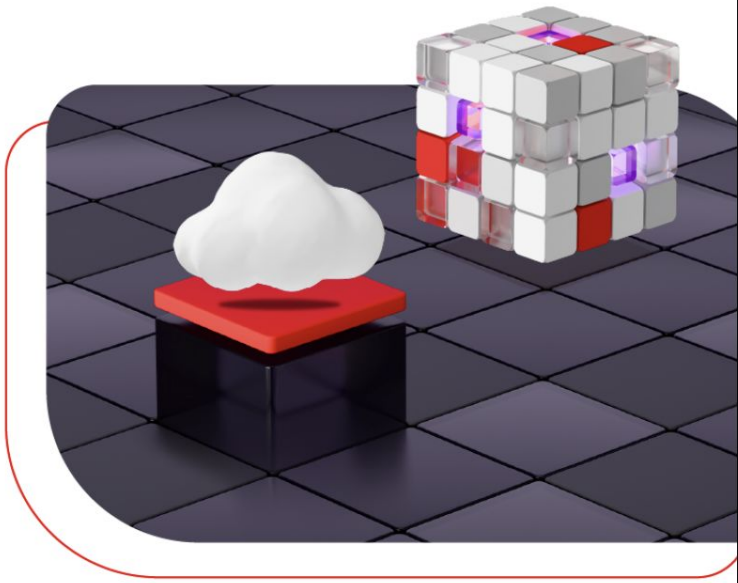
Red Hat

Security for an AI-Driven World



Building a Resilient Foundation for the AI-Ready Enterprise

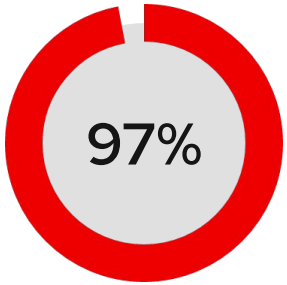
(**The Security Symposium**, Redmond, WA, August 2026)

Paul Turnbull
Solution Architect Lead/Manager
Red Hat (Canada)

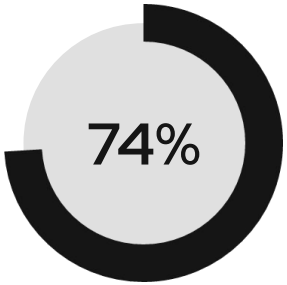
MOBIA CONNECT

STATE OF CLOUD-NATIVE SECURITY

Security incidents are common and costly



experienced at least one security incident



Delayed deployments due to security issues

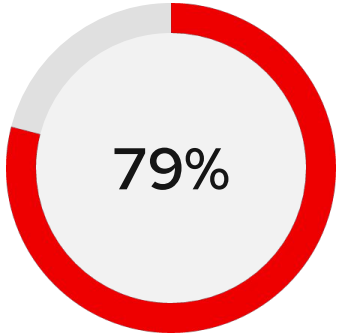
MOST COMMON INCIDENTS:

1. Misconfigured infrastructure
2. Known vulnerabilities detected
3. Unauthorized access
4. Use of unverified images
5. Failed security audit

Source: Red Hat's State of Cloud Native Security 2026 Report

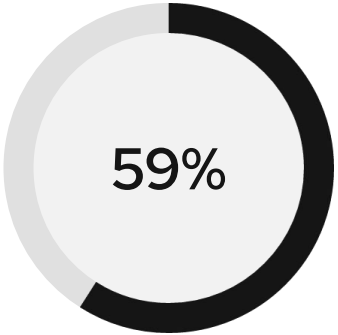
Red Hat

AI-related security risks are adding additional pressure



agree Gen AI creates new security challenges

Governance is struggling. Many organizations don't know if sensitive data is being exposed to insecure AI tools.



have no AI governance framework

Formal policy and documentation around internal AI use is not keeping pace with the rapid adoption of agentic tools.

Source: Red Hat's State of Cloud Native Security 2026 Report

The agentic mandate: autonomy = systemic risk

Moving from simple GenAI (chatbots) to **autonomous agents** that execute high-impact, multi-step tasks across the enterprise.

THE SHIFT



Autonomous agents

Agents operate with limited supervision, autonomously interacting with critical systems including internal tools affecting **Finance, Inventory, HR, Manufacturing .**

THE SYSTEMIC PAIN



Systemic risk

A compromised agent executes unauthorized transactions, cross region, shuts down production lines, or alters supply chain, **instantaneously and world-wide**

OUR CHALLENGE



Embedded security

Security must be implemented in all aspects and artifacts part of an AI system, including model, agents, third-party tools, **rather than just the network perimeter** considering sovereignty and policies.

AI discovers vulnerabilities at **machine speed**

Discovery is exponential. Remediation is linear.

IT and security teams are under pressure:

- AI can discover thousands of vulnerabilities in days—faster than humans.
- Exploit development time has compressed dramatically from public CVE to working exploit in hours not weeks.
- The bottleneck is no longer finding—it's controlling an exposure before a fix exists.

45%

of vulnerabilities in large organizations remain unpatched after 12 months.

Source: Centre for Emerging Technology and Security (Alan Turing Institute), 2025 – cited in CETAS analysis, April 2026.

Frontier AI models

Frontier models (like Mythos) that can identify and autonomously build exploits to 30-year-old zero-days in hours. When previously these vulnerabilities evaded human detection for decades.

Project Glasswing

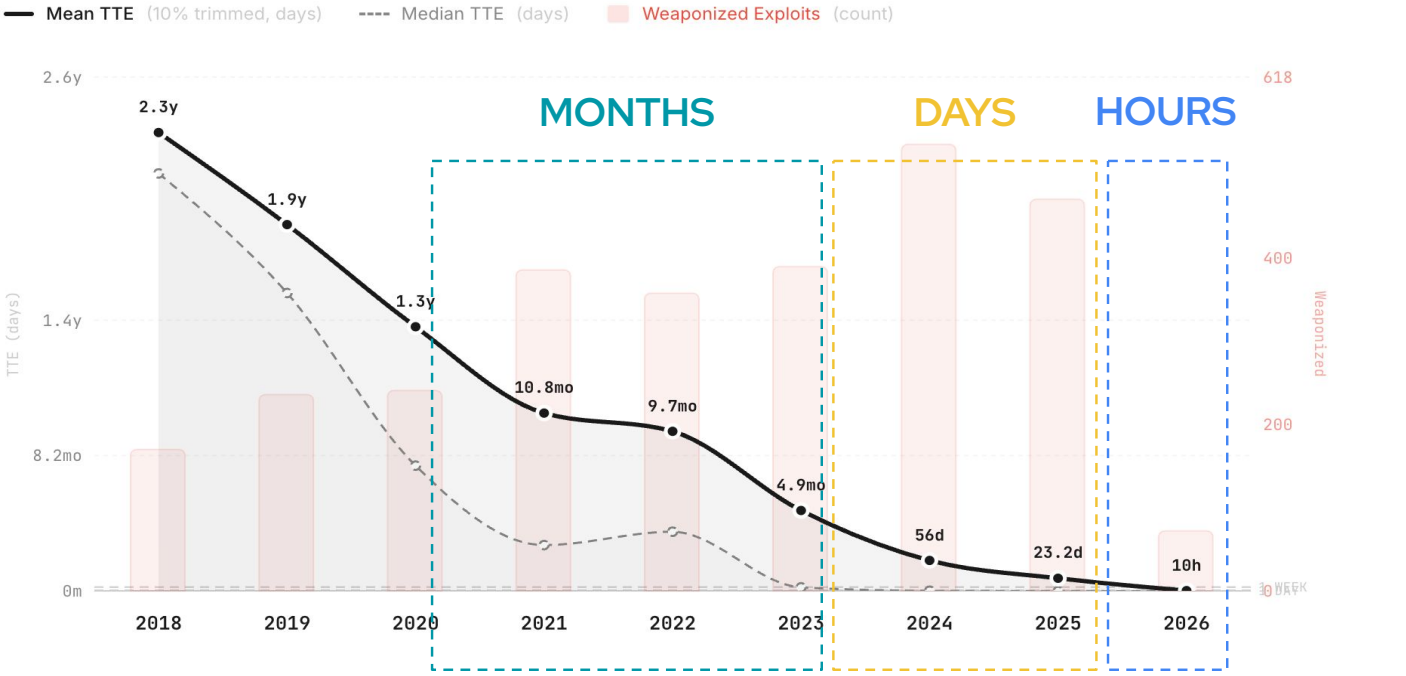
While Glasswing accelerates discovery for Anthropic built models, its limitations mean it can't cover every model or threat surface.

Red Hat

AI is accelerating the security vulnerability landscape

From Vulnerability to Exploitation

TTE (Time-to-Exploit) measures the gap between CVE disclosure and confirmed exploitation

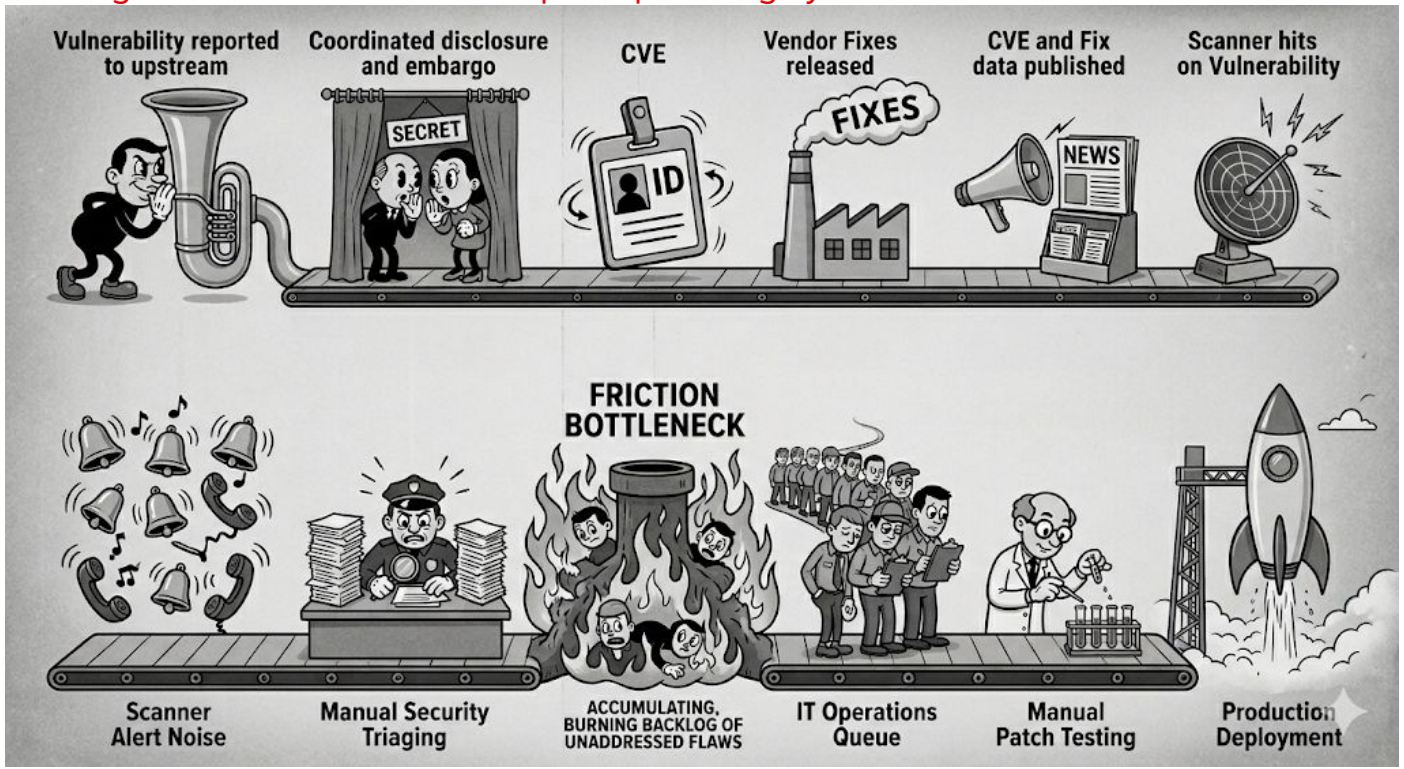


Based on 3,532 CVE-exploit pairs from trusted sources (CISA KEV, VulnCheck KEV & XDB)

zerodayclock.com

The Status Quo

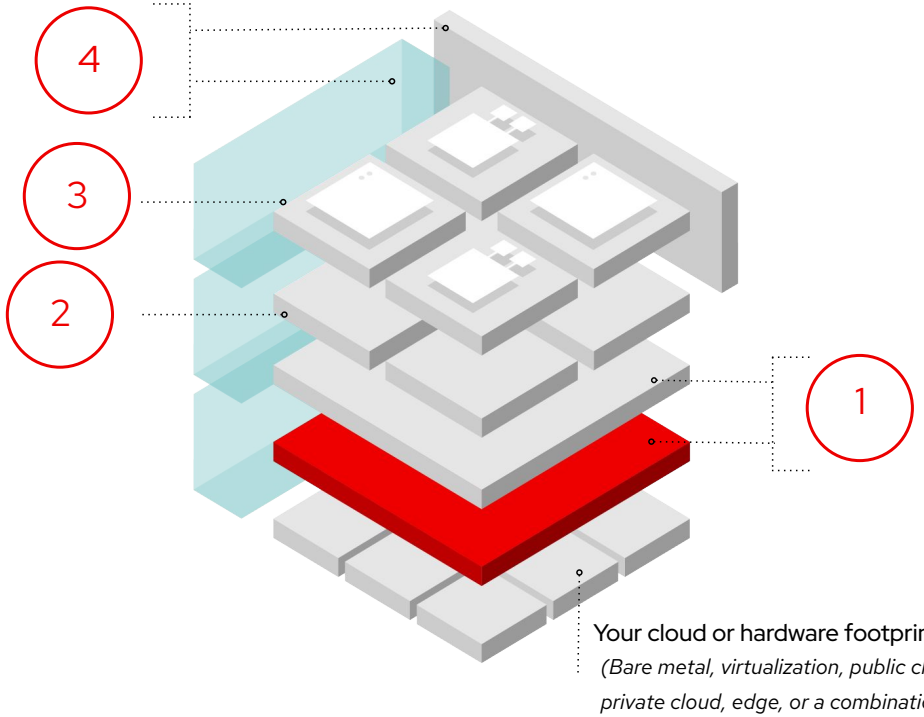
Shifting from a reactive human-speed patching cycle in an AI-accelerated threat landscape



Red Hat

The AI-Ready Architecture is Layered

Security from the OS to Autonomous AI Agents

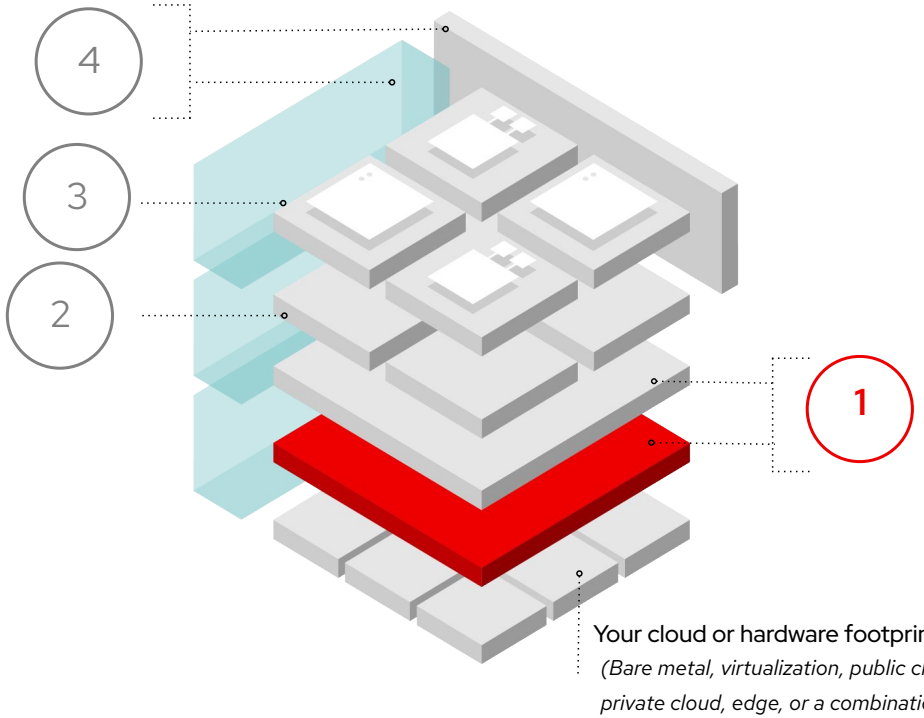


1. Start with a strong foundation
2. Protect the pipeline
3. Implement guardrails for AI agents and models
4. Automate security at scale

Red Hat

Layer 1: Harden the Foundation

Start with a strong foundation



1. **Start with a strong foundation**
2. Protect the pipeline
3. Implement guardrails for AI agents and models
4. Automate security at scale

Red Hat

Hardened Foundation

THE CHALLENGE

Problem: Resource Sharing Risks

AI needs massive compute (GPUs), but sharing resources creates security gaps and performance exposure.

THE ANSWER

Solution: Secure Isolation

We use RHEL and OpenShift to provide a zero-trust, hardened base for the AI workload.

PRODUCT HIGHLIGHTS

- **RHEL & OpenShift:** Decades of Linux and container security best practices applied to AI
- **Hardware vendors Integration:** Offload network and security tasks, physically isolating critical AI workloads and support AI performance.

"You cannot secure an AI system if you cannot secure the metal it runs on. Isolation is the first defense."

Red Hat

Confidential Computing

Protect user data and models from the host

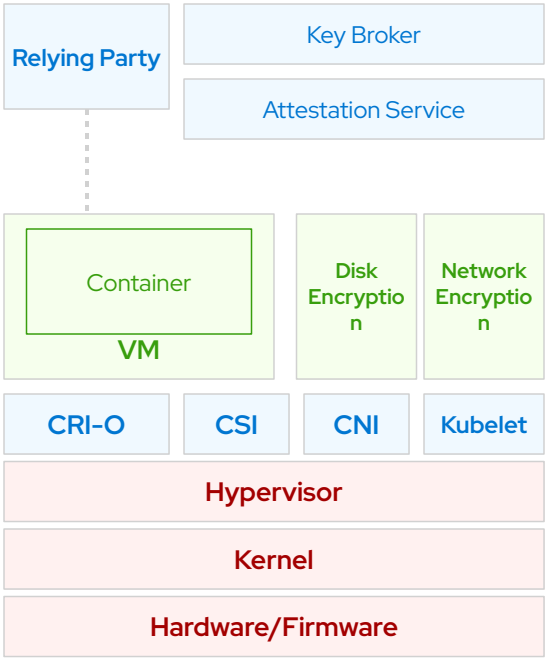
Data Owner/Tenant
Confidential area, inaccessible to host even when running on it, includes attestation services

Host
Manages and offers physical resources but is not trusted with data that resides on it

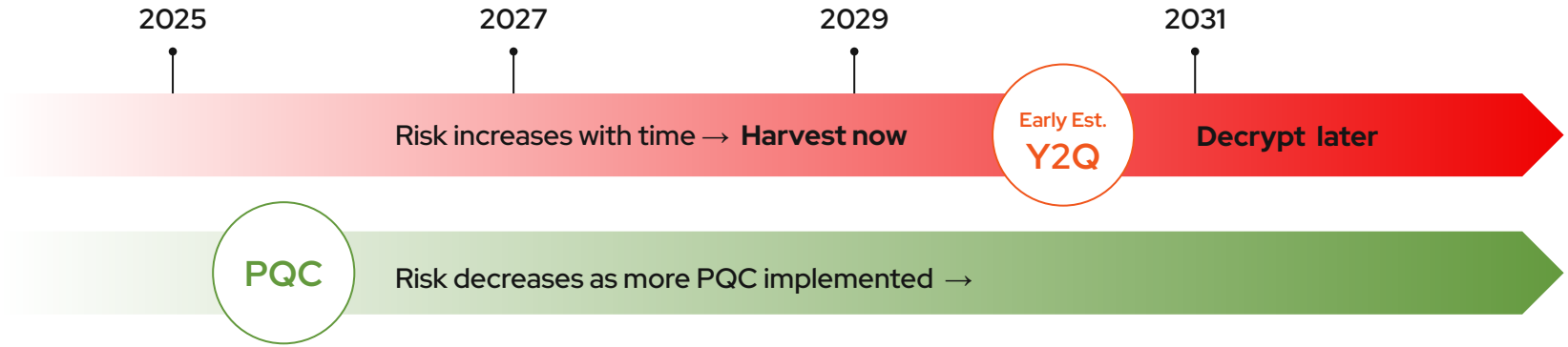
Trusted Platform
Offers confidential guarantees using hardware-level cryptographic enforcement

Technology

- Confidential Containers (CoCo)
- Built on hardware-level cryptography: Intel SGX/TDX, AMD SEV, IBM Z Secure Execution, **NVIDIA GPU TEE**



Post-quantum cryptography is a future challenge, but the threat is already here



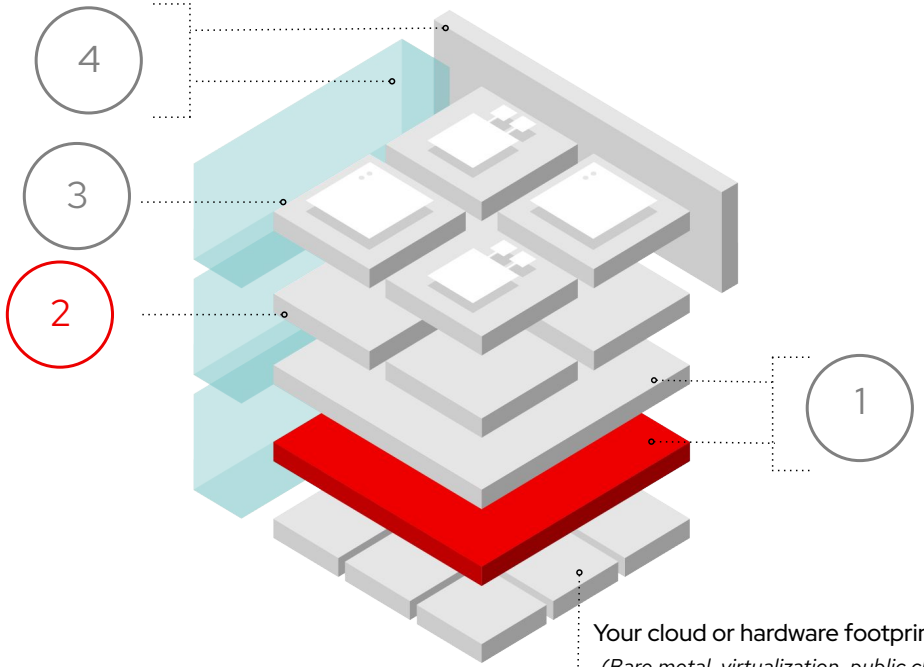
Red Hat Enterprise Linux 10 includes NIST-approved quantum-resistant algorithms[1] that enable key-exchange, encryption, and signing.

Y2Q Year when quantum starts defeating current encryption | **PQC** Start of PQC adoption

1 FIPS 203 ML-KEM
FIPS 204 ML-DSA
FIPS 205 SLH-DSA

Layer 2: Protect the Pipeline

Knowing exactly what is in your software

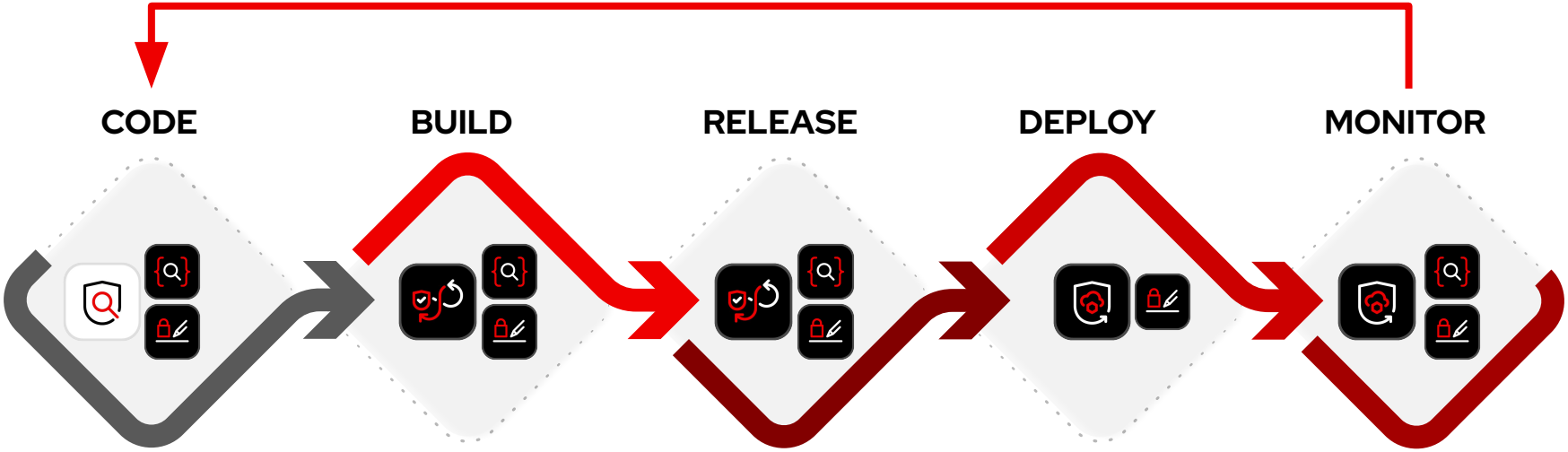


1. Start with a strong foundation
2. **Protect the pipeline**
3. Implement guardrails for AI agents and models
4. Automate security at scale

Red Hat

Software Supply Chain Security

From Code Commit to Runtime Enforcement



Identify dependencies and recommend trusted content before coding

Provide guidance, templates, tools to catch known vulnerabilities and misconfigurations improving dev productivity, reducing cost to fix, securing the pipeline

Vulnerability Scanning

Policy Enforcement

Artifact Signing

Automate guardrails to prevent deployment of problematic code and deliver risk mitigation

Detect and respond to runtime threats with prioritized risk assessments

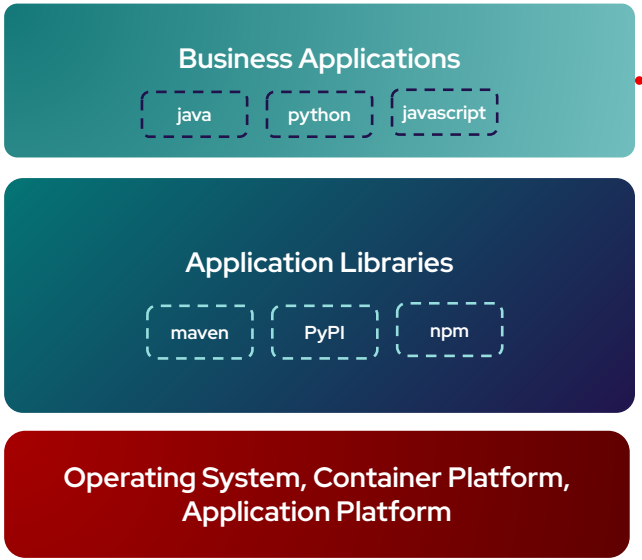
RH Trusted Profile Analyzer

RH Trusted Artifact Signer

RH Advanced Developer Suite

RH Advanced Cluster Security

Closing the vulnerability gap in production

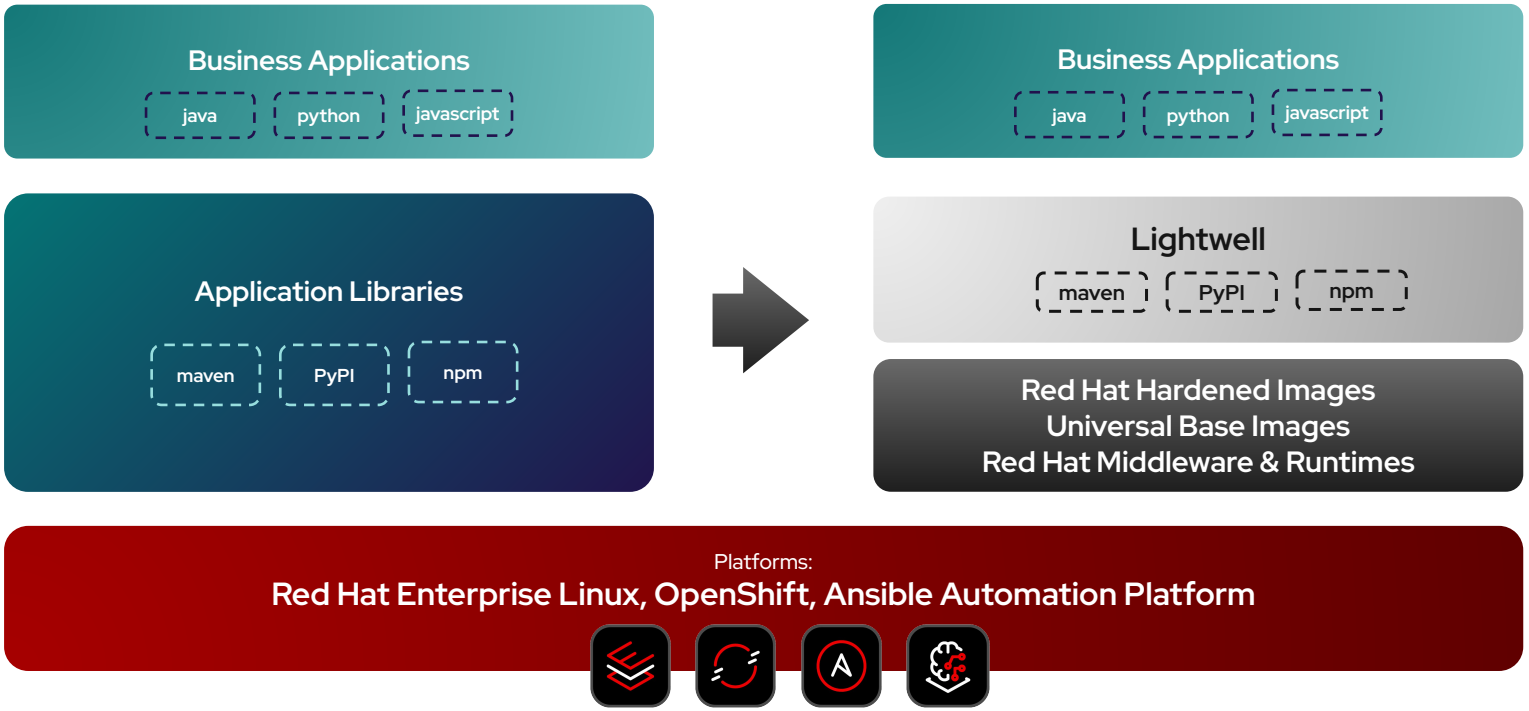


You are responsible for the **Business Applications.**

Who is responsible for the **Open Source libraries** in your applications?

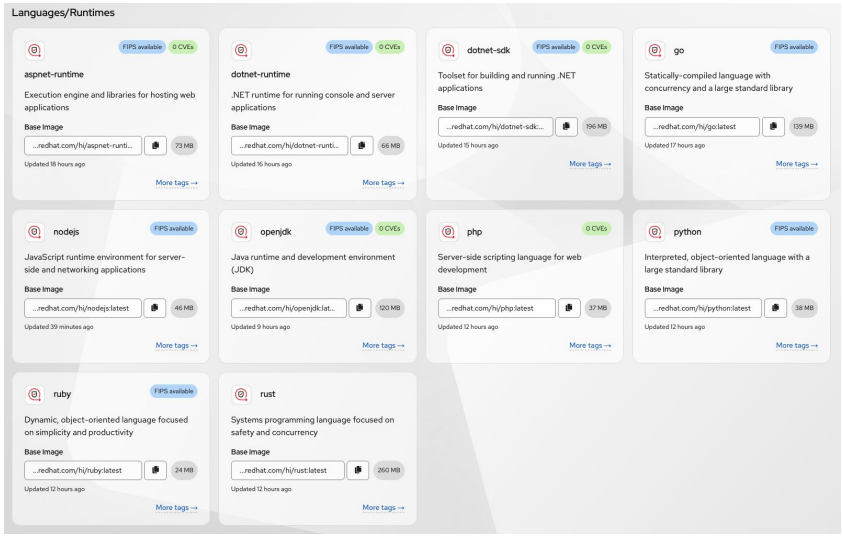
Who is responsible for the code that provides the **platform** you applications run on?

Closing the vulnerability gap in production



Red Hat Hardened Images

Modern software, delivered as hardened, verifiable container images



Foundational set of open source languages, runtimes, databases, web, & utilities aiming to reduce toil around CVEs.

- Built on a minimal, hardened Linux userspace optimized for containerized environments.
- Available at no cost. Support is included with Red Hat Enterprise Linux & OpenShift Container Platform
- Delivered through Red Hat's SLSA 3 trusted supply chain.

Project Lightwell

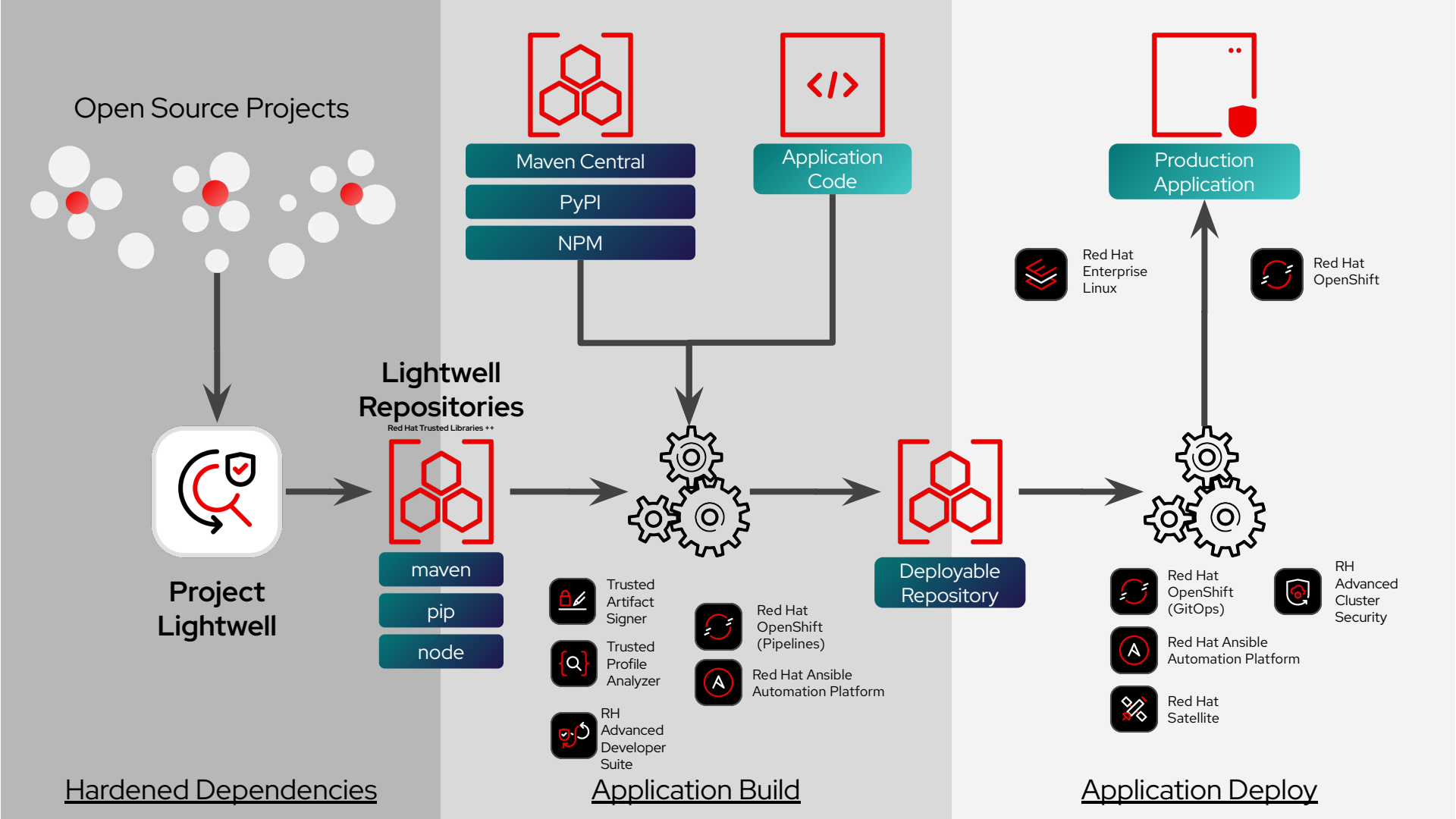
Securing the ecosystems you depend on

- Curated package registries for Java, Python, and JavaScript that are scanned, patched, signed, certified
- Patched packages at the **exact versions** you already run
- SLSA Level 3 builds, HSM-backed signing, SBOMs for every package
- Remediations will be submitted to upstreams to fix bugs at the source
- AI-powered vulnerability triage with automated severity assessment and remediation prioritization

Red Hat Hardened Images - secures the container, including the dependencies inside it.

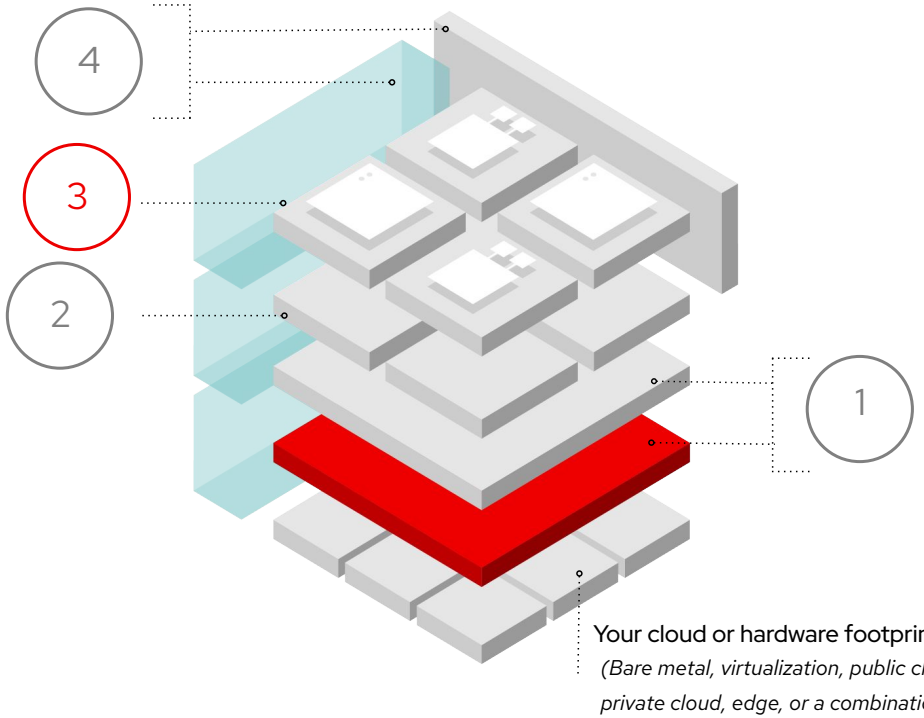
Red Hat Trusted Libraries - available today, with a curated and expanding set of common python libraries.




Open Source Projects
Maven Central
PyPI
NPM
Application Code
Production Application
Red Hat Enterprise Linux
Red Hat OpenShift
Lightwell Repositories
Red Hat Trusted Libraries ++
Project Lightwell
maven
pip
node
Trusted Artifact Signer
Trusted Profile Analyzer
RH Advanced Developer Suite
Red Hat OpenShift (Pipelines)
Red Hat Ansible Automation Platform
Deployable Repository
Red Hat OpenShift (GitOps)
RH Advanced Cluster Security
Red Hat Ansible Automation Platform
Red Hat Satellite
Hardened Dependencies
Application Build
Application Deploy

Layer 3: Secure the Intelligence

Implement guardrails for AI agents and models



1. Start with a strong foundation
2. Protect the pipeline
3. **Implement guardrails for AI agents and models**
4. Automate security at scale

Red Hat

Defense in depth for AI agents

A six-layer security framework provided by Red Hat AI

1. Start with a hardened platform
2. Isolate the runtime
3. Sandbox the process
4. Guard the conversation
5. Attack yourself first
6. Deploy with identity and observability

Red Hat

LAYER + COMPONENTS

WHAT IF THIS LAYER IS ALL YOU HAVE?

Red Hat AI

EACH LAYER ASSUMES BREACH BELOW

1 Start with a hardened platform

OpenShift ships locked down by default. Keep restricted-v2 SCCs, SELinux enforcing, and add DNS-based egress filtering.

Red Hat OpenShift | SCCs (restricted-v2) | SELinux

Without this layer:

Weakened SCCs let agent pods run as root, mount service account tokens, and reach any IP

2 Isolate the runtime

If an agent triggers a container escape, the attacker lands in a throwaway VM, not the host node.

OpenShift sandboxed containers | Kata Containers

Without this layer:

Container escape = host node compromise. Attacker accesses all pods on that node.

3 Sandbox the process

Control what individual binaries can do inside the container. Per-binary network policy, filesystem allowlisting, credential-free inference.

Without this layer:

node and curl have the same permissions. Prompt injection spawns curl to exfiltrate data.

4 Guard the conversation

Intercept prompt injections before the model. Filter unsafe outputs. Gate every tool call.

Without this layer:

Prompt injections reach the model unchecked, agent calls any tool without authorization

5 Attack yourself first

Proactive vulnerability scanning. Run in CI. Feed findings back into guardrails.

Garak (120+ probes) | Chatterbox Labs

Without this layer:

Vulnerabilities discovered by adversaries first, guardrails never improve from real attacks

6 Deploy with identity and observability

Every agent gets a cryptographic identity. Every tool call is traced with full lineage.

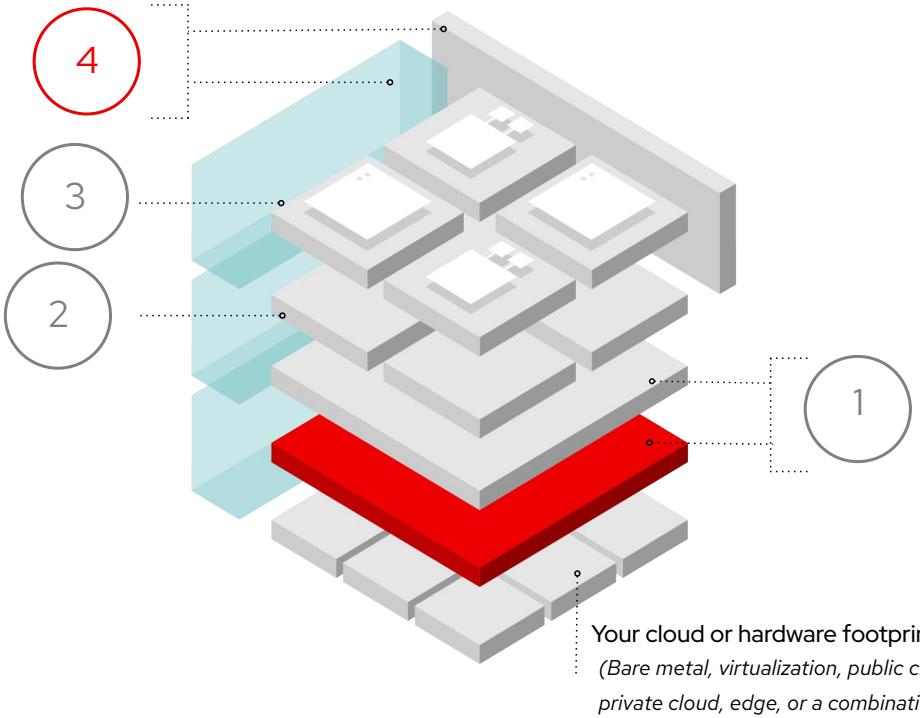
Without this layer:

Shared API keys, no audit trail, no way to trace which agent made which call

The one thing to remember: Start with the assumption that the agent will be compromised. Not might. Will. The question is not whether, but whether your stack limits the impact when it does.

Layer 4: Automate by Default

Automate security at scale



1. Start with a strong foundation
2. Protect the pipeline
3. Implement guardrails for AI agents and models
4. **Automate security at scale**

Red Hat

How Automation can help with security in an AI-Driven world



Speed

From weeks to hours for configuration mitigation and critical patches. Greatly reduces exposure with actions taken in minutes and seconds



Scale

Handle thousands of vulnerabilities. Orchestrate across global hybrid infrastructure.



Intelligence

Supports context-aware prioritization and recommendations for threat mitigation



Governance

Audit trails for every action. Role-based access and change management.

Red Hat

Vulnerability Detection to Remediation with Automation

When a vulnerability drops, the window between discovery and remediation is where the risk lives. **Ansible closes that gap.**

Vulnerability lands. Ansible mitigates.

Respond before the patch is ready

- Rapidly identify affected systems for threat assessment.
- Enable workflows that isolate affected systems and disable exposed services automatically.
- Trigger increased alerting without human intervention.

Patch with confidence, when ready.

The full pipeline tested and validated.

- Patch vulnerabilities consistently when patches are ready using Ansible Automation Platform pipelines.
- Enable validations to ensure systems are back online before declaring success.
- Enable the ability to return to a known state with automated rollbacks.

Prove it worked. Every time.

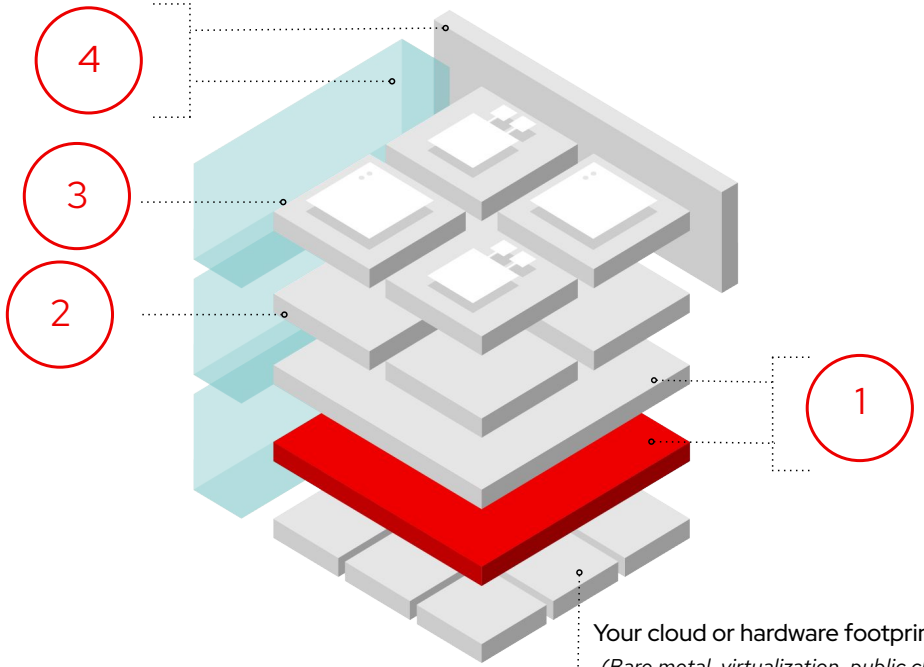
Formalize governance before you scale

- Record every action in your ITSM (detection, remediation, validation, rollback)
- Know exactly what's automated, what needs human approval, and who owns what.

Red Hat

In Summary - Trustworthy by Design

Security from the OS to Autonomous AI Agents



1. Start with a strong foundation
2. Protect the pipeline
3. Implement guardrails for AI agents and models
4. Automate security at scale

Red Hat



Thank you

Red Hat is the world's leading provider of enterprise open source software solutions. Award-winning support, training, and consulting services make Red Hat a trusted adviser to the Fortune 500.

linkedin.com/company/red-hat

facebook.com/redhat

youtube.com/@redhat

x.com/RedHat

Red Hat