images, and Kubernetes API calls. Detective correlates EKS audit logs with runtime findings to show the full attack chain within a cluster.
- **Security Lake integration.** Query raw CloudTrail management events, VPC Flow Logs, and EKS audit log data stored in Security Lake directly from the Detective console, giving investigators access to the underlying events behind summarized findings.

Detective is not a tool you configure and forget. It is the investigation layer that turns alerts into answers. When a GuardDuty finding says "anomalous API call from IAM role X," Detective shows you every action that role took, which resources it accessed, and how the behavior compares to the role's historical baseline.

CloudWatch: Unified Data Management

The Problem

Security operations teams deal with data from dozens of sources: AWS services, SaaS platforms (identity providers, EDR tools, vulnerability scanners), and on-premises systems. Each source has its own format, its own storage, and its own query tool. Correlating a suspicious login from Okta with a GuardDuty finding and a CloudTrail event means stitching data across three different platforms.

Unified Data Management (re:Invent 2025)

Announced at re:Invent 2025, CloudWatch unified data management brings operations, security, and compliance data into a single management layer:

- **Built-in normalization.** Ingested data normalizes to OCSF and OpenTelemetry schemas automatically, so events from different sources share a common format.
- **Pre-built connectors.** Connectors for third-party sources including CrowdStrike, SentinelOne, Okta, Entra ID, Wiz, Zscaler, Palo Alto Networks, Microsoft Office 365, Windows Event Logs, GitHub, and ServiceNow CMDB. Each connector handles authentication, data extraction, and schema mapping.
- **Managed S3 Tables.** Ingested data lands in Apache Iceberg tables in managed S3 Tables at no additional storage charge. Query via Amazon Athena, Amazon Redshift, SageMaker Unified Studio, or any Iceberg-compatible tool.

Unified data management eliminates the fragmentation that makes security investigation slow. Instead of context-switching between five consoles, analysts query all data sources from one interface with a shared schema.

AWS Config: Continuous Configuration Assessment

AWS Config continuously records resource configuration changes and evaluates them against rules. When a resource drifts from its desired state -- an S3 bucket goes public, a security group opens port 22 to 0.0.0.0/0, an RDS instance loses encryption -- Config flags the deviation.

Managed Rules Expansion

AWS Config maintains a library of managed rules covering common security and compliance checks. Recent expansions:

- **New managed rules added in November 2025 and January 2026**, covering additional resource types and configuration checks.

- **Conformance packs.** Pre-built collections of rules aligned to frameworks (CIS, NIST, PCI DSS). Expanded to 5 new regions in November 2025.
- **Organization-level deployment.** Deploy Config rules and conformance packs across all accounts in an organization from the delegated administrator account.

Config rules serve as the evaluation engine behind many Security Hub controls. When Security Hub checks whether S3 buckets are encrypted, an AWS Config rule performs the actual evaluation.

AWS Security Agent (Preview)

Announced December 2, 2025 at re:Invent, AWS Security Agent is an AI-powered agent performing automated penetration testing against web applications.

Key capabilities:

- **Automated penetration testing.** Simulates real-world attack scenarios against web applications, covering SQL injection, cross-site scripting (XSS), and server-side request forgery (SSRF).
- **Domain verification.** Requires domain ownership verification before testing begins, preventing unauthorized testing.
- **VPC support.** Tests both public-facing and VPC-hosted applications.
- **Automatic code remediation.** Generates pull requests with fixes for discovered vulnerabilities, integrating directly with GitHub.
- **Integration.** Uses IAM roles for permissions, CloudWatch Logs for test output, and Secrets Manager for authentication credentials.

AWS Security Agent is currently in preview and available in US East (N. Virginia) only.

Amazon Security Lake: Centralized Security Data

Amazon Security Lake automatically centralizes security data from AWS services, SaaS providers, and custom sources into a purpose-built data lake stored in your S3 buckets.

Architecture

- **Schema.** All data normalizes to OCSF format and stores in Apache Parquet columnar format, optimized for analytical queries.
- **Cross-account and cross-region.** Security Lake aggregates data from multiple accounts and rolls it up to delegated regions, creating a single queryable repository.

- **Sources.** Native AWS sources include CloudTrail (management and data events), VPC Flow Logs, Route 53 Resolver query logs, EKS audit logs, Security Hub findings, and WAF logs. Third-party sources integrate through custom source registration.
- **Subscribers.** Query subscribers (Athena, Amazon OpenSearch, third-party SIEMs) and data access subscribers (direct S3 access) consume the normalized data.

Integration Points

- **Amazon Detective.** Queries Security Lake data directly during investigations, providing access to raw events behind summarized findings.
- **Amazon Athena.** Run SQL queries across the full security data lake for ad hoc investigation.
- **SIEM integrations.** Over 30 subscriber integrations and over 50 source integrations from the partner network.

Security Lake is not a detection tool. It is the data foundation that makes detection, investigation, and compliance reporting possible at scale.

Building Automated Remediation

Detection without response is monitoring. Detection with automated response is security. AWS provides the building blocks for closed-loop remediation.

EventBridge + Lambda

The most common remediation pattern:

1. **GuardDuty** or **Security Hub** generates a finding.
2. **EventBridge** matches the finding against a rule (e.g., finding type = "UnauthorizedAccess:IAMUser/MaliciousIPCaller").
3. **Lambda** executes the remediation action (e.g., revoke temporary credentials, isolate the instance by modifying its security group, disable the IAM user's access keys).

SSM Automation Runbooks

For standardized remediation, AWS Systems Manager Automation provides pre-built runbooks:

- **AWS-DisablePublicAccessForSecurityGroup.** Disables SSH and RDP ports (22, 3389) that are open to all IP addresses (0.0.0.0/0 or ::/0).
- **AWSConfigRemediation-EnableEncryptionOnDynamoDbTable.** Switches a DynamoDB table to use a customer-managed KMS key, replacing the default AWS-owned encryption key.
- **Automated Security Response on AWS.** An AWS Solutions Implementation that deploys remediation playbooks for Security Hub findings, providing pre-built runbooks aligned to CIS, FSBP, and PCI DSS standards.

Step Functions for Complex Workflows

When remediation requires multiple steps (isolate instance, capture memory dump, snapshot volumes, notify security team, create JIRA ticket), Step Functions orchestrates the workflow with state management, error handling, and human approval gates.

Design Principles for Automated Remediation

- **Start with alerting, graduate to remediation.** Run every remediation in dry-run mode (alert only) before enabling automatic action. False positives that generate alerts are annoying; false positives that revoke production access are outages.
- **Scope narrowly.** Automate remediation for findings with low false-positive rates and high blast-radius (e.g., public S3 buckets, open security groups). Leave nuanced findings for human review.
- **Log everything.** Every automated remediation action must generate its own CloudTrail event and write to a dedicated remediation log for audit.

The Toc Consulting Monitoring Stack

Built from real-world implementations across multi-account AWS environments, this is the monitoring architecture we deploy:

Foundation Layer

- **CloudTrail organization trail.** Multi-region, management + selective data events, log file integrity validation, delivered to a centralized logging account with S3 Object Lock.
- **AWS Config.** Enabled in all accounts and regions with organization-level conformance packs aligned to CIS v5.0.0 and FSBP.
- **VPC Flow Logs.** Enabled on all VPCs, published to S3 in the centralized logging account (use Amazon Data Firehose for near real-time delivery to Security Lake or a SIEM).

Detection Layer

- **GuardDuty.** Enabled in all accounts and regions with a delegated administrator. All protection plans enabled. Runtime Monitoring enabled for production workloads running containers or sensitive EC2 instances.
- **Security Hub.** Enabled with CIS v5.0.0 and FSBP standards. Automation rules configured to suppress known false positives and enrich findings with organizational context.
- **Amazon Inspector.** Continuous vulnerability scanning for EC2, Lambda, and ECR container images.

Investigation Layer

- **Amazon Detective.** Enabled in all accounts, integrated with Security Lake.
- **CloudTrail Lake.** Organization event data store with enrichment enabled. Security analysts use AI natural language queries for rapid investigation.
- **Security Lake.** Centralized OCSF data lake, aggregating CloudTrail, VPC Flow Logs, Route 53 DNS logs, and third-party sources.

Response Layer

- **EventBridge rules.** Trigger Lambda functions for high-confidence, low-false-positive findings (public S3 bucket, overly permissive security group, disabled CloudTrail).
- **SSM Automation runbooks.** Standardized remediation for Config rule violations.
- **Step Functions.** Orchestrate multi-step incident response workflows with human approval gates.
- **SNS topics.** Notify security teams via email, Slack (via AWS Chatbot), or PagerDuty for findings that require human judgment.

Operational Layer

- **CloudWatch unified data management.** Ingest third-party security data (EDR, identity provider, vulnerability scanner) into a single queryable platform with OCSF normalization.
- **CloudWatch dashboards.** Operational visibility into detection coverage, finding volumes, and remediation success rates.

Incident Response Runbook Template

A runbook is a documented procedure for responding to a specific type of security incident. Every organization must maintain runbooks for at least these scenarios:

1. Compromised IAM Credentials

1. Identify the compromised principal (user, role, access key).
2. Revoke active sessions: attach an inline deny-all policy with an `aws:TokenIssueTime` condition.
3. Disable or delete the compromised access keys.
4. Review CloudTrail for all actions taken by the compromised credentials.
5. Assess impact: what resources were accessed, modified, or deleted?
6. Remediate: rotate credentials, repair any damage, notify affected parties.
7. Post-incident: determine root cause (phishing, exposed key, overly broad permissions) and put preventive controls in place.

2. Public S3 Bucket

1. Confirm the bucket is unintentionally public (some buckets are legitimately public, e.g., static website hosting).
2. Remove the public access grant (bucket policy, ACL, or Block Public Access setting).
3. Enable S3 Block Public Access at the account level if not already enabled.
4. Review CloudTrail S3 data events for unauthorized access.
5. Assess data exposure: what data was in the bucket, and was it accessed?

3. Compromised EC2 Instance

1. Isolate the instance: replace its security group with one that allows no inbound or outbound traffic (do not terminate; preserve evidence).
2. Capture a memory dump (if tooling is available).
3. Snapshot the EBS volumes for forensic analysis.
4. Review VPC Flow Logs and GuardDuty findings for the instance.
5. Determine the attack vector: vulnerable application, compromised key pair, or lateral movement from another resource.
6. Terminate and replace the instance from a known-good AMI.

4. Unauthorized API Activity from Unknown Region

1. Identify the source: CloudTrail shows the region, IP, and user agent.
2. If the activity is from a region you do not use, consider enabling an SCP that denies all actions in unused regions.
3. Investigate the IAM principal: is it a legitimate user operating from an unexpected location, or a compromised credential?
4. Follow the compromised credentials runbook if credential compromise is confirmed.

Audit Checklist

#	Control	How to Verify
1	CloudTrail organization trail enabled, multi-region, management + data events	`aws cloudtrail describe-trails`. Verify `IsOrganizationTrail`, `IsMultiRegionTrail`, and event selectors
2	CloudTrail log file integrity validation enabled	`aws cloudtrail describe-trails`. Verify `LogFileValidationEnabled: true`

#	Control	How to Verify
3	CloudTrail logs delivered to a centralized logging account with S3 Object Lock	Check S3 bucket policy and Object Lock configuration in the logging account
4	GuardDuty enabled in all accounts and regions	`aws guardduty list-detectors` across all accounts; should return a detector ID
5	All GuardDuty protection plans enabled (including Runtime Monitoring for container workloads)	`aws guardduty get-detector`. Check each feature status
6	Security Hub CSPM enabled with CIS v5.0.0 and FSBP standards	`aws securityhub get-enabled-standards`. Verify both standards active
7	Security Hub automation rules configured for known false positives	`aws securityhub list-automation-rules`. Review suppression rules
8	AWS Config enabled in all accounts and regions with organization conformance packs	`aws configservice describe-configuration-recorders` across all accounts
9	Amazon Detective enabled and integrated with Security Lake	`aws detective list-graphs` and verify Security Lake data source
10	VPC Flow Logs enabled on all VPCs	`aws ec2 describe-flow-logs`. Verify all VPCs have associated flow logs
11	Security Lake enabled with CloudTrail, VPC Flow Logs, and Route 53 DNS sources	Check Security Lake console for enabled sources
12	Automated remediation in place for high-confidence findings	Review EventBridge rules targeting Security Hub and GuardDuty events
13	Incident response runbooks documented and tested	Review runbook repository; verify last tabletop exercise date
14	CloudTrail Lake event data store created for investigative queries	`aws cloudtrail list-event-data-stores`. Verify active store with enrichment

Chapter 6: Securing Agentic AI on AWS

The Agentic AI Inflection Point

AI agents, autonomous systems that plan, decide, and act without step-by-step human instructions, are no longer research prototypes. Gartner predicts that 40% of enterprise applications will feature task-specific AI agents by end of 2026, up from less than 5% in 2025. According to Akto's 2025 State of Agentic AI Security report, 69% of enterprises are piloting or have deployed AI agents at scale. The Cisco State of AI Security 2026 report found that 83% of organizations planned to deploy agentic AI capabilities, but only 29% felt truly ready to do so securely.

The security gap is real. Akto found that only 21% of enterprises have full visibility into agent actions, MCP tool invocations, or data access, meaning 79% operate with blindspots where agents invoke tools the security team cannot observe. Only 38% monitor AI traffic end-to-end.

Only 17% monitor agent-to-agent interactions. And Gartner predicts that over 40% of agentic AI projects will be canceled by end of 2027, citing escalating costs, unclear business value, or inadequate risk controls.

This chapter covers the threat surface for agentic AI, the AWS services that secure it, and the architecture patterns that make agent deployments production-ready.

The AWS Agentic AI Stack

Securing AI agents requires understanding the platform they run on. AWS shipped a full stack for agentic AI development in 2025-2026:

Amazon Bedrock AgentCore

AgentCore is AWS's managed platform for building, deploying, and operating AI agents at scale without infrastructure management. It went from preview (July 2025, AWS Summit New York) to general availability (October 2025) and added Policy Controls and Evaluations in preview at re:Invent (December 2, 2025).

AgentCore consists of nine modular services:

Service	Capability
Runtime	Serverless agent execution supporting workloads up to 8 hours, with A2A (Agent-to-Agent) protocol support
Memory	Episodic memory so agents learn and adapt from past experiences
Gateway	Converts APIs, Lambda functions, and existing services into MCP-compatible tools
Browser	Enterprise sandbox environment for web navigation
Code Interpreter	Sandboxed code execution for Python, JavaScript, and TypeScript, with S3 integration and AWS CLI access
Identity	Agent identity and credential management with OAuth 2.0 support, integrating with Cognito, Okta, and Entra ID
Observability	OpenTelemetry-compatible tracing with CloudWatch integration
Policy (preview)	Cedar-based policy controls for agent tool calls (covered below)
Evaluations (preview)	Continuous and on-demand quality assessment (covered below)

AgentCore is framework-agnostic and works with Strands Agents, CrewAI, LangGraph, LlamaIndex, Google ADK, and OpenAI Agents SDK. It supports VPC, PrivateLink, CloudFormation, and resource tagging. Available in 9 regions at GA (October 2025) with consumption-based pricing.

Strands Agents SDK

Strands is AWS's open-source (Apache-2.0) SDK for building AI agents where the LLM handles planning and tool usage on its own. Open-sourced in May 2025, it reached v1.0 in July 2025 and added a TypeScript preview in December 2025. Strands runs in production inside Amazon Q Developer, AWS Glue, VPC Reachability Analyzer, and Kiro.

Strands v1.0 introduced multi-agent primitives and support for the A2A protocol. It includes 20+ pre-built tools, supports MCP servers, and works with multiple LLM providers including Amazon Bedrock, Anthropic, Meta Llama, and Ollama.

Amazon Nova Act

Nova Act is a service for building UI-based workflow automation agents, powered by the Nova 2 Lite model. Generally available since December 2, 2025, it achieves 90% reliability on UI-based workflows (updating CRM records, filling forms, navigating multi-step web processes). Nova Act performs at the top of the WorkArena L1 and REAL Bench V1/V2 benchmarks.

Kiro

Kiro is an AI-powered IDE built on Code OSS, centered on spec-driven development. It entered public preview on July 14, 2025 (AWS Summit New York), with an Autonomous Agent preview announced December 2, 2025 and Kiro Powers (partner integrations via MCP servers) announced at re:Invent 2025.

OWASP Top 10 for Agentic Applications

Released December 2025 by the OWASP GenAI Security Project, this is the industry-standard risk taxonomy for AI agent security. Use it as your threat modeling framework.

ID	Risk	What Can Go Wrong
ASI01	**Agent Goal Hijack**	Hidden prompts in data sources redirect the agent's objectives. A crafted email, document, or web page causes the agent to exfiltrate data instead of summarizing it.
ASI02	**Tool Misuse & Exploitation**	The agent calls legitimate tools in unintended ways, deleting production databases from misunderstood instructions, or escalating privileges through authorized API calls.

ID	Risk	What Can Go Wrong
ASI03	**Identity & Privilege Abuse**	Leaked or cached credentials let agents operate beyond their intended scope. Session tokens persist across tasks.
ASI04	**Agentic Supply Chain Vulnerabilities**	MCP and A2A ecosystems are poisoned: typosquatting, rug pulls (tools silently replaced post-installation), hallucinated dependencies.
ASI05	**Unexpected Code Execution**	Natural-language interactions unlock remote code execution. Agents generate and execute attacker-controlled code.
ASI06	**Memory & Context Poisoning**	Persistent memory or RAG data stores are infected with malicious content that biases future agent reasoning across sessions.
ASI07	**Insecure Inter-Agent Communication**	Spoofed or intercepted messages between agents misdirect entire agent clusters.
ASI08	**Cascading Failures**	A small misstep in one agent propagates through multi-agent workflows, amplifying the impact at each step.
ASI09	**Human-Agent Trust Exploitation**	Confident, well-structured agent explanations mislead operators into approving harmful actions (automation bias).
ASI10	**Rogue Agents**	Misaligned agents deviate from their defined scope while appearing to operate normally.

Real-World Agent Attacks: This Is Not Theoretical

These risks are not hypothetical. The following incidents are documented:

Incident	What Happened	OWASP Risk
EchoLeak (CVE-2025-32711)	Zero-click prompt injection in Microsoft 365 Copilot. Crafted documents (emails, Word files, PowerPoint presentations), processed through Copilot's RAG pipeline, caused automatic data exfiltration without user interaction. CVSS 9.3 (Microsoft CNA rating).	ASI01
GitHub Copilot RCE (CVE-2025-53773)	Prompt injection in GitHub Copilot manipulated VS Code settings to enable auto-approve for shell commands, achieving remote code execution.	ASI05

Incident	What Happened	OWASP Risk
Cursor IDE (CVE-2025-59944)	Case-sensitivity bug in protected file path checking on Windows/macOS allowed attackers to bypass protections on MCP configuration files, escalating to RCE.	ASI04, ASI05
GitHub MCP Server	A malicious public issue contained prompt injection that hijacked the developer's AI assistant, exfiltrating data from private repositories through a broadly-scoped Personal Access Token.	ASI01, ASI02
Supabase Cursor Agent	SQL instructions hidden in support tickets caused the Cursor agent, operating with a service role key that bypassed Row-Level Security, to read and exfiltrate integration tokens from the database.	ASI01, ASI03
Devin AI	Security researchers demonstrated fundamental design weaknesses: attackers exposed ports, leaked tokens, and connected to a command-and-control server through prompt injection.	ASI01, ASI05, ASI10

The pattern across all these incidents: **the agent operated within its technical permissions but was directed to perform actions outside its intended scope.** This is the core challenge of agentic AI security.

The Confused Deputy Problem for AI Agents

The confused deputy problem -- a less-privileged entity coercing a more-privileged entity into performing unauthorized actions -- is a classic security concept. With AI agents, it becomes the default failure mode.

When a Bedrock agent makes downstream API calls, it operates using its service role for orchestration and resource-based policies on downstream services (such as Lambda functions that trust `bedrock.amazonaws.com` as service principal). IAM evaluates these permissions, not the IAM identity of the human user who invoked the agent. The human user only needs `bedrock:InvokeAgent` permission. The agent's subsequent actions may have broader effective permissions than the human user.

The result is a structural vulnerability: an agent with broad permissions can be directed (via prompt injection, manipulated context, or misunderstood instructions) to perform actions that the requesting user was never authorized to do. Acuvity describes this as **semantic privilege escalation**: the agent operates within its technical permissions but beyond the semantic scope of the task.

At enterprise scale, where agents may process tens of thousands of transactions per day, human review of every action is impossible. The defense is not fewer permissions (which would make the agent non-functional). It is **policy controls applied outside the agent's reasoning loop**.

Amazon Bedrock Guardrails: Content-Level Protection

Bedrock Guardrails provides six configurable safeguard policies that filter content flowing into and out of foundation models:

1. **Content Filters.** Detect and block harmful content across five categories: Hate, Insults, Sexual, Violence, and Misconduct. Each category has configurable strength levels. **Prompt Attack** is an additional content filter category that detects prompt injection and jailbreak attempts on input, and can also be applied to detect attacks in third-party model responses.
2. **Denied Topics.** Define topics the model should refuse to engage with (e.g., competitor product recommendations, financial advice).
3. **Word Filters.** Block specific words, phrases, and profanity.
4. **Sensitive Information Filters.** Detect PII (names, addresses, credit card numbers) and custom patterns defined via regex.
5. **Contextual Grounding Checks.** Detect hallucinations by comparing model output against source material and the original prompt.
6. **Automated Reasoning Checks.** GA since August 2025, this is the first and only generative AI safeguard designed to help prevent factual errors from hallucinations, using formal logic and mathematical reasoning to deliver up to 99% accuracy at detecting correct responses from LLMs. Automated Reasoning translates domain rules into formal logical models and mathematically verifies whether the model's output satisfies those rules.

Coding Use Case Support (November 2025)

Guardrails now detects harmful content within code: comments, variable and function names, and string literals. It supports 12 programming languages and can identify harmful content embedded in code output across the supported content filter categories.

Cross-Model Support

The ApplyGuardrail API works with any foundation model, whether Bedrock-hosted, self-hosted on SageMaker or EC2, or third-party (OpenAI, Google Gemini). Guardrails is not limited to Bedrock models.

AgentCore Policy Controls: Action-Level Protection

Guardrails protects content. Policy Controls protects actions. The distinction matters: an agent can generate perfectly safe content while making devastating tool calls.

Cedar Policy Language

AgentCore Policy Controls use Cedar, an open-source policy language designed for fine-grained authorization. Cedar policies are applied outside the agent's reasoning loop, integrated with AgentCore Gateway to intercept every tool call in real time. The agent cannot bypass, override, or reason its way around a Cedar policy.

Key Cedar properties:

- **"Forbids override permits."** A `forbid` policy denies the action regardless of any `permit` policy. This is a fundamental safety property: you can express absolute boundaries that cannot be overridden.
- **Default deny.** No action is allowed unless a specific `permit` policy grants it.
- **Sub-millisecond latency.** Cedar is designed for less than 1ms typical evaluation time, even with hundreds of policies.
- **No loops, no side effects.** Cedar policies are purely declarative. They cannot execute code, make network calls, or modify state.

Natural Language to Cedar

Developers can describe rules in plain English (for example, "the agent should never delete any production database") and the system generates candidate Cedar policies, validates them against the tool schema, and uses automated reasoning to check safety conditions (identifying policies that are always-allow, always-deny, or contain unsatisfiable conditions).

Why This Matters

Consider an agent with access to a database tool. Without policy controls, the agent might execute `DROP TABLE` if a prompt injection instructs it to. With Cedar:

```
forbid (
  principal,
  action == Action::"database.execute",
  resource
)
when { context.sql_statement like "*DROP*" };
```

This policy blocks destructive SQL regardless of what the agent's reasoning loop decides. The control is external, deterministic, and auditable.

A More Realistic Example: Finance Agent Policies

The `DROP` blocker above illustrates the mechanism, but production Cedar policies need to express business logic, not just block dangerous keywords. Consider a finance agent that can query transactions but must not initiate large transfers or bulk-export records:

```
// Illustrative Cedar policies for a finance agent.
// Syntax follows Cedar's permit/forbid model; adapt to your schema.
// Entity types (Agent, Action, Account) and schema depend on your
// AgentCore Gateway configuration -- these are not built-in types.

// Allow the finance agent to query transactions in the production account
permit(
  principal == Agent::"finance-analyst",
  action == Action::"query_transactions",
  resource in Account::"prod-finance"
);

// Block transfers above the approved threshold
forbid(
  principal == Agent::"finance-analyst",
  action == Action::"initiate_transfer",
  resource
) when { context.amount > 10000 };

// Block bulk data exports that could be exfiltration
forbid(
  principal == Agent::"finance-analyst",
  action == Action::"export_data",
  resource
) when { context.record_count > 100 };
```

This is closer to real-world policy design because it combines three distinct controls: a scoped `permit` that limits which resources the agent can access, a `forbid` with a financial threshold that prevents high-value unauthorized transfers, and a `forbid` with a volume threshold that blocks bulk data exfiltration. Because Cedar's "forbids override permits" rule is absolute, the agent cannot reason its way around these boundaries regardless of prompt content.

AgentCore Evaluations: Continuous Quality Assessment

Policy controls define what agents cannot do. Evaluations measure what agents actually do, and how well.

AgentCore Evaluations (preview) provides two modes:

- **Online Evaluation.** Continuously monitors deployed agents using live production traffic, assessing quality across multiple dimensions in real time.

- **On-Demand Evaluation.** Targeted assessments of specific traces or spans, useful for investigating customer-reported issues or validating fixes.

Evaluations includes 13 built-in evaluators covering dimensions such as correctness, helpfulness, tool selection accuracy, safety, goal success rate, and context relevance. Custom evaluators are supported, including LLM-as-a-Judge techniques. Evaluations integrates with Strands and LangGraph via OpenTelemetry and OpenInference.

AgentCore Evaluations is available in preview in select regions. See the AWS supported regions page for current availability.

AgentCore Identity: Who Is the Agent?

Traditional IAM answers "who is this user?" For AI agents, the question expands: who is this agent, who authorized it, and what credentials should it use?

AgentCore Identity provides:

- **Workload identities.** A centralized registry for managing agent identities. Each agent is implemented as a workload identity with specialized attributes that can be tracked across services.
- **Inbound JWT authorizer.** Validates incoming JWT bearer tokens to determine whether a user or service is allowed to invoke a specific agent, adding an authorization layer before the agent begins executing.
- **OAuth 2.0 support.** Agents can authenticate to third-party services using standard OAuth flows, with built-in credential providers for services integrated via Cognito, Okta, and Entra ID.

Without agent-level identity, you cannot distinguish between an agent acting on behalf of an authorized user and one that has been hijacked. Agent identity closes that gap and creates the audit trail needed for forensic investigation and compliance.

MCP Security Risks

The Model Context Protocol (MCP), now the standard for connecting agents to tools, introduces its own attack surface:

- **Tool Poisoning.** Malicious instructions hidden in MCP tool descriptions are invisible to users (many clients truncate or hide long descriptions) but fully visible to the LLM, which reads and obeys them. Research from Invariant Labs and Acuvity demonstrates that this attack is practical and effective.

- **Rug Pull Attacks.** A legitimate MCP tool is silently replaced with a malicious version after installation. Most MCP clients do not re-prompt the user for approval after the initial install, so the modified tool executes without user awareness.
- **Prompt Injection via MCP.** Untrusted content (documents, tickets, database records) enters the LLM's context window through MCP tool responses and is interpreted as instructions, coercing the model into unintended tool calls.

Defenses

- Pin MCP tool versions and verify tool integrity on every connection.
- Apply Cedar policies via AgentCore Gateway to restrict which MCP tools an agent can call and under what conditions.
- Treat all MCP tool output as untrusted data. Never allow it to modify agent system prompts or override policy controls.
- Use AgentCore Identity to authenticate MCP servers, not just the tools they expose.

Bedrock Model Invocation Logging

Log every AI model call. Bedrock Model Invocation Logging captures the full request/response payload and metadata for all model invocations in an account and region.

Key details:

- **Disabled by default.** This must be explicitly enabled. Without it, you have no audit trail of what prompts were sent to models or what responses were generated.
- **Destinations.** CloudWatch Logs (JSON format, payloads up to 100 KB) and S3 (for larger data including images). Both destinations can be enabled simultaneously.

Bedrock also publishes CloudWatch runtime metrics automatically (invocation count, input/output token counts, error rates) under the `AWS/Bedrock` namespace. These are separate from invocation logging and require no additional configuration. Use CloudWatch alarms on these metrics to detect anomalies (sudden spikes in token consumption may indicate prompt injection attacks or agent loops).

Model invocation logging is the AI equivalent of CloudTrail. Enable it in every account and region where Bedrock is used.

AWS Security Incident Response: Agentic AI Investigation

AWS Security Incident Response (launched December 2024) added agentic AI-powered investigation in November 2025. It automatically gathers evidence across CloudTrail, IAM, EC2, and Cost Explorer, correlates the data, and presents actionable summaries. This capability is enabled for all Security Incident Response customers at no additional cost.

It cuts the time from alert to understanding by automating evidence collection that analysts normally do by hand: querying CloudTrail for API activity, checking IAM for newly created users or roles, reviewing EC2 for launched instances, and pulling Cost Explorer data for unusual resource consumption.

Securing the AI Data Pipeline

Agentic AI security extends beyond runtime controls. The data that trains, fine-tunes, and grounds AI models requires its own protection layer:

- **Training data on S3.** Encrypt with KMS (SSE-KMS), enforce bucket policies that restrict access to the training pipeline's IAM roles, enable S3 Object Lock for immutable training datasets, and enable Macie to detect sensitive data in training corpora.
- **Model artifacts.** Store in S3 with versioning and access logging. For container-based models, store images in ECR with image scanning enabled and enforce image signing.
- **VPC endpoints for Bedrock.** Route all Bedrock API calls through VPC endpoints (PrivateLink) to keep prompts and responses off the public internet. This is essential for data residency requirements; prompts and responses stay within the AWS network.
- **Prompt and response logging.** Enable Bedrock Model Invocation Logging (covered above) and deliver logs to a centralized security account with S3 Object Lock for immutability.

Cost of the Agentic AI Security Stack

The security layers described in this chapter add operational cost on top of base model invocation charges. The primary cost drivers are Bedrock Guardrails (charged per text unit assessed), model invocation logging (CloudWatch Logs ingestion and storage), CloudWatch alarms, and AgentCore service consumption. For a production agent handling moderate traffic, expect the security overhead to add roughly 10-20% to base model invocation costs. The exact figure depends on prompt length, invocation frequency, and how many guardrail policies are evaluated per request. The highest-leverage controls -- IAM least privilege, Cedar policies, and permissions boundaries -- add zero marginal cost. Start there, then layer on Guardrails and logging as the agent moves from development to production.

Architecture Pattern: Secure Agentic AI on AWS

This reference architecture combines the services covered in this chapter into a defense-in-depth deployment for production AI agents:

Layer 1: Identity and Authorization

- **AgentCore Identity.** Every agent has a verifiable identity in the agent directory.
- **IAM least privilege.** Agent execution roles follow least-privilege principles. Use AgentCore Policy Controls (Cedar) to constrain agent tool calls at runtime.
- **Permissions boundaries.** Set maximum permissions boundaries on agent roles to prevent privilege escalation, even if the role policy is misconfigured.

Layer 2: Content Protection

- **Bedrock Guardrails.** Applied to all model invocations. Content filters block harmful input/output. Sensitive information filters prevent PII leakage. Automated Reasoning validates compliance with domain-specific rules using formal logic. Coding safeguards detect malicious code.
- **Input validation.** Validate and sanitize all data entering the agent's context window. Treat MCP tool responses as untrusted input.

Layer 3: Action Control

- **AgentCore Policy Controls (Cedar).** Define absolute boundaries for agent tool calls. Block destructive operations. Restrict data access by scope. Enforce rate limits.
- **AgentCore Gateway.** Intercepts every tool call in real time, evaluating it against Cedar policies before execution.

Layer 4: Monitoring and Detection

- **Bedrock Model Invocation Logging.** Full audit trail of all model interactions.
- **AgentCore Evaluations.** Continuous quality monitoring of deployed agents.
- **AgentCore Observability.** OpenTelemetry traces for every agent execution.
- **GuardDuty.** Detect anomalous API activity from agent execution roles.
- **CloudTrail.** Audit trail for all AWS API calls made by agents.
- **CloudWatch alarms.** Detect anomalies in token consumption, error rates, and tool call frequency.

Layer 5: Emergency Controls (Kill Switch)

A misbehaving agent can cause damage in seconds. The kill switch is not a single button; it is a sequence of containment actions executed in order of urgency.

Step 1: Immediate containment. Revoke the agent's IAM role session by applying a deny-all inline policy to the execution role. This takes effect on the next API call. If the agent runs in AgentCore Runtime, terminate the runtime session directly. Both actions can be triggered by a Lambda function invoked through EventBridge, with no human in the loop.

Step 2: Automated circuit breakers. CloudWatch alarms should monitor three signals: token consumption rate (a sudden spike may indicate a prompt injection loop), API call volume from the agent's execution role (runaway tool invocations), and error rates (repeated failures suggest the agent is attempting actions outside its permissions). When any alarm breaches its threshold, an EventBridge rule triggers a Lambda function that disables the execution role entirely, cutting off all AWS API access.

Step 3: Network isolation. Modify the security group attached to the agent's compute environment (Lambda VPC configuration, ECS task, or EC2 instance) to remove all inbound and outbound rules. This severs the agent's network access, preventing data exfiltration or lateral movement even if IAM containment is delayed.

Step 4: Bedrock model access revocation. Remove the agent's access to foundation models by revoking Bedrock model access permissions on the execution role. This prevents the agent from making any further model invocations, stopping the reasoning loop at its source.

Step 5: Post-incident forensics. Preserve all evidence before making changes to the environment. CloudTrail logs capture every API call the agent made. Bedrock Model Invocation Logs capture every prompt and response. AgentCore Observability traces capture the full execution path. Copy these logs to an immutable S3 bucket (Object Lock) in a separate security account. Document the agent's last known state, the triggering event, and the containment timeline.

Test this procedure quarterly. Run through all five steps in a staging environment with a deliberately misbehaving test agent. A kill switch you have never tested is a kill switch that fails during a real incident.

Human approval gates. For high-risk actions (financial transactions, data deletion, production deployments), require explicit human approval via Step Functions before the agent proceeds. This is not a kill switch; it is a preventive control that reduces the need for one.

For organizations running agents across multiple cloud providers, the IAM-based controls described here should be complemented by a centralized policy engine and unified audit trail spanning all environments.

Checklist: 15 Controls for Agentic AI Security

#	Control	How to Implement
1	Every agent has a unique, verifiable identity	Use AgentCore Identity agent directory; avoid shared service roles
2	Agent execution roles follow least privilege	Scope IAM policies to specific resources; use permissions boundaries; use Cedar policies to constrain tool calls per agent
3	Cedar policies define absolute action boundaries	Deploy AgentCore Policy Controls; start with `forbid` policies for destructive operations
4	Bedrock Guardrails applied to all model invocations	Turn on all six safeguard policies; configure Automated Reasoning for domain-specific rules
5	Model invocation logging enabled in every account/region	Enable Bedrock Model Invocation Logging with CloudWatch Logs and S3 destinations
6	All Bedrock API calls route through VPC endpoints	Create interface VPC endpoints for Bedrock; enforce via IAM condition key `aws:SourceVpce`
7	MCP tool versions pinned and integrity verified	Verify tool definitions on every connection; block unauthorized tool changes via Cedar policies
8	Agent-to-agent communication authenticated	Use A2A protocol with identity verification; authenticate MCP servers via AgentCore Identity
9	Training data and model artifacts encrypted and access-controlled	SSE-KMS on S3; ECR image scanning; bucket policies restricting access to pipeline roles
10	Agent quality continuously monitored	Activate AgentCore Evaluations (online mode) for production agents
11	Anomaly detection on agent behavior	CloudWatch alarms on token consumption, error rates, and tool call frequency; GuardDuty for API anomalies
12	Kill switch implemented for all production agents	EventBridge + Lambda to revoke credentials or terminate sessions on guardrail violations
13	Human approval required for high-risk actions	Step Functions approval gates for financial, deletion, and production-impacting operations
14	Prompt injection defenses in place	Guardrails Prompt Attack filter enabled; input sanitization on all MCP tool responses; output validation
15	Incident response runbook includes agent scenarios	Add compromised agent, prompt injection, and data exfiltration via agent to IR playbooks

Chapter 7: Compliance & Audit: Frameworks That Matter

Why Compliance Is a Security Accelerator

Compliance is not a checkbox exercise. Frameworks like SOC 2, ISO 27001, PCI DSS, and NIS2 encode decades of collective security experience into structured control sets. Adopting them forces the architectural discipline that prevents breaches: logging, access control, encryption, incident response, and continuous monitoring. Done well, compliance hardens your AWS environment by design. Done poorly, it creates a paper trail that obscures real risk.

AWS supports 143 security standards and compliance certifications. But AWS certifications cover AWS infrastructure only. Your workloads, configurations, IAM policies, and data handling are your responsibility under the Shared Responsibility Model. This chapter maps the major frameworks to AWS services, highlights the requirements that catch organizations off guard, and describes the audit methodology Toc Consulting uses to close the gap between certification and actual security.

SOC 2 Type II

SOC 2 Type II reports assess the operating effectiveness of controls over a period of time (typically 6–12 months) across up to five Trust Services Criteria: Security, Availability, Processing Integrity, Confidentiality, and Privacy.

AWS scope. AWS publishes SOC 1, SOC 2, and SOC 3 reports. The Fall 2025 reports (covering October 2024 through September 2025) include **185 services** in scope, up from 184 in the Spring 2025 reports. SOC 1 and SOC 2 reports are available through AWS Artifact under NDA. SOC 3 reports are publicly available.

What AWS covers, and what you must cover. AWS's SOC 2 report covers physical security, host operating systems, virtualization infrastructure, and service-level controls. Everything above the hypervisor -- your application logic, IAM policies, encryption configuration, logging, and monitoring -- is your responsibility. Each SOC 2 report lists **Complementary User Entity Controls (CUECs)** that customers must implement. CUECs are not optional; your auditor will test them.

AWS published the AICPA SOC 2 Compliance Guide on AWS in July 2025, providing detailed mapping of Trust Services Criteria to AWS services and customer controls.

Key AWS services for SOC 2 evidence:

Trust Services Criteria	AWS Services
Security (CC1–CC9)	IAM, CloudTrail, GuardDuty, Security Hub, AWS Config
Availability (A1)	Route 53, ELB, Auto Scaling, Multi-AZ deployments, Resilience Hub
Processing Integrity (PI1)	CloudWatch, X-Ray, Lambda dead-letter queues
Confidentiality (C1)	KMS, S3 encryption, Secrets Manager, Macie
Privacy (P1–P8)	Macie, IAM Access Analyzer, S3 Object Lock

ISO 27001:2022

ISO/IEC 27001 specifies requirements for building, running, and continually improving an information security management system (ISMS). AWS transitioned from ISO 27001:2013 to **ISO/IEC 27001:2022**; certificates were reissued on November 22, 2023 by Ernst and Young CertifyPoint. AWS also holds ISO/IEC 27017:2015 (cloud security controls) and ISO/IEC 27018:2019 (PII in public cloud).

AWS certification does not equal your certification. AWS's ISO 27001 certificate covers AWS infrastructure and managed services. Your organization must build and certify its own ISMS, with AWS as a component. Your auditor will examine how you apply Annex A controls within your AWS environment: access management, cryptography, operational security, and supplier relationships.

AWS Audit Manager provides a prebuilt framework for **ISO/IEC 27001:2013 Annex A** (not the 2022 revision). This framework automates evidence collection from CloudTrail, Config, and Security Hub against the 2013 Annex A controls. For the 2022 revision, you will need to map the new Annex A structure (which consolidated 114 controls into 93) to the automated evidence collected by Audit Manager, supplementing with custom controls where needed.

CIS AWS Foundations Benchmark v5.0.0

The Center for Internet Security (CIS) AWS Foundations Benchmark provides prescriptive hardening guidance for AWS accounts. CIS has released **v6.0.0**, but the latest version supported by AWS Security Hub CSPM is **v5.0.0**.

Security Hub supports CIS AWS Foundations Benchmark **v5.0.0, v3.0.0, v1.4.0, and v1.2.0**. AWS recommends v5.0.0, which includes **40 automated controls** and is available in all AWS Regions where Security Hub CSPM is available, including GovCloud (US) and China Regions.

The benchmark covers identity and access management, logging, monitoring, networking, and storage controls. Unlike compliance frameworks (SOC 2, ISO 27001), CIS benchmarks are technical: they specify exact configuration settings, CLI commands, and console steps -- ideal starting points for organizations that need actionable security baselines rather than policy-level requirements.

GDPR on AWS

The General Data Protection Regulation (GDPR) applies to any organization that processes personal data of individuals in the European Union, regardless of where the organization is located.

AWS's role under GDPR. AWS acts as a data processor with respect to customer content processed through AWS services, and as a data controller for account and operational data (such as registration and billing information). The AWS Data Processing Addendum (DPA) applies automatically to all customers who require it under GDPR. The DPA includes Standard Contractual Clauses (SCCs) for transfers of personal data outside the EU.

Data residency. Data stored in an AWS Region stays in that Region unless you explicitly move it. AWS offers eight European Regions for data residency:

Region	Location	EU Member State
eu-west-1	Ireland	Yes
eu-west-2	London	No (UK)
eu-west-3	Paris	Yes
eu-central-1	Frankfurt	Yes
eu-central-2	Zurich	No (Switzerland)
eu-south-1	Milan	Yes
eu-south-2	Spain	Yes
eu-north-1	Stockholm	Yes

For organizations requiring data processing exclusively within the EU: the six EU member state Regions (Ireland, Paris, Frankfurt, Milan, Spain, Stockholm) keep data within EU borders. The UK and Switzerland Regions are European but are not EU member states and require separate adequacy decisions for GDPR transfers.

AWS European Sovereign Cloud. Launched **January 15, 2026**, the AWS European Sovereign Cloud is a more than EUR 7.8 billion investment located in Brandenburg, Germany. It is operated by EU-based personnel (initially a blended team of EU residents and EU citizens, with AWS gradually transitioning to exclusively EU citizens for all operational roles) and is physically and logically separate from all other AWS Regions. Planned expansion includes Sovereign Local Zones in Belgium, Netherlands, and Portugal. This Region is designed for public sector and regulated organizations that require operational sovereignty within the EU.

Key GDPR services. Amazon Macie discovers and classifies PII in S3. IAM Access Analyzer validates that resources are not unintentionally shared externally. S3 Object Lock prevents deletion (supporting retention obligations). CloudTrail provides the audit trail for demonstrating lawful processing.

PCI DSS v4.0.1

PCI DSS v4.0.1 was published June 11, 2024 as a limited revision to v4.0 (no new requirements). **All future-dated requirements became mandatory on March 31, 2025.**

Key requirements that affect AWS environments:

- **Requirement 8.4.2**: MFA is now required for **all access into the Cardholder Data Environment (CDE)**, not just administrative access as under v3.2.1. This includes console access, API access, and any authentication to systems within the CDE boundary.
- **Requirement 8.3.6**: Minimum password length increased to **12 characters** (was 7 under v3.2.1). An exception permits 8 characters minimum for systems that cannot support 12.
- **Requirement 6.4.3**: All payment page scripts must be inventoried, authorized, and integrity-checked.
- **Requirement 5.4.1**: Automated mechanisms to detect and protect against phishing.

Security Hub CSPM provides automated controls for PCI DSS v4.0.1, enabling continuous compliance checking against these requirements.

AWS Artifact provides the Fall 2025 PCI DSS compliance package, including the Attestation of Compliance (AOC) and Responsibility Summary.

HIPAA on AWS

The Health Insurance Portability and Accountability Act (HIPAA) requires a **Business Associate Agreement (BAA)** between covered entities and their service providers. AWS offers a BAA that covers **HIPAA-eligible services** only, not all AWS services.

HIPAA-eligible services now include generative AI capabilities: **Amazon Bedrock**, **Amazon SageMaker AI**, and **Amazon Q Business** are all covered under the BAA. Healthcare organizations can now build GenAI applications on AWS without breaking HIPAA compliance.

You must configure security controls within eligible services yourself. Enabling a service does not make it HIPAA-compliant. Apply encryption, access controls, audit logging, and data handling procedures as specified in the HIPAA Security Rule.

NIS2 Directive

The Network and Information Security Directive 2 (NIS2) has been applicable across the EU since **October 18, 2024**, succeeding the original 2016 NIS Directive. It introduces heightened security requirements and stricter incident reporting obligations for essential and important entities across 18 sectors.

Transposition is uneven. NIS2 is an EU Directive, meaning each member state must transpose it into national law. As of early 2026, not all member states have completed transposition. Monitor national implementation timelines in every EU country where you operate.

What NIS2 requires. Entities must implement risk management measures covering incident handling, business continuity, supply chain security, network security, and vulnerability disclosure. Report security incidents to the CSIRT or competent authority within 24 hours (early warning) and 72 hours (incident notification).

AWS published the NIS 2 Considerations for AWS Customers whitepaper (originally November 2023, updated December 2025), which maps NIS2 Article 21 requirements to corresponding AWS services. The December 2025 update added mappings to the ENISA Technical Implementation Guidance published in June 2025. AWS's 143 security standards and compliance certifications support the supply chain assurance requirements of Article 21(2)(d).

Key AWS services for NIS2: Security Hub (continuous security monitoring), GuardDuty (threat detection for incident handling), AWS Config (configuration compliance), CloudTrail (audit logging), AWS Security Incident Response (incident notification support).

DORA (Digital Operational Resilience Act)

The Digital Operational Resilience Act (DORA) became effective **January 17, 2025**. It applies to approximately 22,000 financial entities and ICT service providers across the EU, including banks, insurers, investment firms, and their key ICT third-party providers.

AWS as Critical Third-Party Provider. On November 18, 2025, the European Supervisory Authorities (ESAs) designated **AWS as one of 19 Critical Third-Party Providers (CTPPs)** (as designated by ESAs in their initial register) under DORA. This subjects AWS to direct ESA oversight in areas including risk management and governance, incident reporting, subcontracting, and ICT security.

AWS has released three resources for DORA compliance:
1. **DORA Financial Services Addendum (DORA FSA)**: contractual terms addressing DORA Article 30 requirements
2. **Level 1 DORA Workbook**: maps DORA requirements to AWS Shared Responsibility Model
3. **Level 2 Guide**: detailed deployment guidance

Recommended AWS services for DORA: AWS Audit Manager (evidence collection and framework mapping), Security Hub (continuous security monitoring), AWS Resilience Hub (resilience testing and assessment), Trusted Advisor (operational best practices), AWS Health (service event notifications), and AWS Incident Detection and Response (managed incident handling).

What Auditors Actually Ask For

Technical controls get the most attention during preparation. But auditors spend equal or more time on evidence that no AWS service generates automatically. They want documented policies and procedures -- not just that CloudTrail is enabled, but that a written logging policy defines what gets logged, who reviews it, and how long it is retained. They want training records proving staff completed security awareness training. They want risk assessments performed at defined intervals, not just a Security Hub score. They want incident response test documentation showing you ran a tabletop exercise or simulation, not just that you wrote a playbook. They want change management logs tying deployments to approved change tickets. Security Hub compliance scores impress dashboards, not auditors. An auditor will ask "show me your last three change requests and the approval chain" before they ask about your FSBP score.

AWS Audit Manager

AWS Audit Manager automates evidence collection and framework assessment for continuous compliance. A **July 2025** update enhanced evidence collection across 14 standard frameworks.

Key prebuilt frameworks include:

Framework	Version
PCI DSS	v4.0 (v4.0.1 not yet available)
SOC 2	Trust Services Criteria
NIST SP 800-53	Rev. 5
HIPAA	Security Rule
CIS AWS Foundations Benchmark	Multiple versions
FedRAMP	Moderate (r4)
GDPR	Data Protection
GxP	EU Annex 11
ISO/IEC 27001	2013 Annex A

Framework	Version
NIST Cybersecurity Framework	v1.1
NIST SP 800-171	Rev. 2
GLBA	Financial Privacy
ACSC Essential Eight	Australian Cyber Security Centre

Audit Manager automates evidence from **CloudTrail** (API activity logs), **AWS Config** (configuration compliance snapshots), and **Security Hub** (security control findings). It also supports custom frameworks for organization-specific control sets.

Important limitation: The ISO/IEC 27001 framework in Audit Manager covers the **2013 Annex A** only. If your certification scope is the 2022 revision, you must manually map the consolidated 93-control Annex A structure to the automated evidence collected against the 2013 framework.

> **What Goes Wrong in Practice**
>
> Four audit failures come up repeatedly in AWS environments. First, insufficient change management evidence: CloudTrail proves what changed but not why. Auditors want a change ticket linked to each deployment, with an approval chain and rollback plan. Without a correlation between CloudTrail events and change management records, you fail the change management control. Second, missing data flow diagrams: auditors ask where data enters, moves through, and leaves your environment. Most teams have architecture diagrams but not data flow diagrams -- they are not the same thing. Third, no evidence of periodic access reviews: IAM Access Analyzer flags unused access, but auditors want proof that someone reviewed access quarterly (or whatever your policy states), made decisions, and documented the outcome. An Access Analyzer report sitting in an S3 bucket is not a review. Fourth, the gap between Security Hub "compliant" and actual audit readiness: Security Hub checks technical configurations, but compliance requires policies, procedures, training, testing, and documented evidence of all four. A 95% Security Hub score with no written incident response procedure is still a finding.

AWS Artifact and Control Tower

AWS Artifact

AWS Artifact is the self-service portal for downloading AWS compliance reports: SOC 1/2/3, ISO certificates, PCI DSS AOC, FedRAMP packages, and C5 attestations.

Since **December 2025**, Artifact provides direct access to **previous report versions** -- no more contacting AWS Support. This requires the `artifact:ListReportVersions` IAM permission.

AWS Control Tower Compliance Frameworks

AWS Control Tower added compliance framework support throughout 2025:

- **2025**: Multiple compliance frameworks added to the Control Catalog, including CIS, FedRAMP, ISO/IEC 27001, NIST CSF, NIST SP 800-171, PCI DSS, and SOC 2
- **November 2025**: Additional managed AWS Config rules added
- **December 2025**: 176 Security Hub controls added to the Control Catalog

These additions let Control Tower enforce compliance-aligned controls across multi-account environments from a single governance pane.

The Compliance Landscape at a Glance

Framework	Scope	Key AWS Tooling	Update Frequency
SOC 2 Type II	Trust Services Criteria	Audit Manager, Config, CloudTrail	Semi-annual reports
ISO 27001:2022	ISMS	Audit Manager (2013 Annex A), Config	3-year certification cycle
CIS v5.0.0	Baseline security configuration	Security Hub CSPM	Updated per benchmark release
GDPR	EU personal data	Macie, IAM Access Analyzer, S3 Object Lock	Ongoing obligation
PCI DSS v4.0.1	Cardholder data	Security Hub CSPM, Config	Annual assessment
HIPAA	Protected health information	BAA, eligible services, CloudTrail	Ongoing obligation
NIS2	Essential/important entities	Security Hub, GuardDuty, Config	National transposition varies
DORA	Financial entities + ICT providers	Audit Manager, Resilience Hub, Security Hub	General ICT testing annually (Art. 24); TLPT every 3 years (Art. 26)

How Toc Consulting Conducts Security Audits

Compliance frameworks tell you what to assess. This section describes how Toc Consulting assesses it: automated checks combined with manual review and evidence collection, producing findings you can act on -- not just compliance reports.

Phase 1: Automated Baseline Assessment

Tool	What It Checks
Security Hub CSPM	FSBP, CIS v5.0.0, PCI DSS v4.0.1, NIST SP 800-53: continuous compliance scoring
AWS Config	Resource configuration compliance against managed and custom rules
Audit Manager	Evidence collection against target framework (SOC 2, ISO 27001, PCI DSS, etc.)
IAM Access Analyzer	External and unused access findings across accounts
Trusted Advisor	Security, cost, performance, and fault tolerance checks

Automated checks provide breadth: hundreds of controls across all accounts in minutes. They cannot assess architecture decisions, business logic, or operational procedures.

Phase 2: Manual Review

Manual review covers what automation cannot:

- **Architecture review**: Network segmentation, blast radius design, data flow mapping, encryption in transit and at rest
- **IAM deep dive**: Policy analysis beyond Access Analyzer, including custom policies, permissions boundaries, cross-account trust relationships, service-linked roles
- **Operational procedures**: Incident response playbooks, change management processes, backup and recovery testing, on-call rotations
- **Business context**: Data classification accuracy, regulatory scope determination, third-party risk assessment

Phase 3: Evidence Collection and Reporting

Every finding is documented with:

1. **Control reference**: mapped to the applicable framework control (e.g., CIS 2.1.1, PCI DSS 8.4.2, ISO 27001 A.8.24)
2. **Current state**: what the automated check or manual review found

3. **Risk rating**: Critical, High, Medium, Low based on exploitability and business impact
4. **Remediation guidance**: specific AWS CLI commands, IAM policy changes, or architecture modifications
5. **Evidence**: screenshots, Config snapshots, CloudTrail events, Security Hub findings

The deliverable is a prioritized remediation roadmap, not a compliance certificate. Certificates come from accredited auditors. Toc Consulting prepares you to pass their assessment with no surprises.

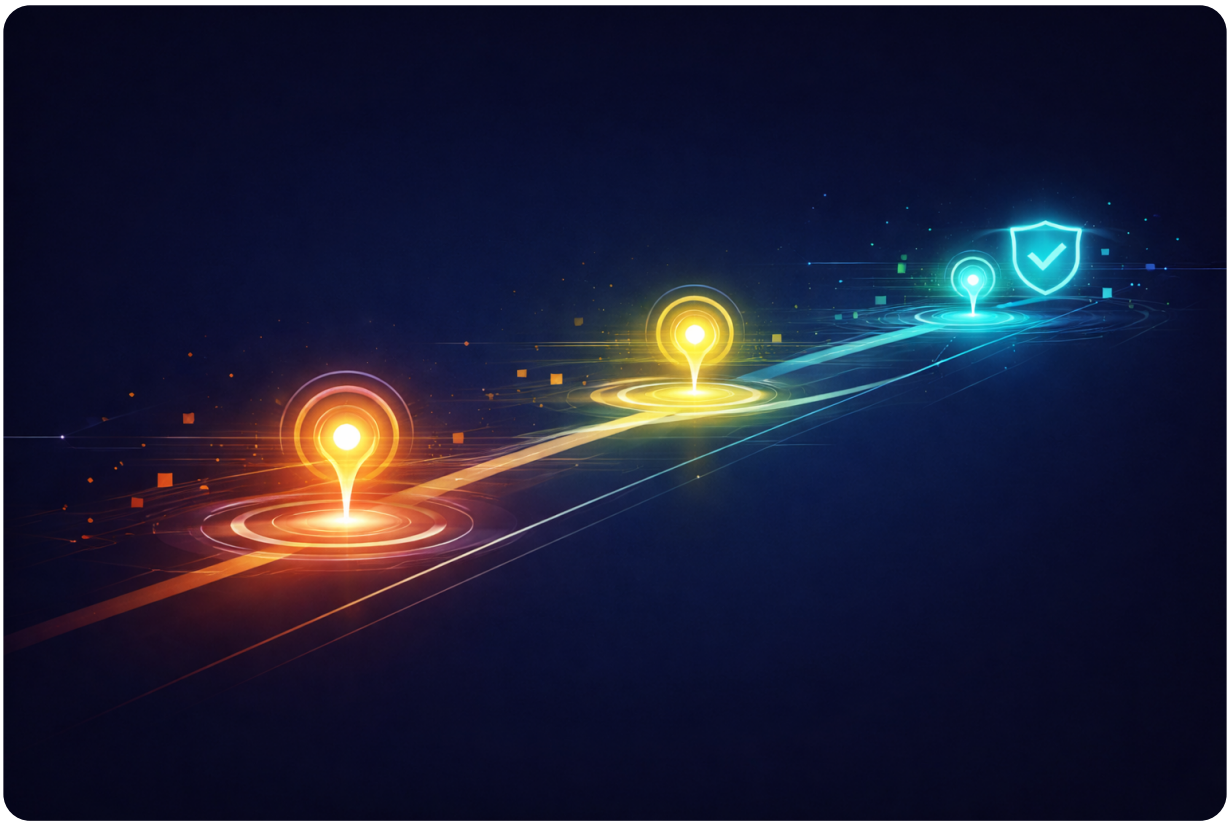
Audit Checklist

Use this checklist as a starting point for compliance readiness:

- [ ] **CloudTrail**: Organization trail enabled, multi-region, log file integrity validation active, delivered to a centralized S3 bucket with restricted access
- [ ] **AWS Config**: Enabled in all regions and accounts, recording all resource types, conformance packs deployed for target framework
- [ ] **Security Hub**: Enabled with FSBP and applicable compliance standard (CIS v5.0.0, PCI DSS v4.0.1), cross-account aggregation configured
- [ ] **GuardDuty**: Enabled in all accounts and regions, delegated administrator configured, Runtime Monitoring evaluated for sensitive workloads
- [ ] **IAM**: Root user MFA enabled (hardware key), no root access keys, IAM Access Analyzer enabled in all accounts, password policy meets framework requirements (12 characters minimum for PCI DSS v4.0.1)
- [ ] **Encryption**: Default EBS encryption enabled in all regions, S3 bucket default encryption (SSE-S3 or SSE-KMS), RDS encryption at rest, TLS 1.2+ enforced for data in transit
- [ ] **Audit Manager**: Assessment created for target framework, evidence collection validated, delegated administrator configured for multi-account
- [ ] **Artifact**: Current compliance reports downloaded and reviewed (SOC 2, ISO, PCI DSS AOC as applicable)
- [ ] **Network**: VPC Flow Logs enabled, security groups restrict ingress to required ports only, public access blocked on S3 buckets (S3 Block Public Access at account level)
- [ ] **Incident Response**: Documented incident response plan, tested at least annually, Security Hub integrated with notification channels (SNS, EventBridge)
- [ ] **Data Classification**: Macie enabled for PII/PHI discovery, S3 inventory complete, data flow diagrams current
- [ ] **Backup**: AWS Backup plans covering key resources, cross-region backup for disaster recovery, restore testing documented

---

Chapter 8: Your AWS Security Action Plan



From Knowledge to Action

The previous seven chapters covered threats, identity, data protection, network security, monitoring, agentic AI security, and compliance. This chapter distills everything into an action plan: what to do first, what to do next, and how to measure progress.

Security has no finish line. The plan below is phased: immediate quick wins that cut the most risk with the least effort, then systematic hardening, then ongoing maturity improvement.

The 30-Day Quick Wins

These actions close the highest-risk gaps with minimal effort and cost. Most take a single session. Several use AWS Free Tier or no-cost features.

Week 1: Identity and Access

#	Action	Why It Matters	Cost
1	Enable MFA on all root users (hardware key preferred)	Root access in the management account bypasses all IAM controls. In member accounts, SCPs can restrict root, but a compromised root credential is still the highest-privilege identity in that account. Consider centralized root access management via AWS IAM and Organizations (launched November 2024), which can disable root credentials entirely in member accounts.	Basic FIDO2 security key: ~$25–$30
2	Delete root access keys in every account	Root access keys are the single most dangerous credential in AWS.	Free
3	Enable IAM Access Analyzer in every account (external access)	Finds resources shared with external principals: S3 buckets, IAM roles, KMS keys, Lambda functions, SQS queues, and many more resource types.	No additional charge for external access findings
4	Enforce MFA for all human IAM users	Credential theft is the single most frequent initial access vector at 22% of breaches (Chapter 1).	Free (software MFA)
5	Review and remove unused IAM users and roles	Dormant credentials are attacker favorites; they do not trigger alerting on login pattern changes.	Free

Week 2: Logging and Detection

#	Action	Why It Matters	Cost
6	Verify CloudTrail organization trail is enabled, multi-region, with log file integrity validation	CloudTrail underpins all detection. One free copy of management events per region.	Free (management events)
7	Enable GuardDuty in all accounts and regions	Behavioral threat detection across CloudTrail, VPC flow data, and DNS logs.	30-day free trial

#	Action	Why It Matters	Cost
8	Turn on Security Hub CSPM with FSBP standard	Security Hub CSPM runs hundreds of FSBP security checks. (As of December 2025, AWS split Security Hub into two services: the unified Security Hub for findings aggregation, and Security Hub CSPM for compliance standards and security checks. References to compliance controls in this chapter refer to Security Hub CSPM.) Security Hub also aggregates findings from GuardDuty, Inspector, and Macie.	30-day free trial; Security Hub CSPM includes a 30-day free trial; pricing is based on security checks, finding ingestion events, and rule evaluations
9	Enable AWS Config in all regions	Records resource configuration changes. Required by Security Hub and most compliance frameworks.	Per configuration item recorded
10	Set up a notification channel for Critical/High findings	Security Hub → EventBridge → SNS (email or Slack). Detection without notification is detection without response.	Free (SNS first 1M requests)

Week 3: Data Protection

#	Action	Why It Matters	Cost
11	Enable S3 Block Public Access at the account level	Prevents accidental public exposure of any S3 bucket in the account.	Free
12	Turn on default EBS encryption in all regions	All new EBS volumes are encrypted automatically. Does not affect existing unencrypted volumes.	Free (encryption with AWS-managed aws/ebs KMS key)
13	Turn on default S3 bucket encryption (SSE-S3 or SSE-KMS)	All new objects encrypted at rest by default. SSE-S3 is enabled by default since January 2023.	Free (SSE-S3)
14	Run Amazon Macie on sensitive S3 buckets	Discovers PII, PHI, and financial data. First 1 GB of automated data discovery per account per month is free.	Free tier: 1 GB/month
15	Review S3 bucket policies for wildcard principals	`"Principal": "*"` in a bucket policy means public access unless constrained by conditions.	Free

Week 4: Network Baseline

#	Action	Why It Matters	Cost
16	Audit security groups for unrestricted inbound rules	0.0.0.0/0 on SSH (22), RDP (3389), or database ports is the most common network misconfiguration.	Free
17	Enable VPC Flow Logs on all VPCs	Network visibility is required for incident investigation and traffic analysis.	Per GB published to CloudWatch Logs or S3
18	Remove internet gateways from VPCs that do not require internet access	Reduces attack surface. Private workloads should use VPC endpoints for AWS service access.	Free
19	Verify that no EC2 instances have public IPs unless required	Use Elastic Load Balancers or CloudFront for public-facing workloads.	Free
20	Enable AWS WAF on all public-facing ALBs, API Gateways, and CloudFront distributions	Protects against OWASP Top 10 web application attacks.	Per ACL, per rule, per request

The 90-Day Security Hardening Roadmap

After the quick wins, the next 60 days focus on systematic hardening -- closing structural gaps that require planning and cross-team coordination.

Month 2: Architecture and Controls

#	Action	Chapter Reference
21	Deploy SCPs in AWS Organizations to enforce guardrails (deny root actions, restrict regions, prevent disabling of logging)	Chapter 2
22	Deploy Resource Control Policies (RCPs) on sensitive resources (S3 buckets, KMS keys)	Chapter 2
23	Configure IAM Identity Center for all human access; eliminate long-lived IAM user credentials	Chapter 2

#	Action	Chapter Reference
24	Deploy VPC endpoints for key AWS services (S3, DynamoDB, KMS, STS, CloudTrail) to eliminate internet-routable API traffic	Chapter 4
25	Segment networks: separate production, staging, and development into distinct VPCs or accounts	Chapter 4
26	Enable Security Hub cross-account aggregation in the delegated administrator account	Chapter 5
27	Enable Security Hub CSPM compliance standards aligned to your target framework (CIS v5.0.0, PCI DSS v4.0.1) and deploy Config conformance packs for additional coverage. Note: Security Hub CSPM and Config conformance packs may support different framework versions. Evaluate overlap to avoid duplicate Config rule costs.	Chapter 7
28	Enable CloudTrail data events for S3 buckets containing sensitive data	Chapter 5
29	Enable GuardDuty Runtime Monitoring for EC2/EKS workloads processing sensitive data	Chapter 5
30	Turn on KMS key rotation and review key policies for least-privilege access	Chapter 3

Month 3: Operations and Response

#	Action	Chapter Reference
31	Create and test incident response runbooks for the top 5 GuardDuty finding types in your environment	Chapter 5
32	Deploy automated remediation for high-confidence findings (e.g., auto-quarantine compromised EC2 instances, auto-revoke exposed S3 buckets)	Chapter 5
33	Enable AWS Backup with cross-region replication for key workloads	Chapter 3
34	Conduct a Well-Architected Security Pillar review	All chapters
35	Configure Bedrock Guardrails if deploying generative AI workloads	Chapter 6
36	Set up AWS Audit Manager with your target compliance framework	Chapter 7
37	Configure CloudTrail Lake for long-term log retention and SQL-based investigation	Chapter 5

#	Action	Chapter Reference
38	Review and harden Lambda execution roles; each function should have its own role with minimal permissions	IAM best practice
39	Enable Amazon Inspector for continuous vulnerability scanning across EC2, Lambda, and ECR container images	Chapter 5
40	Document data flow diagrams and data classification for regulated workloads	Chapter 7

Cost Reality Check: The full security stack described in the 90-day roadmap (GuardDuty, Security Hub, Config, CloudTrail data events, Inspector) typically costs between 2-5% of total AWS spend for a well-architected environment. The largest variable cost drivers are CloudTrail data events (S3 and Lambda), Config rule evaluations, and Security Hub member account charges. Start with GuardDuty and Security Hub (lowest cost, highest signal) and add services incrementally based on risk assessment. Do not let cost uncertainty delay deployment of the baseline -- the cost of a security incident dwarfs the cost of detection services.

The Security Maturity Model

Security maturity is not binary. The model below defines four levels, each building on the previous one. Skip a level and you get the illusion of security without the substance.

Level 1: Baseline

You have it if: CloudTrail is enabled, MFA is on root, S3 Block Public Access is active, GuardDuty is running, Security Hub CSPM is enabled with FSBP.

What it means: You have visibility and basic preventive controls. You detect common threats and receive alerts but are not yet responding systematically.

Typical gaps: No incident response plan, no automated remediation, no centralized log management, IAM policies are overly permissive, no compliance framework in use.

Level 2: Structured

You have it if: SCPs enforce guardrails across accounts, IAM Identity Center replaces long-lived credentials, network segmentation isolates workloads, AWS Config enforces compliance baselines, incident response runbooks exist and have been tested at least once.

What it means: You have moved from detection to prevention. Security controls are systematic rather than ad hoc. You can demonstrate compliance readiness to auditors.

Typical gaps: Remediation is still manual, no threat hunting, limited data classification, no formal security review process for new workloads.

Level 3: Operationalized

You have it if: Automated remediation handles high-confidence findings, CloudTrail Lake retains logs for investigation, Security Lake centralizes logs from AWS and third-party sources, Amazon Detective provides deep investigation capabilities, Audit Manager collects evidence continuously, data classification is complete, security reviews are part of the deployment pipeline, incident response has been tested through tabletop exercises.

What it means: Security operates as a continuous process, not a periodic assessment. You can investigate incidents end-to-end. Compliance evidence is generated automatically.

Typical gaps: No proactive threat hunting, limited AI/agentic security controls, no cross-team security metrics, no formal security maturity reporting to leadership.

Level 4: Adaptive

You have it if: Agentic AI workloads are secured with Bedrock Guardrails and AgentCore Policy Controls (preview as of February 2026), threat hunting is a regular activity using Security Lake and Detective, security metrics are reported to leadership, the organization adapts controls based on emerging threats and intelligence.

What it means: Security is a competitive advantage. You detect and respond faster than attackers can pivot. Your controls adapt to new threat categories (including agentic AI risks) before they become incidents.

Appendix: Consolidated Security Checklist (50 Controls)

This checklist consolidates the controls from each chapter for quick reference. Print this page and use it as a periodic audit tool. For implementation details, refer to the chapter indicated for each control group.

Identity and Access (Chapter 2)

- [ ] 1. Root user MFA enabled (hardware key) in every account
- [ ] 2. Root access keys deleted in every account
- [ ] 3. IAM Identity Center configured for all human access
- [ ] 4. MFA enforced for all human users
- [ ] 5. SCPs deployed to restrict root actions, enforce region limits, and prevent logging disablement
- [ ] 6. RCPs deployed on sensitive resources

- [ ] 7. IAM Access Analyzer enabled (external access) in every account
- [ ] 8. Unused IAM users, roles, and policies removed
- [ ] 9. Lambda functions use per-function execution roles
- [ ] 10. Cross-account trust relationships reviewed and documented

Data Protection (Chapter 3)

- [ ] 11. S3 Block Public Access enabled at account level
- [ ] 12. Default EBS encryption enabled in all regions
- [ ] 13. S3 default encryption active (SSE-S3 or SSE-KMS)
- [ ] 14. KMS key automatic rotation enabled (configurable 90-2,560 days) with least-privilege key policies
- [ ] 15. RDS encryption at rest enabled
- [ ] 16. TLS 1.2+ enforced for data in transit
- [ ] 17. S3 bucket policies audited for wildcard principals
- [ ] 18. Macie enabled for PII/PHI discovery on sensitive buckets
- [ ] 19. S3 Object Lock configured for regulatory retention requirements
- [ ] 20. AWS Backup plans covering key resources with cross-region replication

Network Security (Chapter 4)

- [ ] 21. Security groups restrict inbound to required ports and sources only
- [ ] 22. No unrestricted SSH (22) or RDP (3389) inbound rules
- [ ] 23. VPC Flow Logs enabled on all VPCs
- [ ] 24. VPC endpoints deployed for high-traffic AWS services
- [ ] 25. Internet gateways removed from private VPCs
- [ ] 26. EC2 instances do not have public IPs unless required
- [ ] 27. AWS WAF enabled on public-facing ALBs, API Gateways, and CloudFront
- [ ] 28. Network ACLs provide defense-in-depth for sensitive subnets
- [ ] 29. Network Firewall or third-party IDS/IPS deployed for east-west traffic inspection
- [ ] 30. Private subnets use NAT gateways (not internet gateways) for outbound access

Monitoring and Detection (Chapter 5)

- [ ] 31. CloudTrail organization trail enabled, multi-region, log file integrity validation active
- [ ] 32. CloudTrail data events enabled for sensitive S3 buckets
- [ ] 33. GuardDuty enabled in all accounts and regions with delegated administrator
- [ ] 34. GuardDuty Runtime Monitoring evaluated for sensitive workloads
- [ ] 35. Security Hub CSPM enabled with FSBP and applicable compliance standard

- [ ] 36. Security Hub cross-account aggregation configured
- [ ] 37. AWS Config enabled in all regions, recording all resource types
- [ ] 38. Security Hub CSPM compliance standards enabled for target framework; Config conformance packs deployed for additional coverage
- [ ] 39. Amazon Inspector enabled for continuous vulnerability scanning across EC2, Lambda, and ECR container images
- [ ] 40. Notification pipeline configured (Security Hub → EventBridge → SNS/Slack)

Agentic AI Security (Chapter 6)

- [ ] 41. Bedrock Guardrails configured for all GenAI applications
- [ ] 42. Bedrock Model Invocation Logging enabled
- [ ] 43. Agent execution roles follow least-privilege (no shared roles across agents)
- [ ] 44. MCP tool sources validated and monitored for changes
- [ ] 45. AgentCore Policy Controls (preview as of February 2026) evaluated for production agent deployments

Compliance and Audit (Chapter 7)

- [ ] 46. Audit Manager assessment created for target compliance framework
- [ ] 47. AWS Artifact reports reviewed (SOC 2, ISO, PCI DSS AOC as applicable)
- [ ] 48. Incident response plan documented and tested at least annually
- [ ] 49. Data classification complete and data flow diagrams current
- [ ] 50. Control Tower compliance frameworks enabled for multi-account governance

---

When to Bring in Expert Help

Turning on CloudTrail, GuardDuty, and Security Hub is the easy part. Configuring them correctly, interpreting the findings, and building a security posture that holds up under real attack pressure requires hands-on expertise:

- **Initial architecture design.** Getting multi-account structure, network segmentation, and IAM design right from the start costs far less than retrofitting after production workloads are running.
- **Compliance gap assessment.** Mapping the delta between your current posture and a target framework (SOC 2, PCI DSS, ISO 27001, NIS2, DORA) requires experience with both the framework and AWS-specific implementation.
- **Incident response readiness.** Building runbooks, testing them through tabletop exercises, and establishing a response chain of command requires hands-on experience with real AWS incidents.

- **Agentic AI security.** The threat surface for AI agents (prompt injection, tool poisoning, confused deputy, semantic privilege escalation) requires specialized knowledge that most organizations have not yet built internally.
- **Post-breach investigation.** The first 48 hours determine the outcome. An experienced team that knows CloudTrail, GuardDuty, Detective, and forensic procedures is not something you build during an active incident.

About Toc Consulting

Toc Consulting is an **AWS Security and Cloud Architecture** consultancy. We help organizations secure their cloud infrastructure so they can focus on building.

What we do:

- **AWS Security Audits.** Full-scope assessments combining automated checks (Security Hub, Config, IAM Access Analyzer, Audit Manager) with deep manual review of IAM policies, network architecture, encryption configuration, and operational procedures.
- **Architecture Reviews.** Multi-account strategy, network design, encryption architecture, and monitoring stack design, aligned with the AWS Well-Architected Security Pillar.
- **Compliance Readiness.** Preparing your AWS environment for SOC 2, ISO 27001, PCI DSS, HIPAA, NIS2, and DORA assessments. We deliver the evidence, gap analysis, and remediation roadmap. Your accredited auditor delivers the certificate.
- **Incident Response.** Investigation, containment, and recovery. CloudTrail analysis, forensic evidence preservation, and root cause determination.
- **Agentic AI Security.** Securing AI agent deployments on AWS: Bedrock Guardrails, AgentCore policy controls, model invocation logging, and MCP tool validation.

KloudSec - Cloud Security Platform. We are building KloudSec, an AI-powered cloud security platform that automates 300+ security checks across 100 AWS services, generates compliance reports (CIS, SOC 2, ISO 27001, PCI DSS, NIST), and delivers copy-paste remediation guides. Full scan in under 5 minutes. Launching March 2026.

Book a free AWS security assessment. We will review your AWS environment, identify the top 10 highest-risk findings, and deliver a prioritized remediation roadmap, at no cost and no commitment.

Contact: tocconsulting.fr | kloudsec.io | @TocConsulting

This concludes the Toc Consulting AWS Security Whitepaper, 2026 Edition. Start with the 30-day quick wins, build toward the 90-day roadmap, and measure progress against the maturity model. If you need help, we are here.