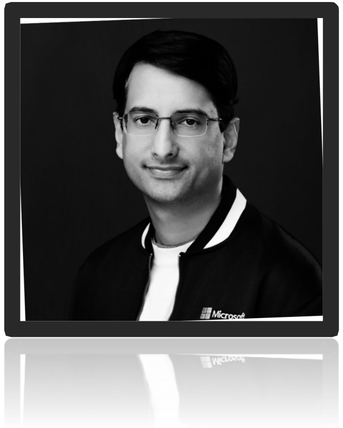
Zero Trust for AI Systems

Razi Rais

Meet Your Instructor – Razi Rais

Cybersecurity Leader | Author | Microsoft Certified Trainer



- **20+ years** in cybersecurity & systems architecture
- **10+ years** at Microsoft in engineering, architecture, and product leadership
- **Global career** in France, Dubai, Singapore with 15+ years in the United States
- **Zero Trust & AI** focus applying AI to modern security architectures
- **Author &Trainer** 5 books, Microsoft Certified Trainer (15+ years), GIAC Advisory Board member, RSAC speaker

Let's connect on LinkedIn
https://linkedin.com/in/razirais

Learning Objectives

- Grasp the fundamentals and core pillars of Zero Trust Security
- Explore the impact of AI on Zero Trust both its benefits and challenges
- Examine AI security frameworks (NIST, OWASP, CSA) and standards aligned with Zero Trust principles
- Engage in practical exercises, interactive discussions, and Q&A sessions to reinforce learning

Prerequisites

No prior Zero Trust or AI knowledge needed just a basic grasp of cybersecurity. We'll share everything else along the way. Bring your curiosity!

Note: *This course is completely vendor-agnostic. The knowledge you gain applies across platforms and cloud environments.*

Resources

Slides and course recording will be shared after the session. Additional resources and references are available on GitHub:
https://github.com/razi-rais/Zero-Trust-for-AI-Systems

Poll

When did you first hear about Zero Trust?

Choose the one that best aligns with your experience.

Choices:

- Recently, within the past year
- A few years ago as the concept gained traction
- A decade ago when early discussions started
- Only after my organization adopted or mentioned it

Zero Trust: How it started?

Year 2010 | Term “Zero Trust” was coined by Forrester

November 5, 2010

Build Security Into Your Network’s DNA: The Zero Trust Network Architecture

by John Kindervag
with Stephanie Balaouras and Lindsey Coit

EXECUTIVE SUMMARY

This report is a deep dive into a potential way in which you could use the concepts of the Zero Trust Model and conceivably implement them in a real-world environment. One of our goals with Zero Trust is to optimize the security architectures and technologies for future flexibility. As we move toward a data-centric world with shifting threats and perimeters, we look at new network designs that integrate connectivity, transport, and security around potentially toxic data. We call this “designing from the inside out.” If we begin to do all those things together we can have a much more strategic infrastructure. If we look at everything from a data-centric perspective, we can design networks from the inside out and make them more efficient, more elegant, simpler, and more cost-effective.

TABLE OF CONTENTS

2 **Forrester’s Zero Trust Network Security Report Collection**

2 **Zero Trust Will Change The Way We Design And Build Networks**

5 **Use Zero Trust To Rebuild The Secure Network**

NOTES & RESOURCES

In developing this report, Forrester drew from a wealth of analyst experience, insight, and research through advisory and inquiry discussions with end users, vendors, and regulators across industry sectors.

Year 2013 | Forrester respond to Cybersecurity Executive Order (EO)

The National Institute of Science and Technology

Developing a Framework to Improve Critical Infrastructure Cybersecurity

In Response to:

RFI# 130208119-3119-01

Submitted On: 04/08/2013

Overview

In February 2013 President Obama's Cybersecurity Executive Order (EO) made public the clear and present danger of cyber warfare. The President called for the Federal Government and its Agencies to lead the fight against cyber criminals. As part of this call to action, President Obama asked the National Institute of Standards and Technology (NIST) to gather industry and Federal feedback to create a set of voluntary policies to help develop the US's cybersecurity framework.

In order to keep up with the continually changing cybersecurity landscape, the Federal Government and organizations in important industries such as finance, utilities, and Federal contractors must fundamentally shift the way in which they think about cybersecurity. The traditional mindset does not take into account the current environment; changes like mobility and big data have made "building stronger walls" an expensive farce that will not adequately protect networks.

To help answer the cybersecurity questions of today while allowing for proactive growth in the future, Forrester has outlined our proprietary "Zero Trust Model" (Zero Trust) of information security. Zero Trust changes the way that organizations think about cybersecurity and better protects valuable information while allowing for free interactions internally. The major benefits of Zero Trust to the Federal Government include:

- **Zero Trust is applicable across all industries and organizations** – It is an easy to implement way to improve safety that any organizations can implement.
- **Zero Trust is not dependent on a specific technology or vendor** – Zero Trust is a vendor neutral design philosophy that allows maximum flexibility to create architectures that meet specific demands.
- **Zero Trust is scalable** – Vital information is protected while public facing data travels freely.
- **There is no chance of violating Civil Liberties** – Zero Trust focuses on keeping internal data safe and would not result in any foreseeable encroachment on Civil Liberties.

Ref: https://www.nist.gov/system/files/documents/2017/06/05/040813_forrester_research.pdf

Year 2014 | Google's Beyond Corp

SECURITY

BeyondCorp: A New Approach to Enterprise Security

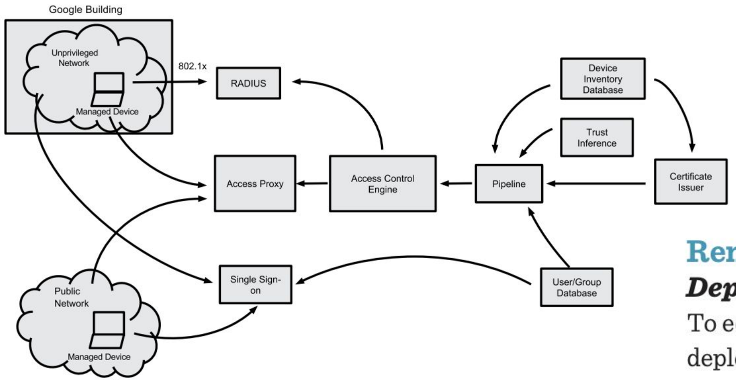


Figure 1: BeyondCorp components and access flow

Removing Trust from the Network

Deployment of an Unprivileged Network

To equate local and remote access, BeyondCorp defines and deploys an unprivileged network that very closely resembles an external network, although within a private address space. The unprivileged network only connects to the Internet, limited infrastructure services (e.g., DNS, DHCP, and NTP), and configuration management systems such as Puppet. All client devices are assigned to this network while physically located in a Google building. There is a strictly managed ACL (Access Control List) between this network and other parts of Google's network.

Ref: https://research.google/pubs/pub43231

Year 2020 – 2022 | US Govt Initiatives

NIST Special Publication 800-207

Zero Trust Architecture

Scott Rose
Oliver Borchert
Stu Mitchell
Sean Connelly

This publication is available free of charge from:
https://doi.org/10.6028/NIST.SP.800-207

C O M P U T E R   S E C U R I T Y

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

NIST's Zero Trust Architecture

2020

THE WHITE HOUSE

Administration Priorities COVID Plan Briefing Room Español MENU

BRIEFING ROOM

Executive Order on Improving the Nation's Cybersecurity

MAY 12, 2021 • PRESIDENTIAL ACTIONS

Sec. 3. Modernizing Federal Government Cybersecurity.

(a) To keep pace with today's dynamic and increasingly sophisticated cyber threat environment, the Federal Government must take decisive steps to modernize its approach to cybersecurity, including by increasing the Federal Government's visibility into threats, while protecting privacy and civil liberties. The Federal Government must adopt security best practices; advance toward Zero Trust Architecture; accelerate movement to secure cloud services, including Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS); centralize and streamline access to cybersecurity data to drive analytics for identifying and managing cybersecurity risks; and invest in both technology and personnel to match these modernization goals.

EO: Adoption of Zero Trust Architecture

2021

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

January 26, 2022

M-22-09

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Shalanda D. Young
Acting Director

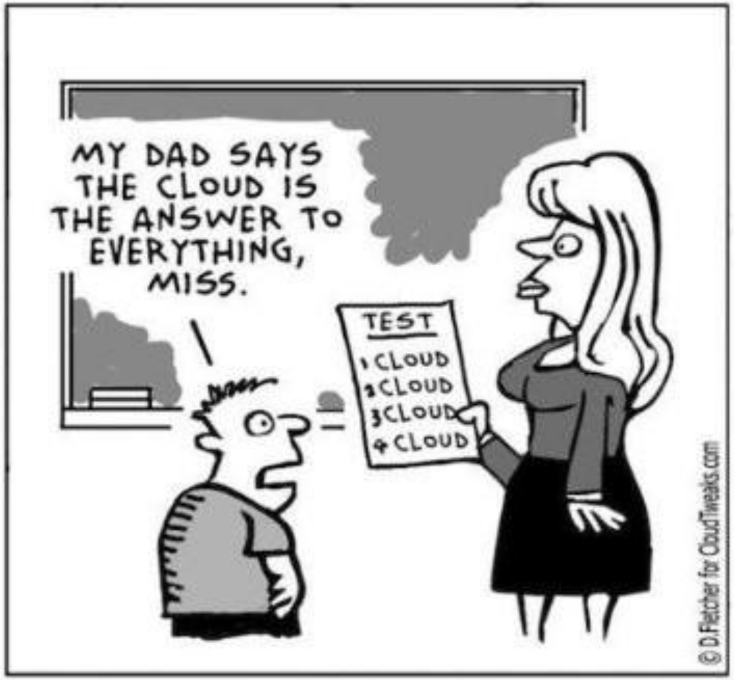
SUBJECT: Moving the U.S. Government Toward Zero Trust Cybersecurity Principles

This memorandum sets forth a Federal zero trust architecture (ZTA) strategy, requiring agencies to meet specific cybersecurity standards and objectives by the end of Fiscal Year (FY) 2024 in order to reinforce the Government's defenses against increasingly sophisticated and persistent threat campaigns. Those campaigns target Federal technology infrastructure, threatening public safety and privacy, damaging the American economy, and weakening trust in Government.

Memorandum: Moving Toward Zero Trust Cybersecurity Principles

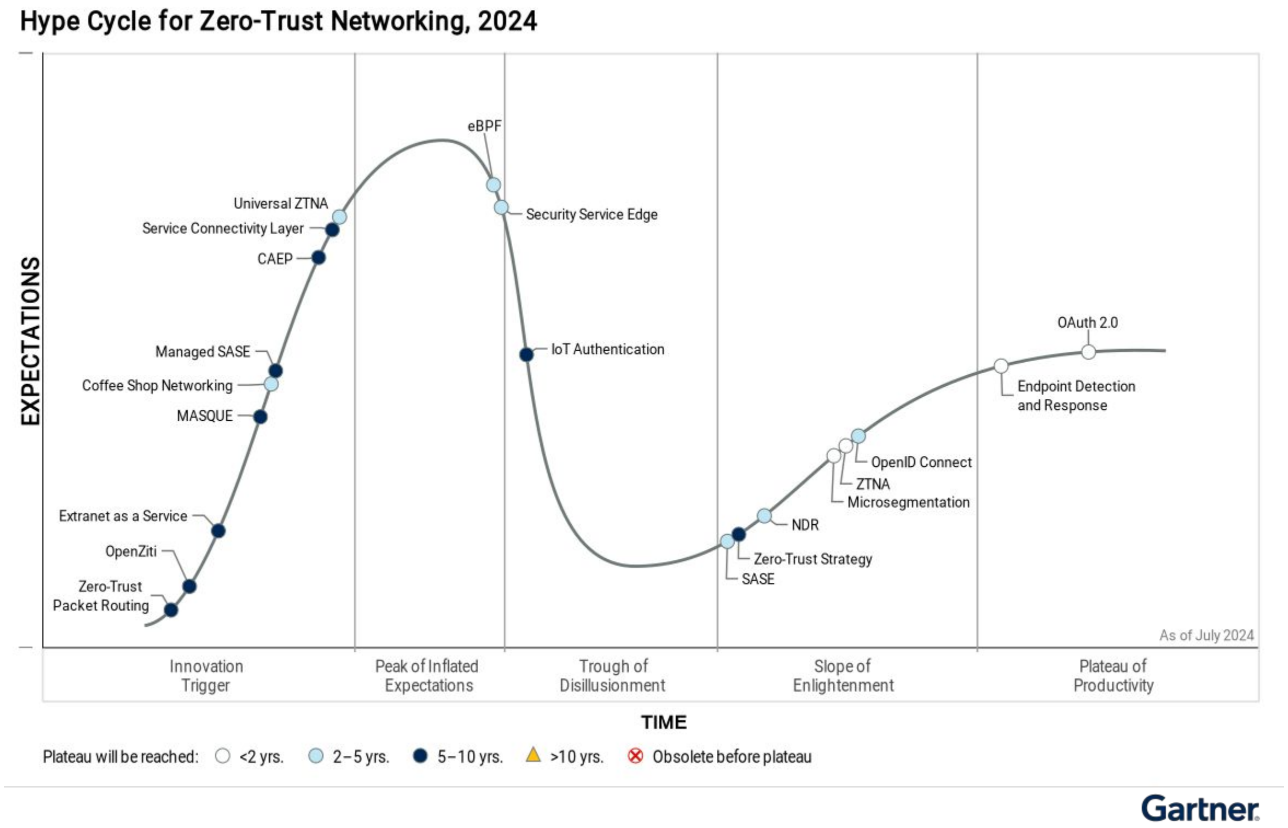
2022

Hype?


MY DAD SAYS THE CLOUD IS THE ANSWER TO EVERYTHING, MISS.
TEST
1 CLOUD
2 CLOUD
3 CLOUD
4 CLOUD
© D.Fletcher for CloudTweaks.com

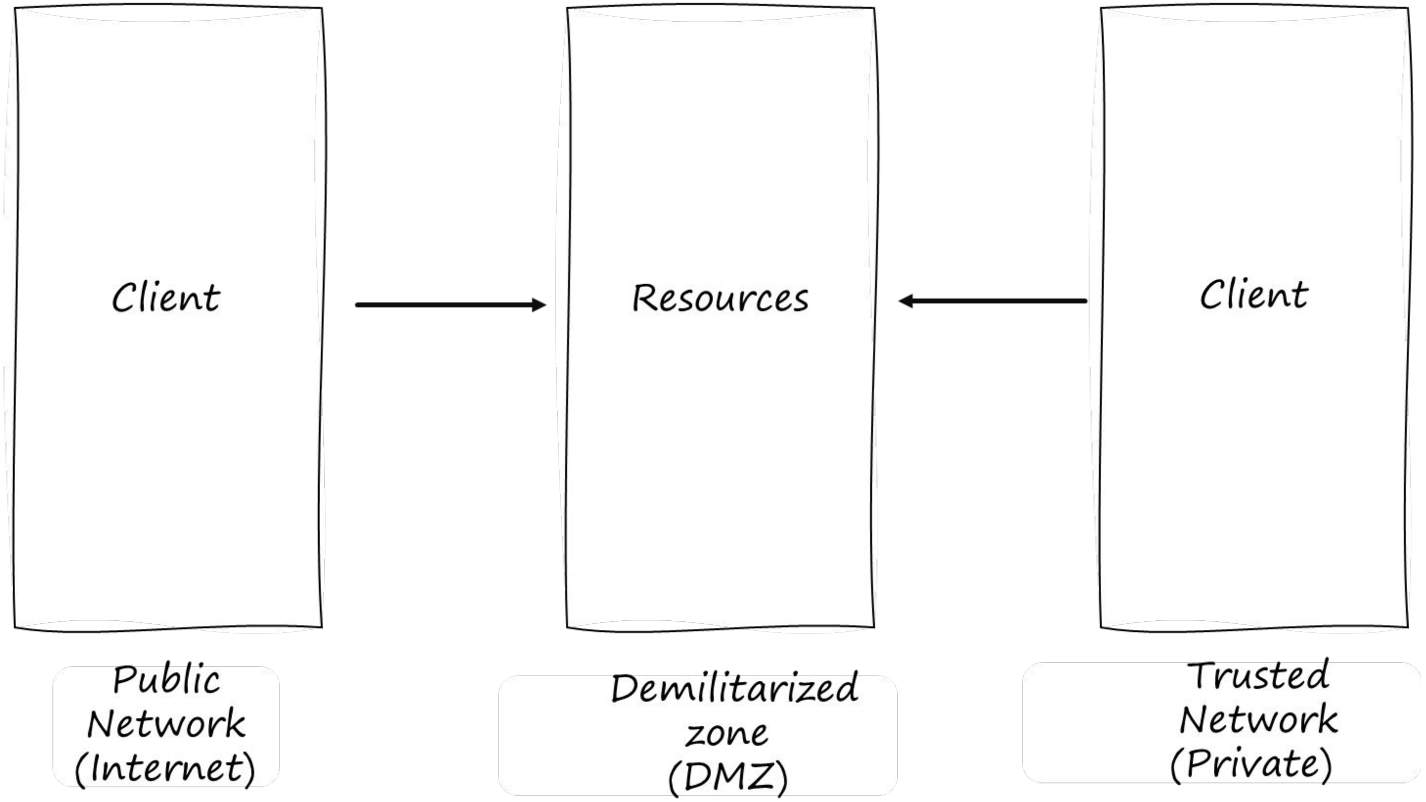

I WAS HOPING FOR A SLIGHTLY MORE DETAILED EXPLANATION OF HOW CLOUD COMPUTING WORKS THAN – "IT'S MAGIC"!
IT
© D.Fletcher for CloudTweaks.com

Gartner: Hype Cycle For Zero Trust Networking

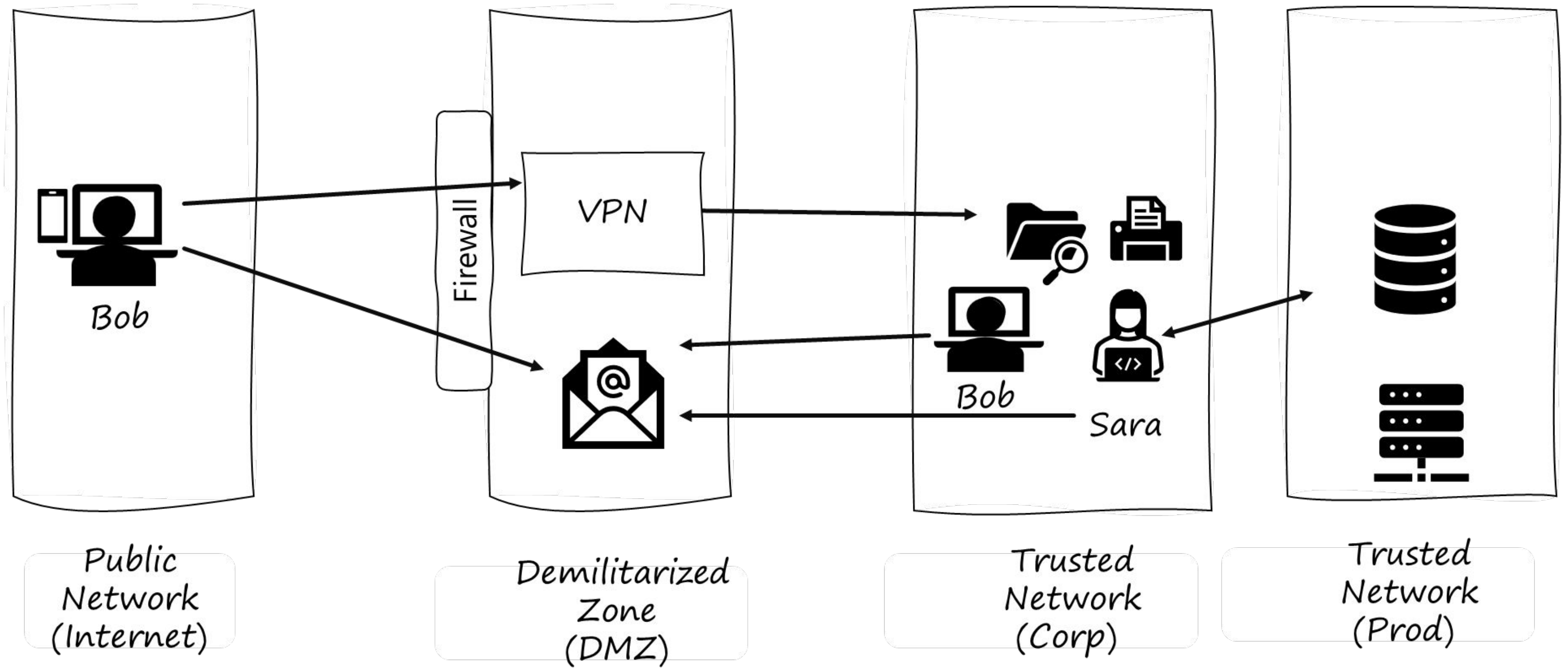


Zero Trust Security: Scenario Walkthrough

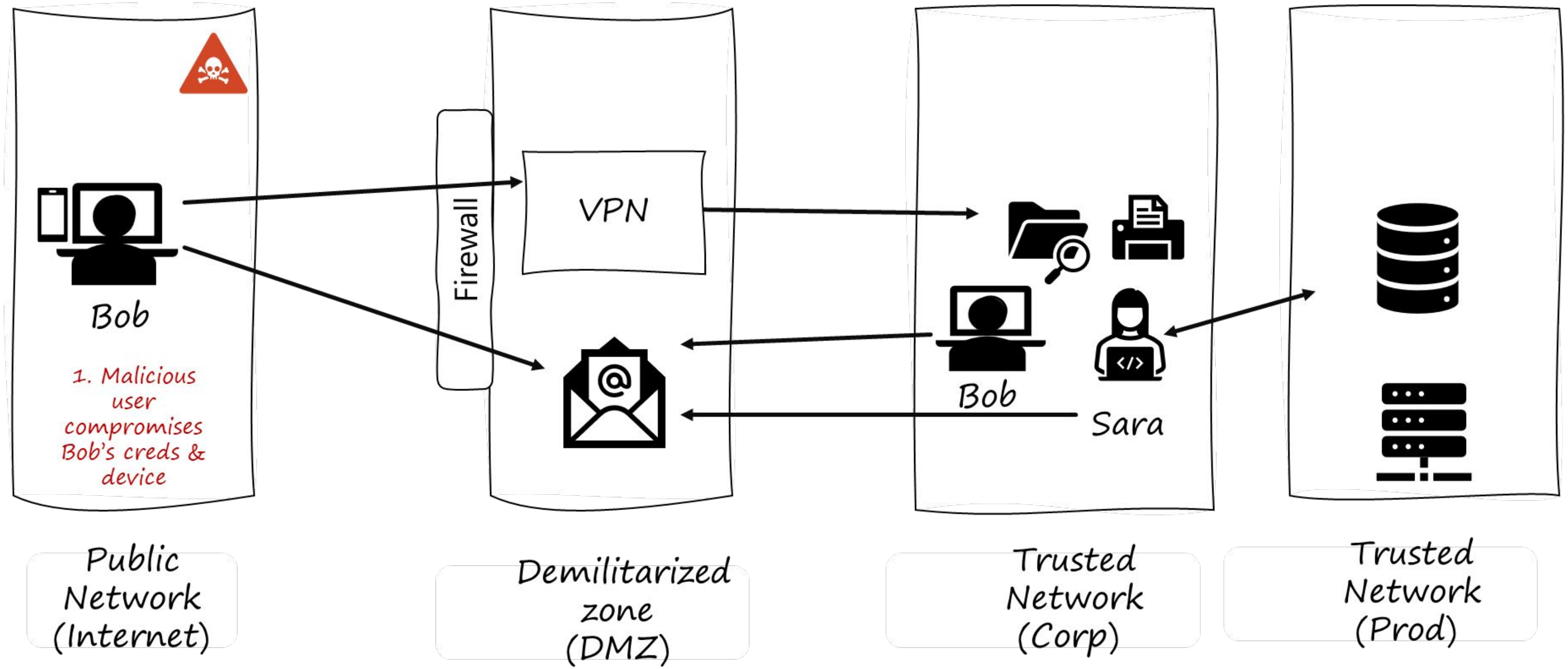
Perimeter based security - Conceptual Model



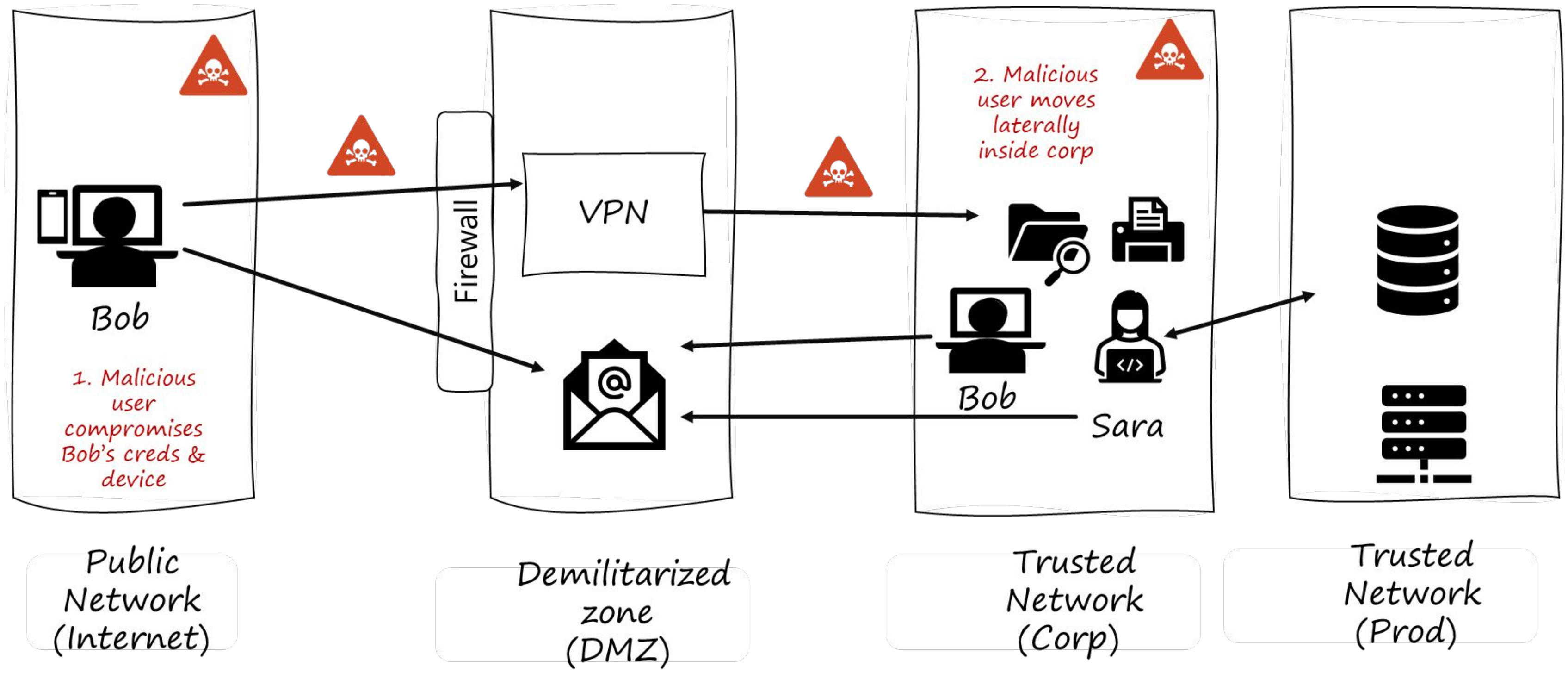
Perimeter based security - Conceptual Model



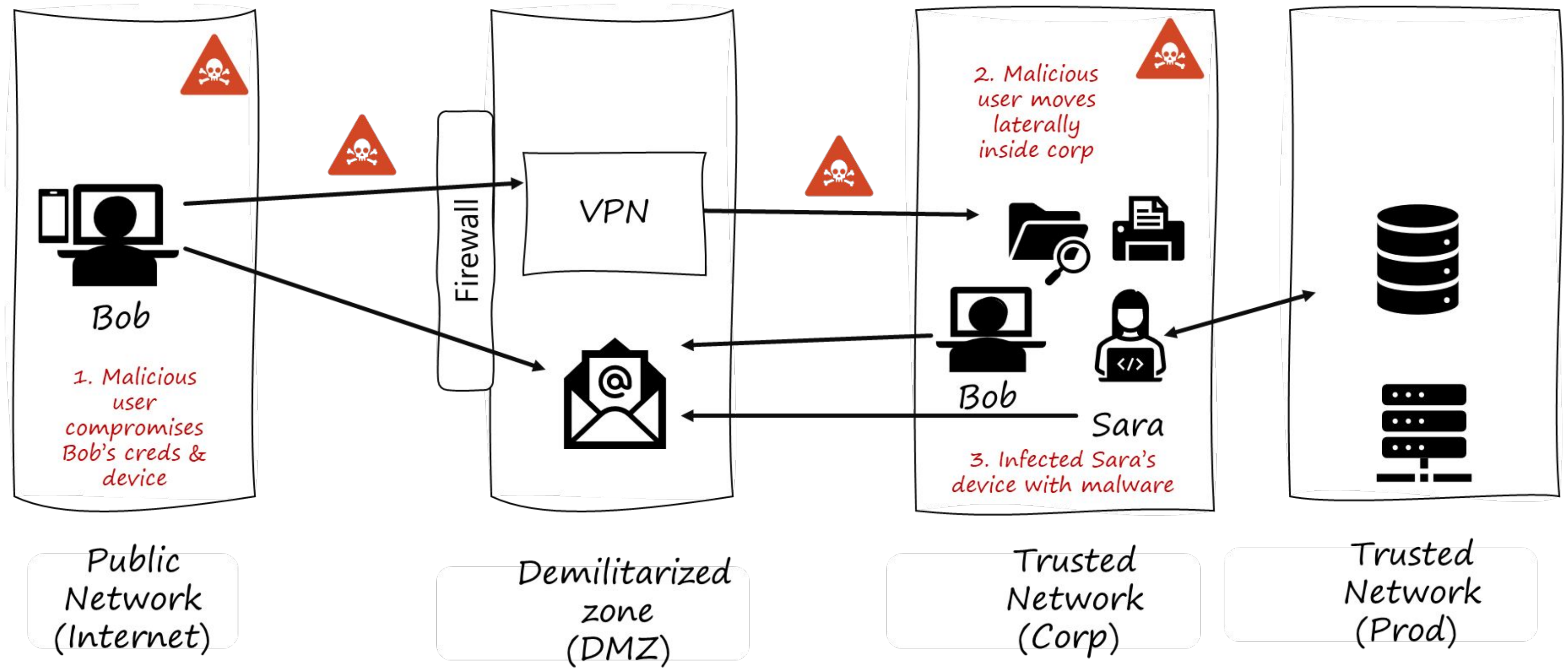
Perimeter based security - Conceptual Model



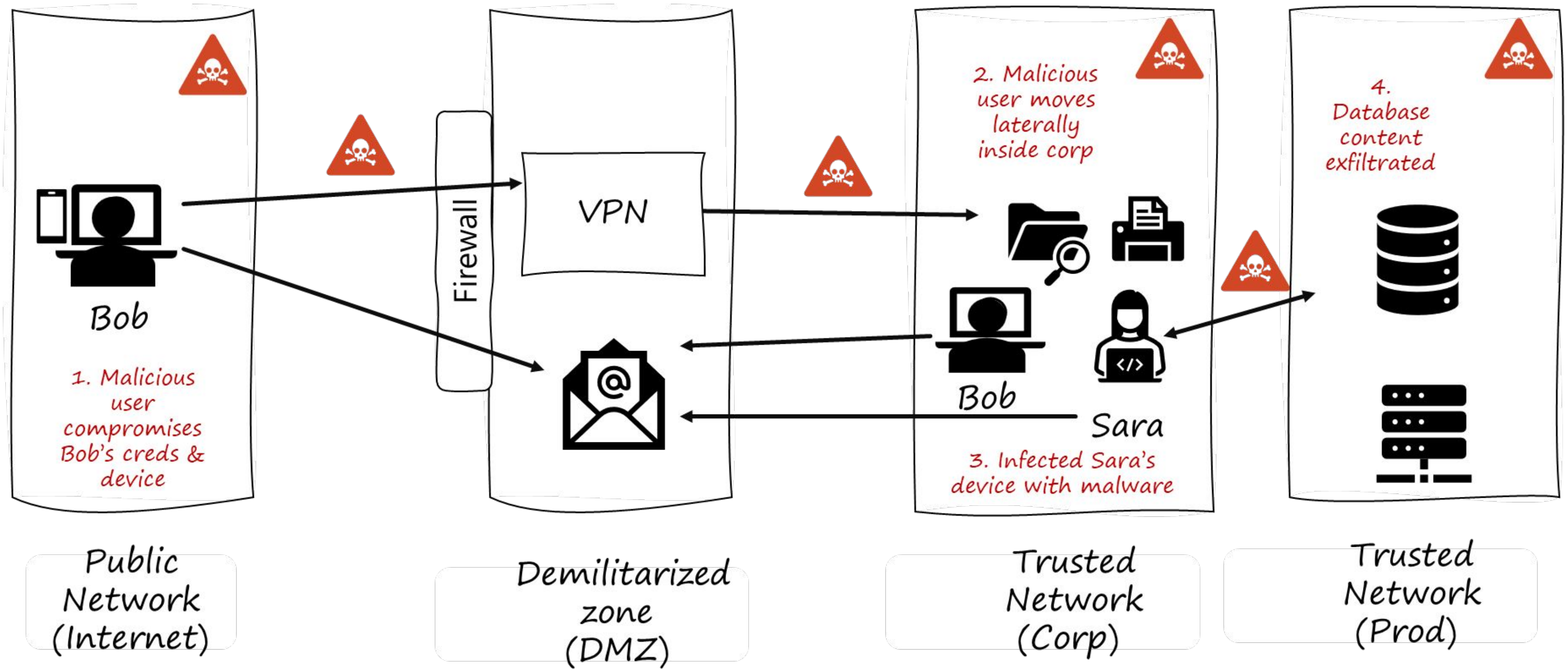
Perimeter based security - Conceptual Model



Perimeter based security - Conceptual Model



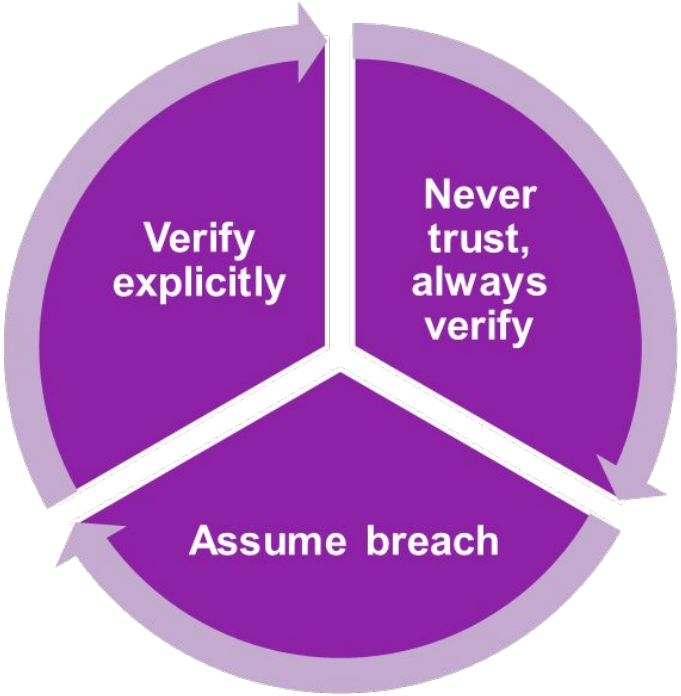
Perimeter based security - Conceptual Model



What went wrong?

1. Zones defines trust boundaries.
2. Malicious user lateral movement went un-detected.
3. Permissions to access privileged resources inside the trusted network were not time-bound.
4. Lack of monitoring in general.

Let’s take a different approach



Zero Trust Approach

Never Trust, Always Verify

Treat every user, device, application/workload, and data flow as untrusted.

Authenticate and explicitly authorize each request to the least privilege.

Verify Explicitly

Access to all resources should be conducted in a consistent and secure manner using multiple attributes (dynamic and static) to derive confidence levels for contextual access decisions to resources.

Access is time-bound.

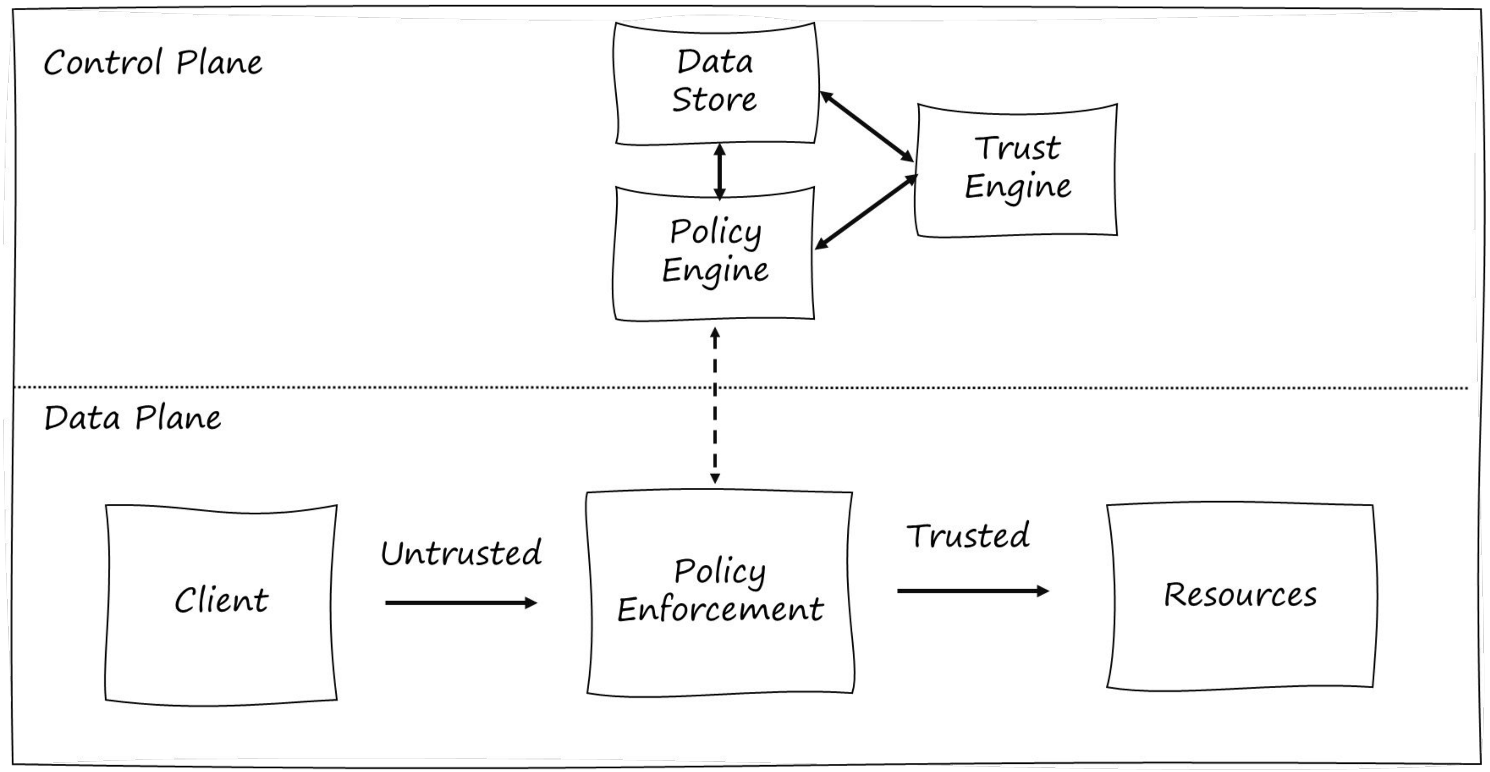
Assume Breach

Operate and defend resources with the assumption that an adversary already has presence within the environment.

Deny by default and heavily scrutinize all users, devices, data flows, and requests for access.

Log, inspect, and continuously monitor all configuration changes, resource accesses, and network traffic for suspicious activity.

Zero Trust Security – Conceptual Model



Zero Trust Security – Conceptual Model

- Act as a **source of truth** for other components in the control plane:
  - User Information (id, name, location, etc.)
  - Device Information (serial, build, os, etc.)
  - Historic Activity (activity logs, analytics via monitoring etc.)

Data Store

- Act as a **source of truth** for other components in the control plane:
  - User Information (id, name, location, etc.)
  - Device Information (serial, build, os, etc.)
  - Historic Activity (activity logs, analytics via monitoring etc.)

Trust Engine

- Performs **risk analysis** and generate a **risk score** against a request
  - Static: Predefined ad-hoc rules based on **user, app, and device** attributes.
  - Dynamic: **Machine learning** is used to calculate the score based on the training data to find **anomalous behavior patterns**.
- Combination of static and dynamic risk analysis works better.
- Trust engine uses data store as a source of truth, that’s why data store should always be kept up-to-date.

Policy Engine

- Responsible for the final decision **to allow or deny** a request
  - Relies on trust engine to obtain a risk score.
  - Relies on data store to fetch data to apply ad-hoc policy rules. Think regulatory, compliance, and organization policies to meet business requirements etc.
- **Example**:
  - Bob is trying to access HR application from un-known device during after hours.

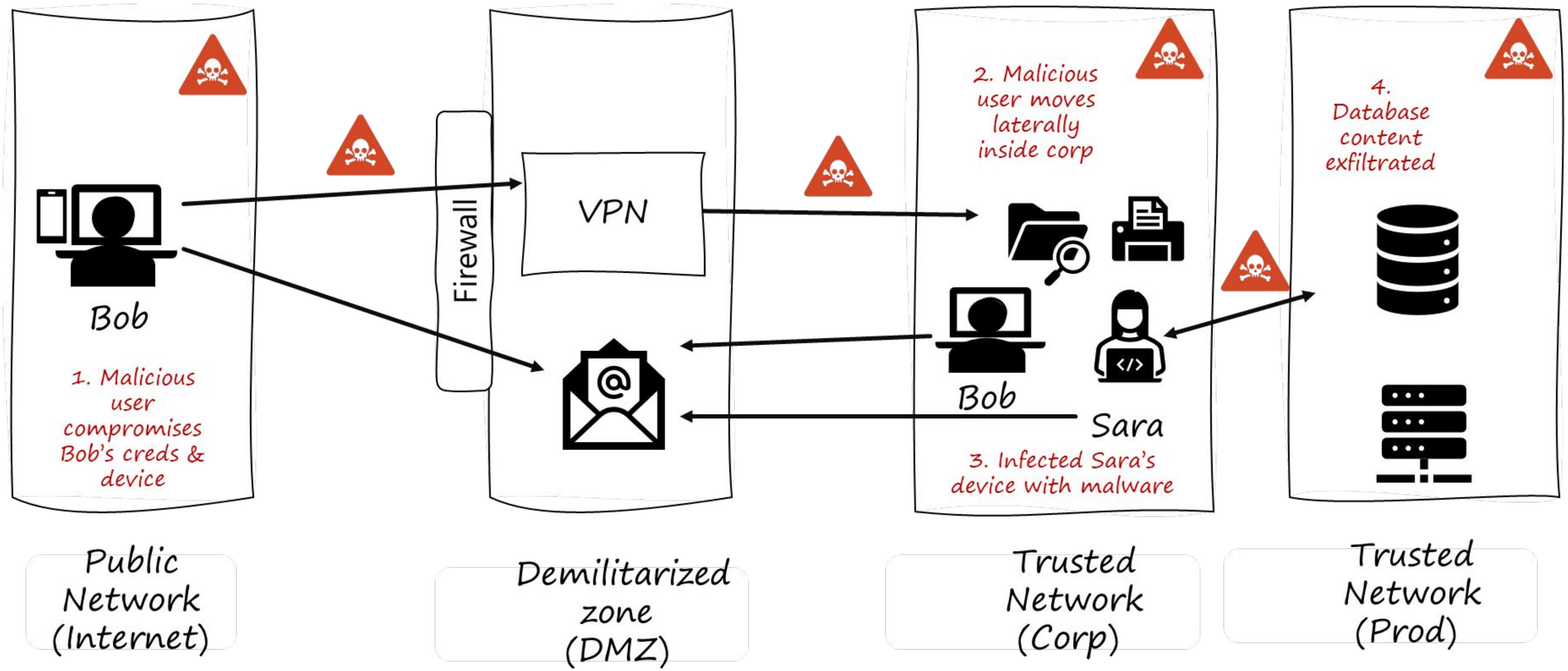
Policy Enforcement

- Responsible for:
  - Obtaining **allow/deny** decision from the policy engine against a request.
  - **Enforcing policy** engine's decision
- Decoupling: Communication to the policy engine and enforcement of the policy can be done by a single component or by multiple components.

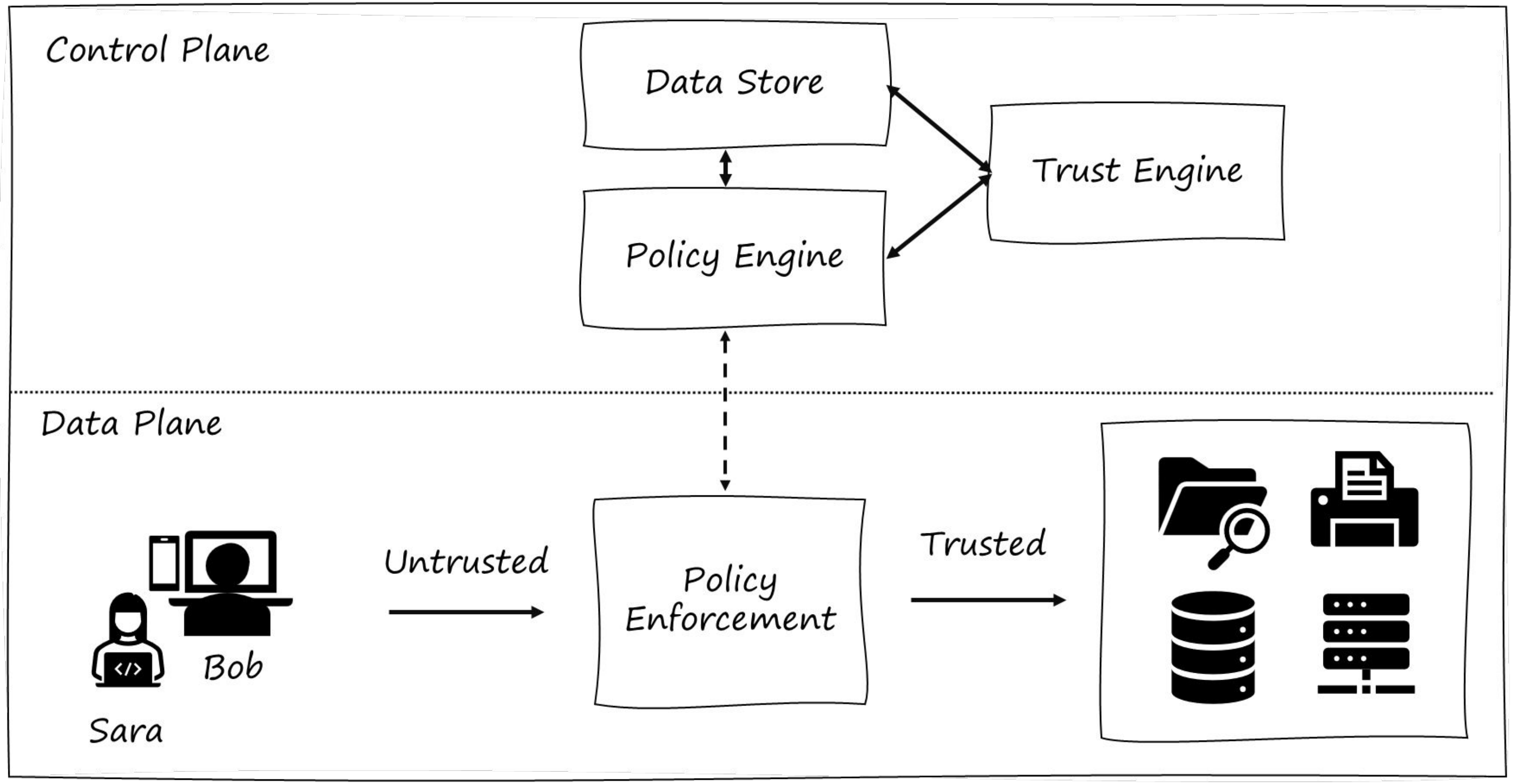
Let me try again to fix!

1. Zones defines trust boundaries.
2. Malicious user lateral movement went un-detected.
3. Permissions to access privileged resources inside the trusted network were not time-bound.
4. Lack of monitoring in general.

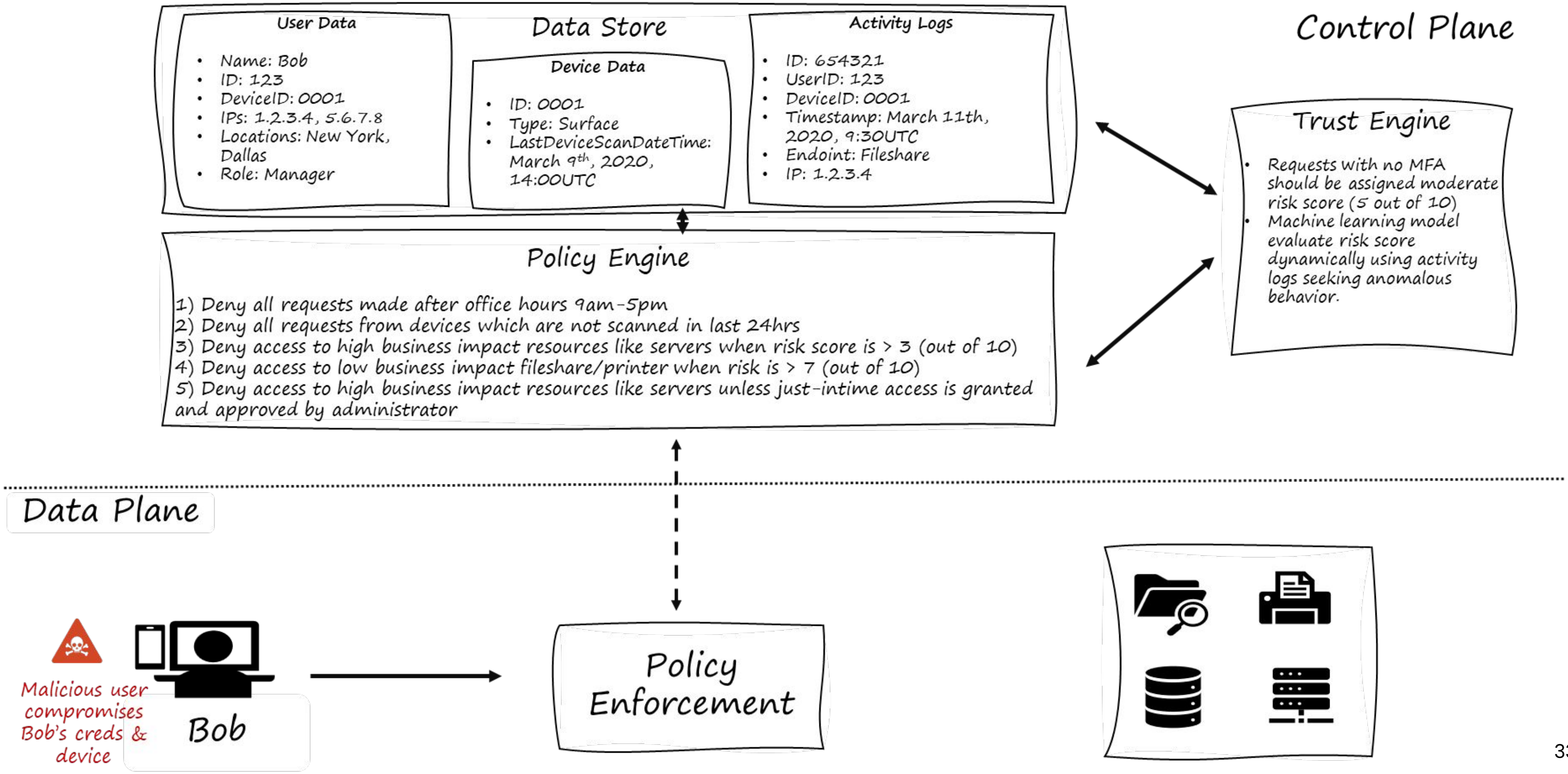
Perimeter based security Recap!



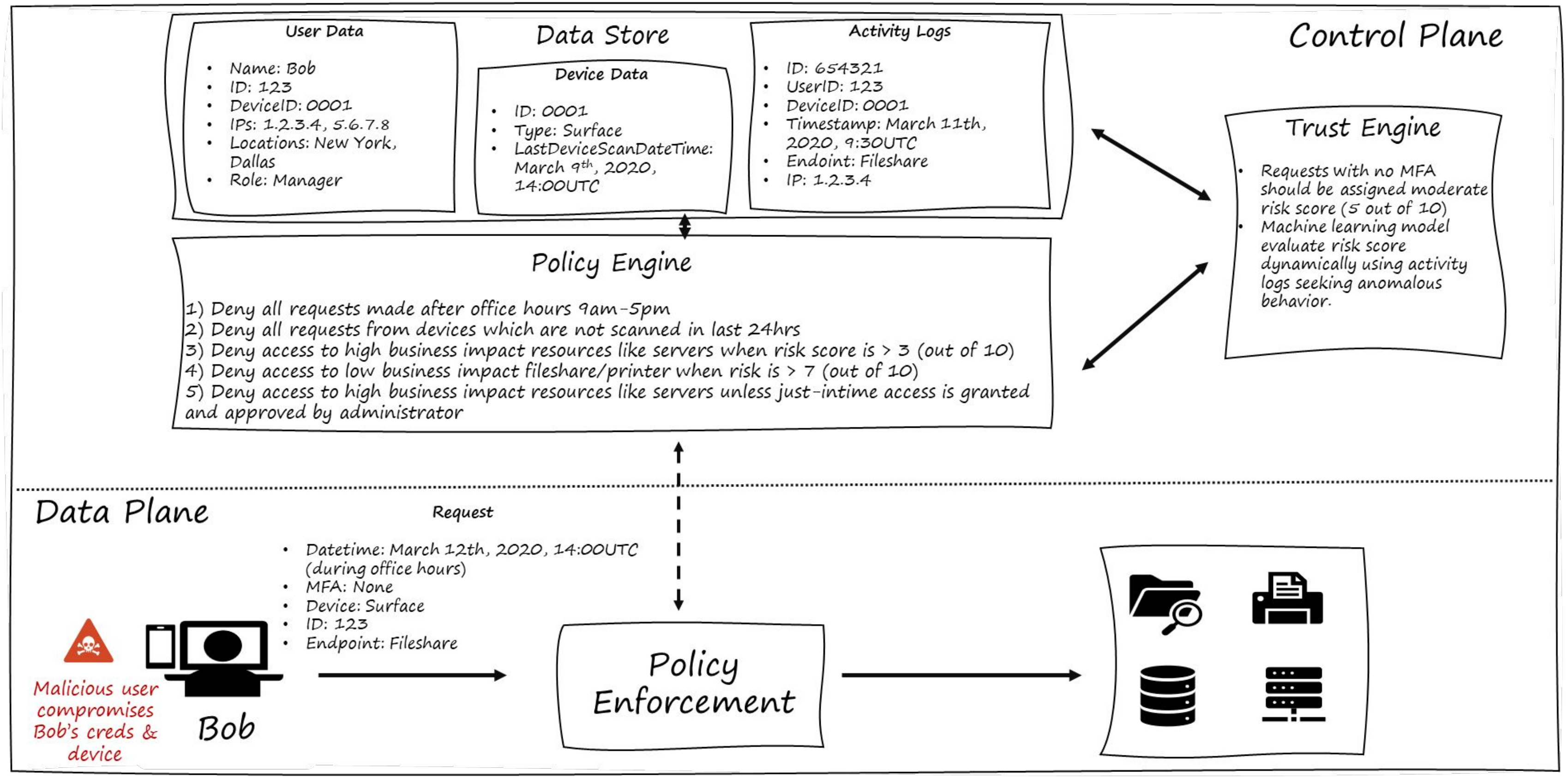
Zero Trust – Scenario walkthrough revisited!



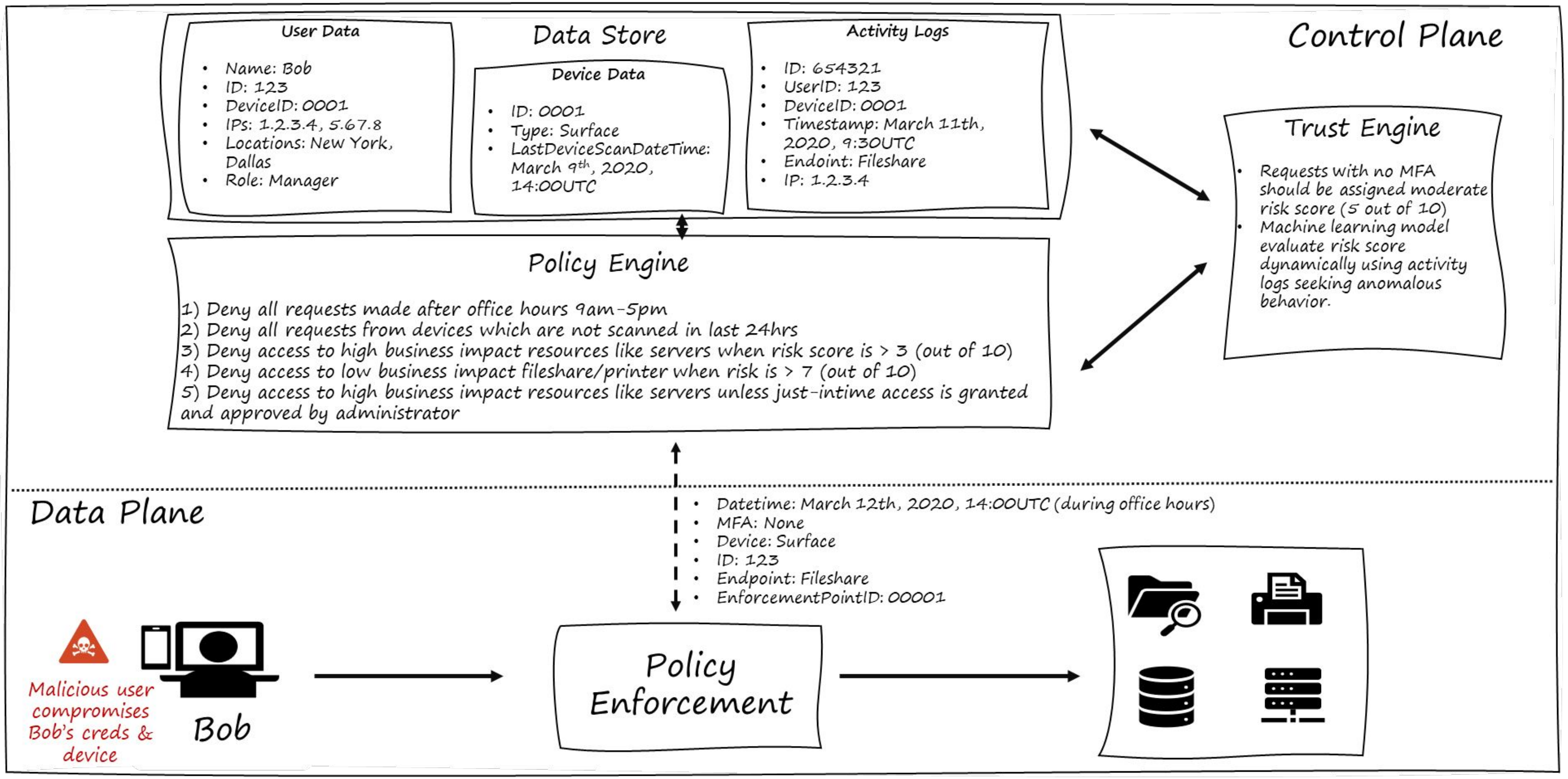
Scenario Walkthrough | Bob’s creds & device is compromised



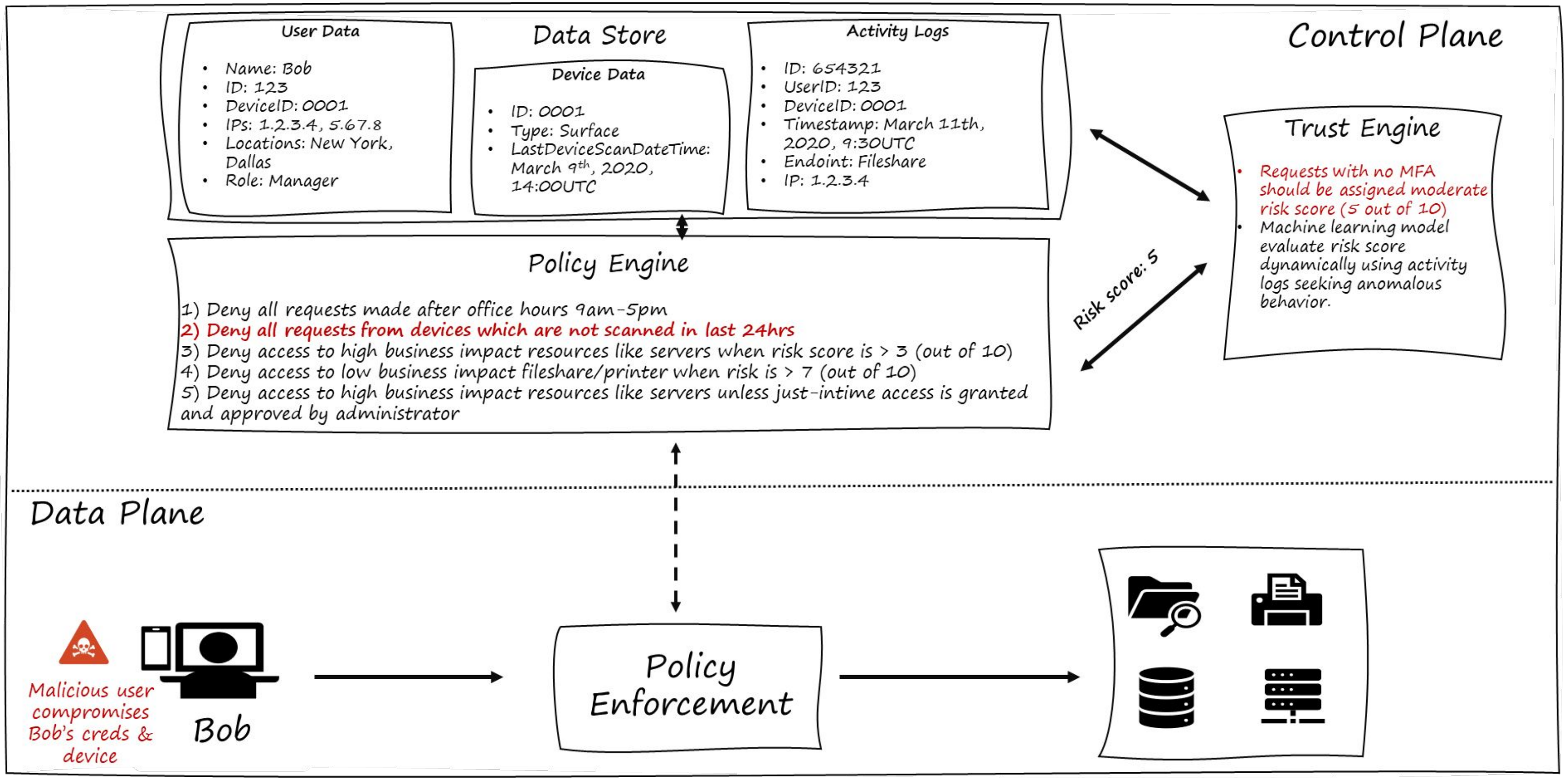
Scenario Walkthrough | Bob's creds & device is compromised



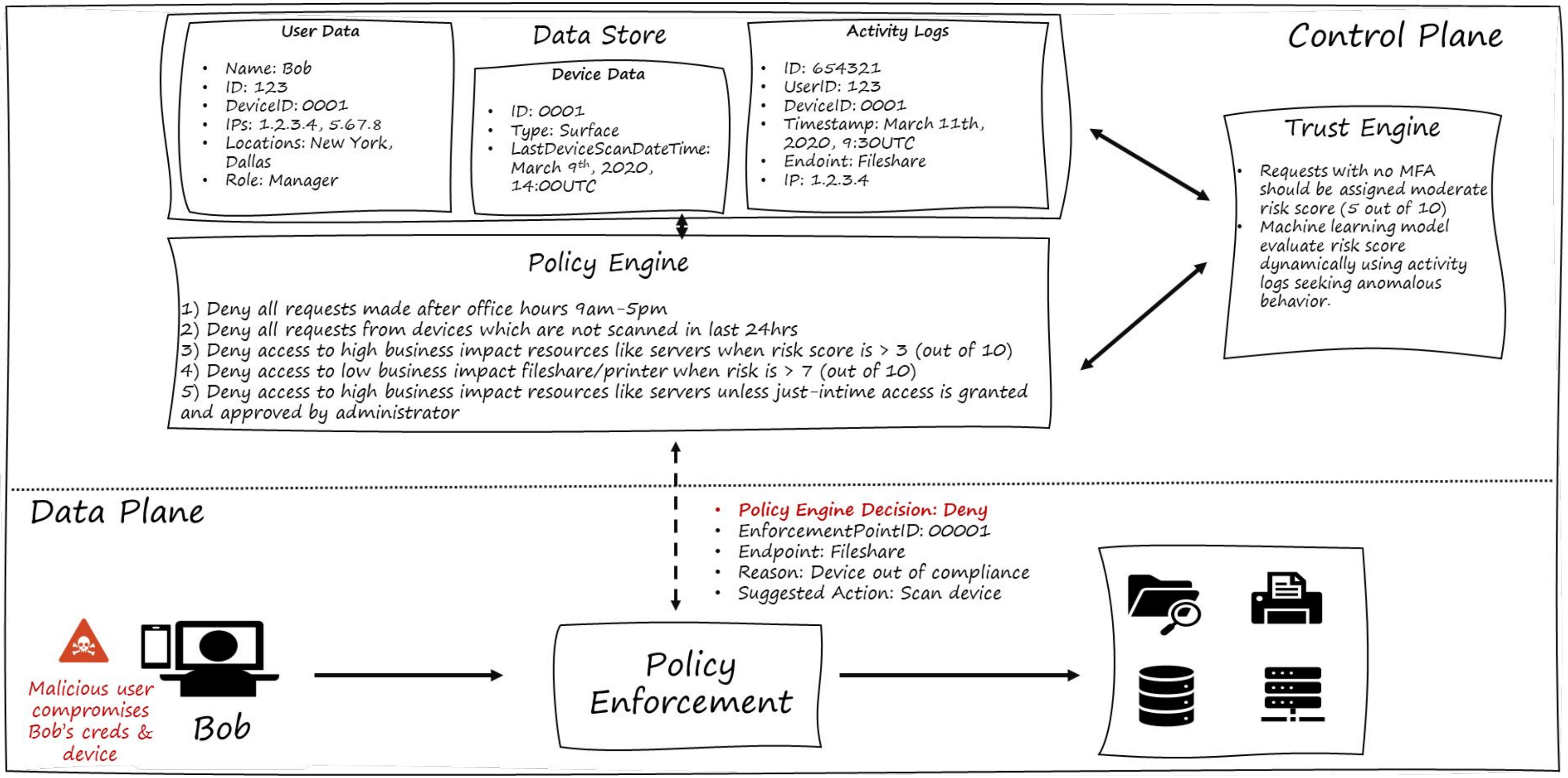
Scenario Walkthrough | Bob's creds & device is compromised



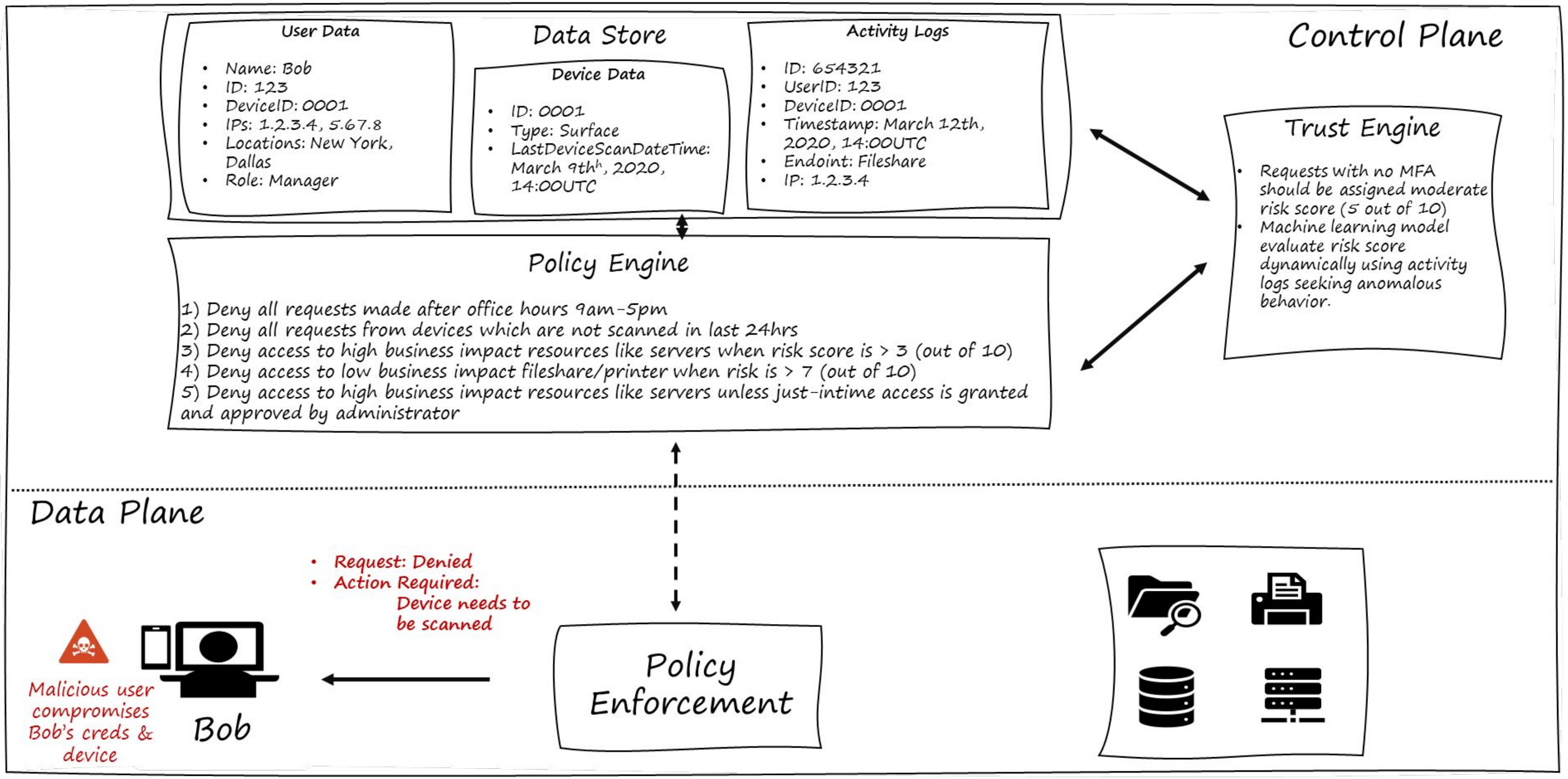
Scenario Walkthrough | Bob's creds & device is compromised



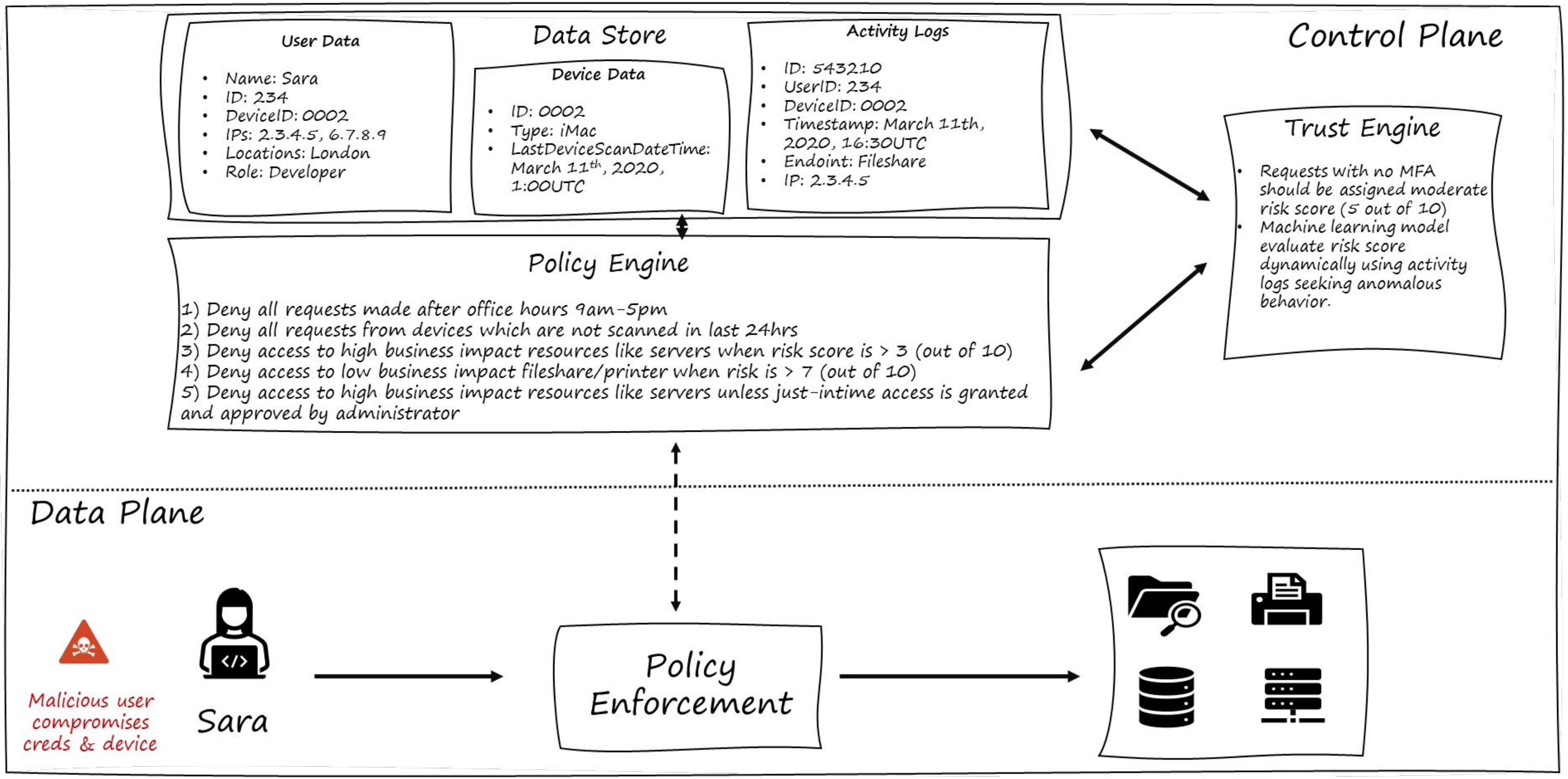
Scenario Walkthrough | Bob's creds & device is compromised



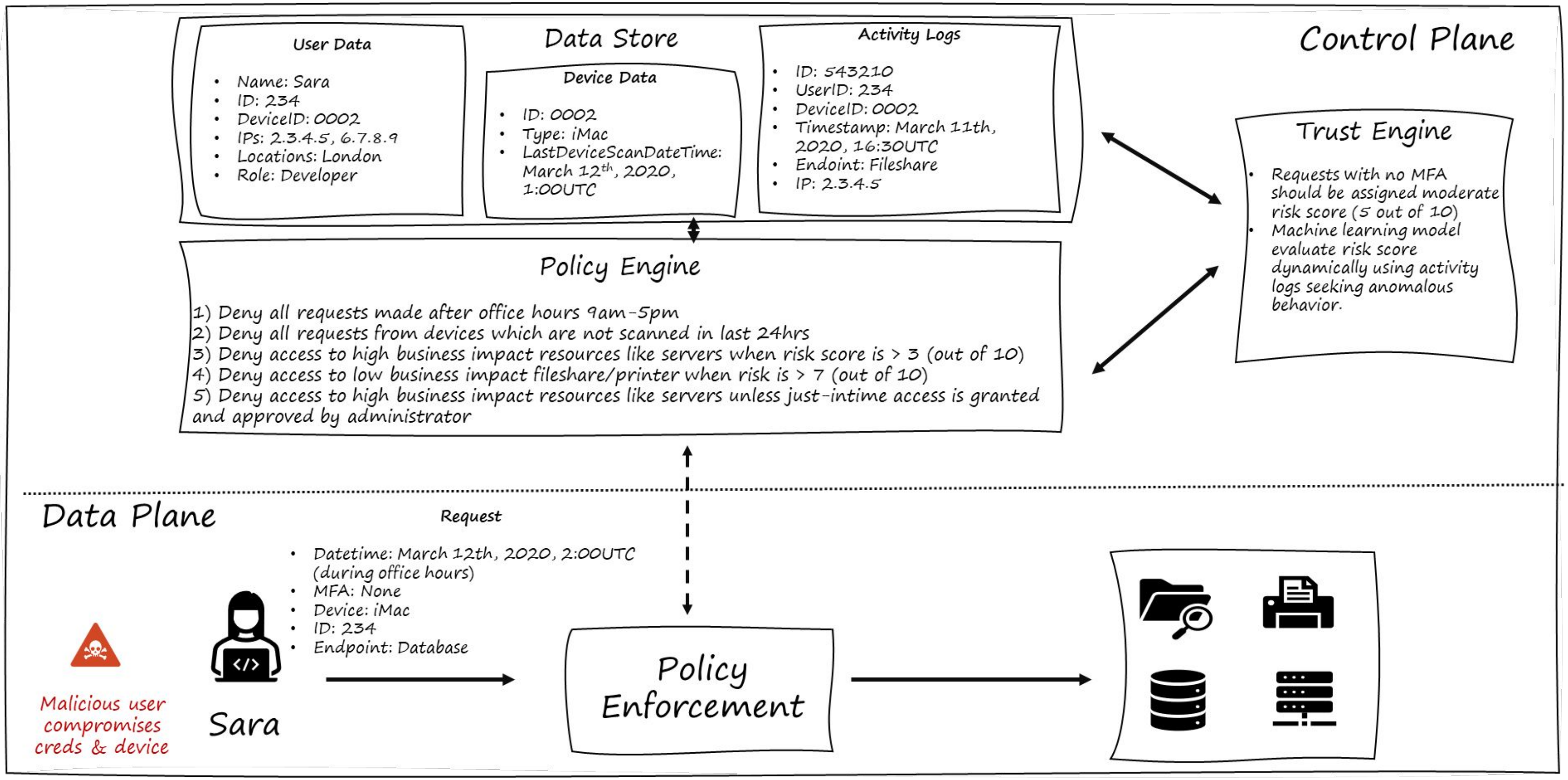
Scenario Walkthrough | Bob’s creds & device is compromised



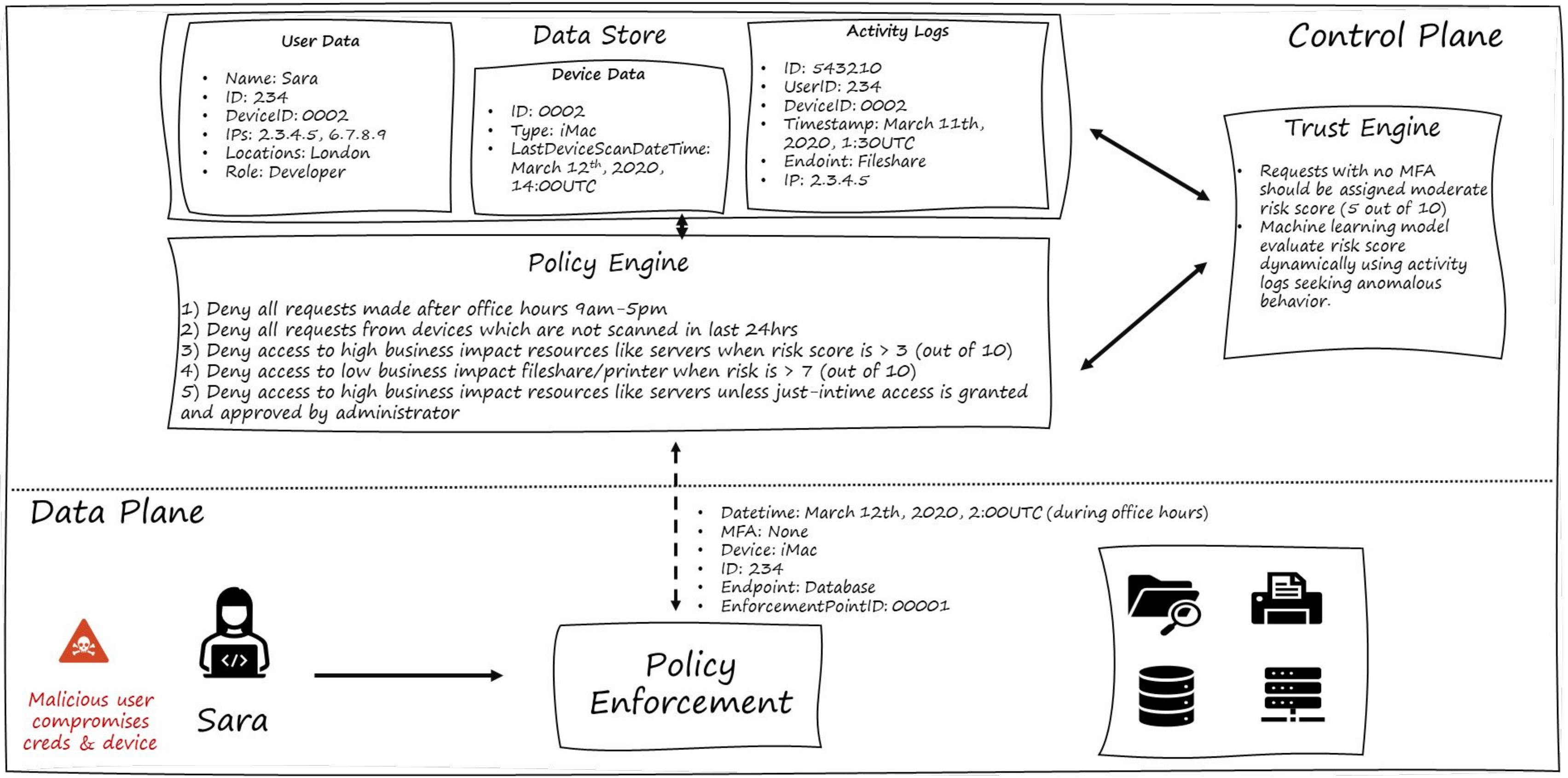
Scenario Walkthrough | Sara's creds & device is compromised



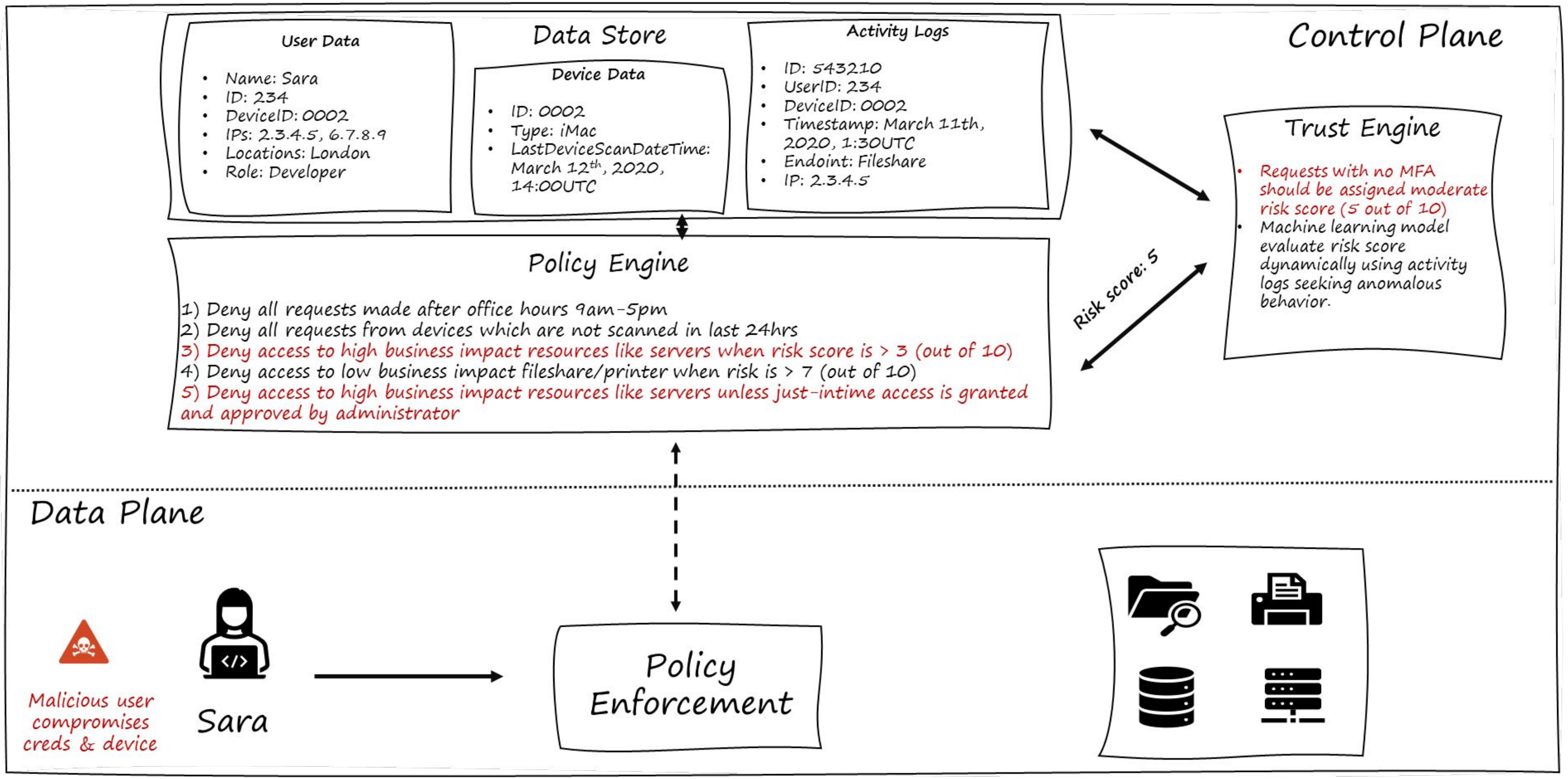
Scenario Walkthrough | Sara's creds & device is compromised



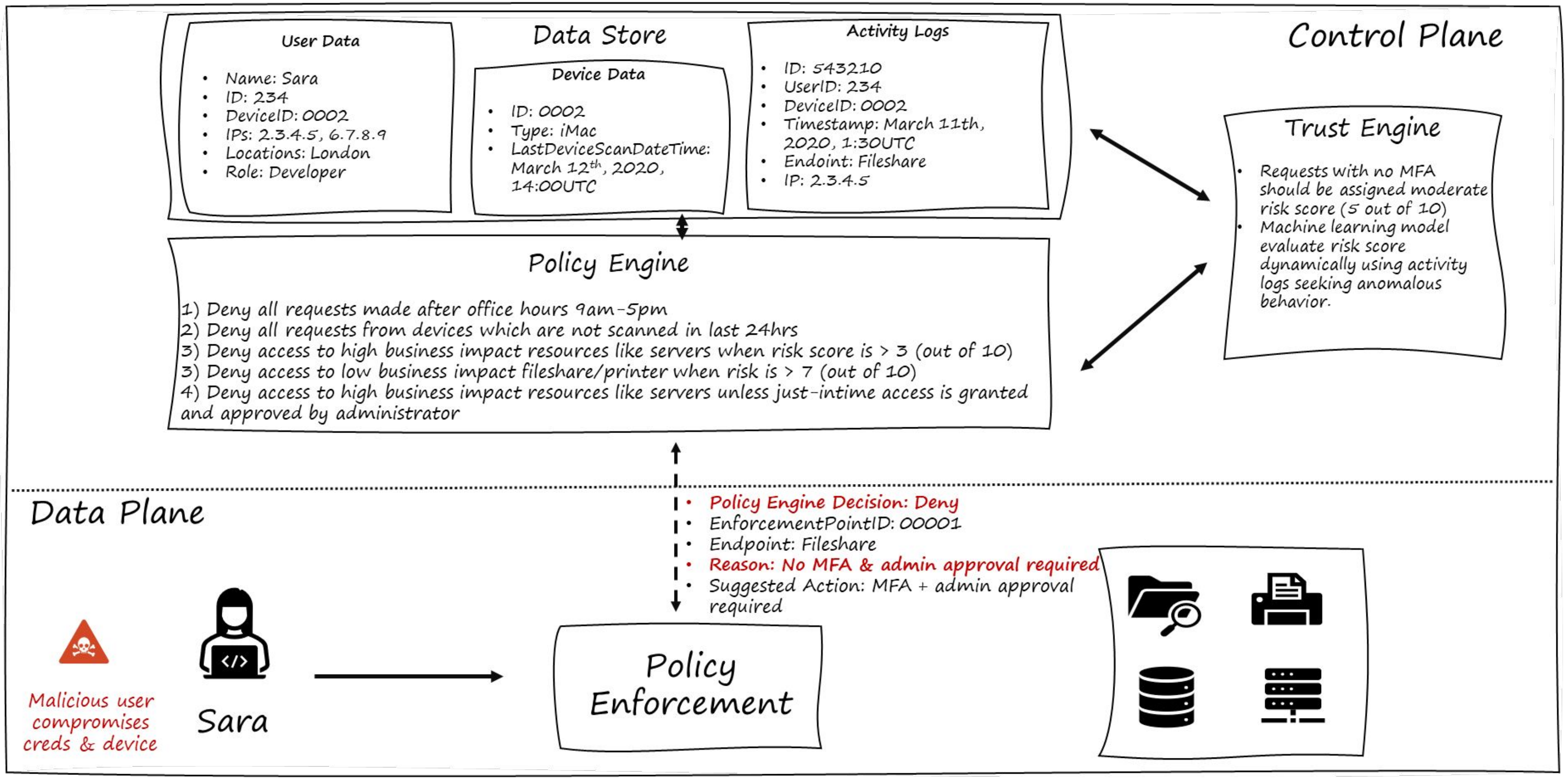
Scenario Walkthrough | Sara’s creds & device is compromised



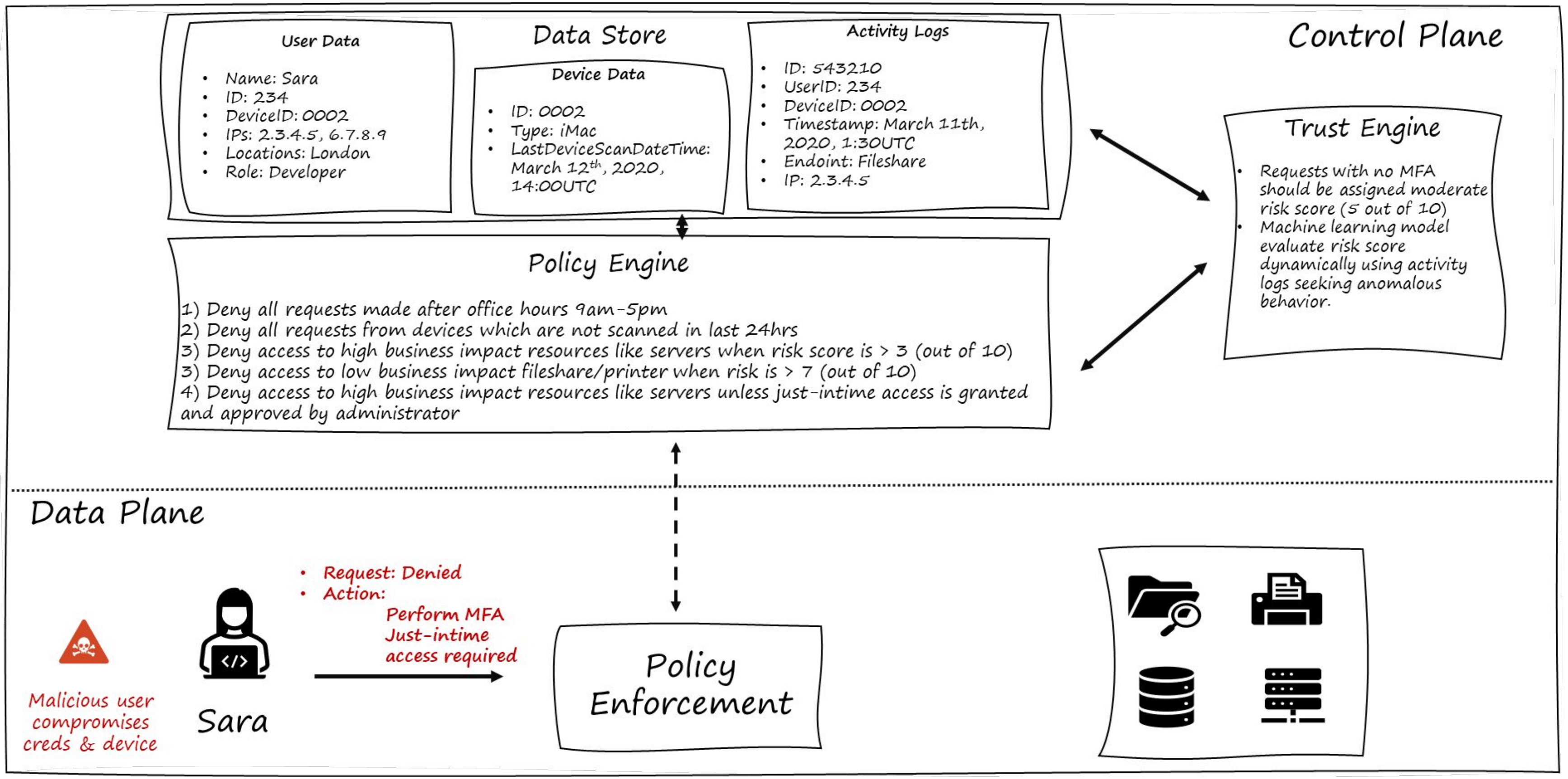
Scenario Walkthrough | Sara's creds & device is compromised



Scenario Walkthrough | Sara's creds & device is compromised



Scenario Walkthrough | Sara’s creds & device is compromised



Zero Trust is a journey

- Zero Trust requires maturity across various practices: governance, automation, analytics, identity, device & change management, data, applications and network.
- Zero Trust requires mindset shift
- Zero Trust requires cross pollination between teams which means culture & process change.
- What Zero Trust is NOT:
    - A security solution that you can buy from a vendor.
    - A single solution fit all model

Poll

Which Zero Trust security control is most important to your organization? Choose the one that best reflects your priorities.

Choices:

- Identity security and access controls
- Device security and posture management
- Network/environment segmentation and protection
- Data security and monitoring of sensitive assets

Discussion Question

In your own words, what does *“never trust, always verify”* mean? Can you share an example (from work or industry) where you think this mindset prevented a risk or could have mitigated one?

Break

Poll

Which of these organizational guidance frameworks do you most commonly use in your day-to-day work? Choose the one that best aligns with your practice.

Choices:

- NIST
- OWASP
- MITRE
- CSA
- None of the above

Zero Trust Security | Frameworks

Organization	Country/Region	Organization Type	Artifact
National Institute of Standards & Technology (NIST)	United States	Govt	Zero Trust Architecture (NIST SP 800-207)
National Cybersecurity Center of Excellence (NCCoE) / NIST	United States	Govt	Implementing a Zero Trust Architecture
National Cybersecurity Center of Excellence (NCCoE) / NIST	United States	Govt	NIST Cybersecurity Practice Guide SP 1800-35 Vol C–D (Implementing a Zero Trust Architecture) - Draft
CISA Zero Trust Maturity Model	United States	Govt	Zero Trust Maturity Model
Cloud Security Alliance (CSA)	United States	Not-for-Profit	Software-Defined Perimeter (SDP) and Zero Trust
Department of Defense	United States	Govt	Department of Defense (DoD) Zero Trust Reference Architecture
National Security Agency	United States	Govt	Embracing a Zero Trust Security Model
National Cyber Security Centre	United Kingdom	Govt	Zero trust architecture design principles

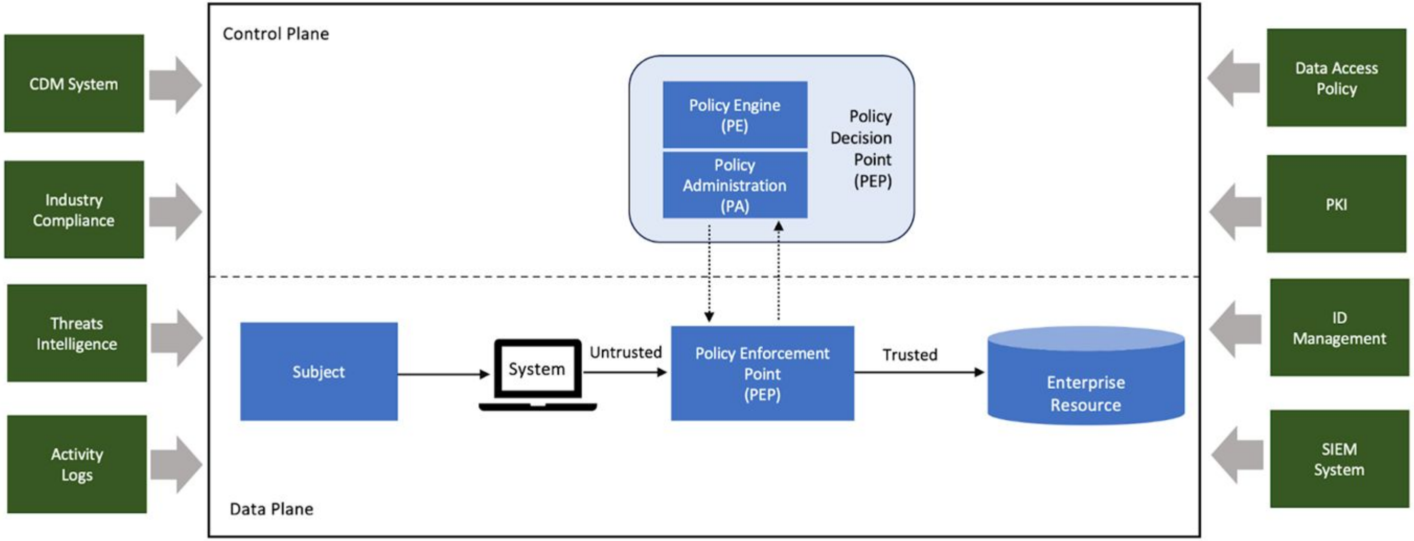
Zero Trust Security | Frameworks

Organization	Country/Region	Organization Type	Artifact
National Institute of Standards & Technology (NIST)	United States	Govt	Zero Trust Architecture (NIST SP 800-207)
National Cybersecurity Center of Excellence (NCCoE) / NIST	United States	Govt	Implementing a Zero Trust Architecture
National Cybersecurity Center of Excellence (NCCoE) / NIST	United States	Govt	NIST Cybersecurity Practice Guide SP 1800-35 Vol C–D (Implementing a Zero Trust Architecture) - Draft
CISA Zero Trust Maturity Model	United States	Govt	Zero Trust Maturity Model
Cloud Security Alliance (CSA)	United States	Not-for-Profit	Software-Defined Perimeter (SDP) and Zero Trust
Department of Defense	United States	Govt	Department of Defense (DoD) Zero Trust Reference Architecture
National Security Agency	United States	Govt	Embracing a Zero Trust Security Model
National Cyber Security Centre	United Kingdom	Govt	Zero trust architecture design principles

Zero Trust Architecture (NIST SP 800-207)

- **Zero trust** provides a collection of concepts and ideas designed to minimize uncertainty in enforcing **accurate, least privilege per-request access decisions** in information systems and services in the face of a network viewed as compromised.

- **Zero Trust Architecture (ZTA)** is an enterprise's cybersecurity plan that uses zero trust concepts and encompasses component relationships, workflow planning, and access policies.



Zero Trust Architecture (NIST SP 800-207)

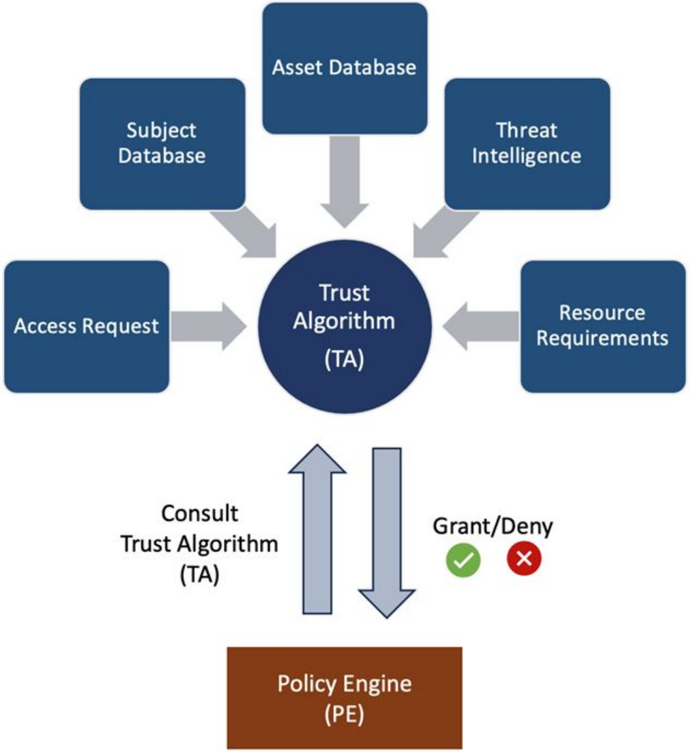
Access request: This is the access request from the subject.

Subject database: Database containing attributes related to the identity of the subject including but not limited to PII, geo-location, entitlements etc.

Asset database: This is the asset inventory database, which contains the known state of every enterprise-owned and/or non-enterprise/BYOD of assets including but not limited to devices, virtual machines etc.

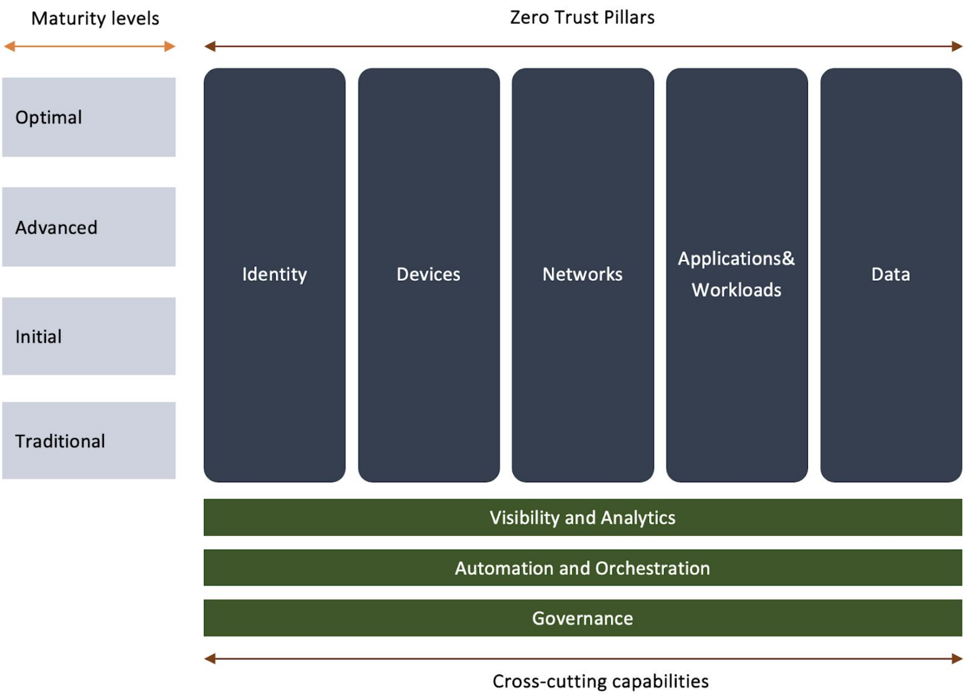
Resource requirements: This includes enterprise rules/policies aligned with business processes and compliance requirements, such as restricting access based on time/day, geo-location, and requiring higher authentication (e.g., MFA) based on the data sensitivity/criticality of the resource.

Threat intelligence: This includes feeds on the most recent threats compiled from various Internet sources, such as the dark web, malware and vulnerability indexes, etc.



Zero Trust Architecture (NIST SP 800-207)

- **Traditional** – Manual configurations and assignment of attributes, static security policies.
- **Initial**: At this maturity level, the lifecycle management of identities, assets, and resources has a basic level of automation
- **Advanced** – Some cross-pillar coordination, centralized visibility, centralized identity control, policy enforcement based on cross-pillar inputs and outputs.
- **Optimal** – Fully automated assigning of attributes to assets and resources, dynamic policies based on automated/observed triggers.



Zero Trust Ecosystem

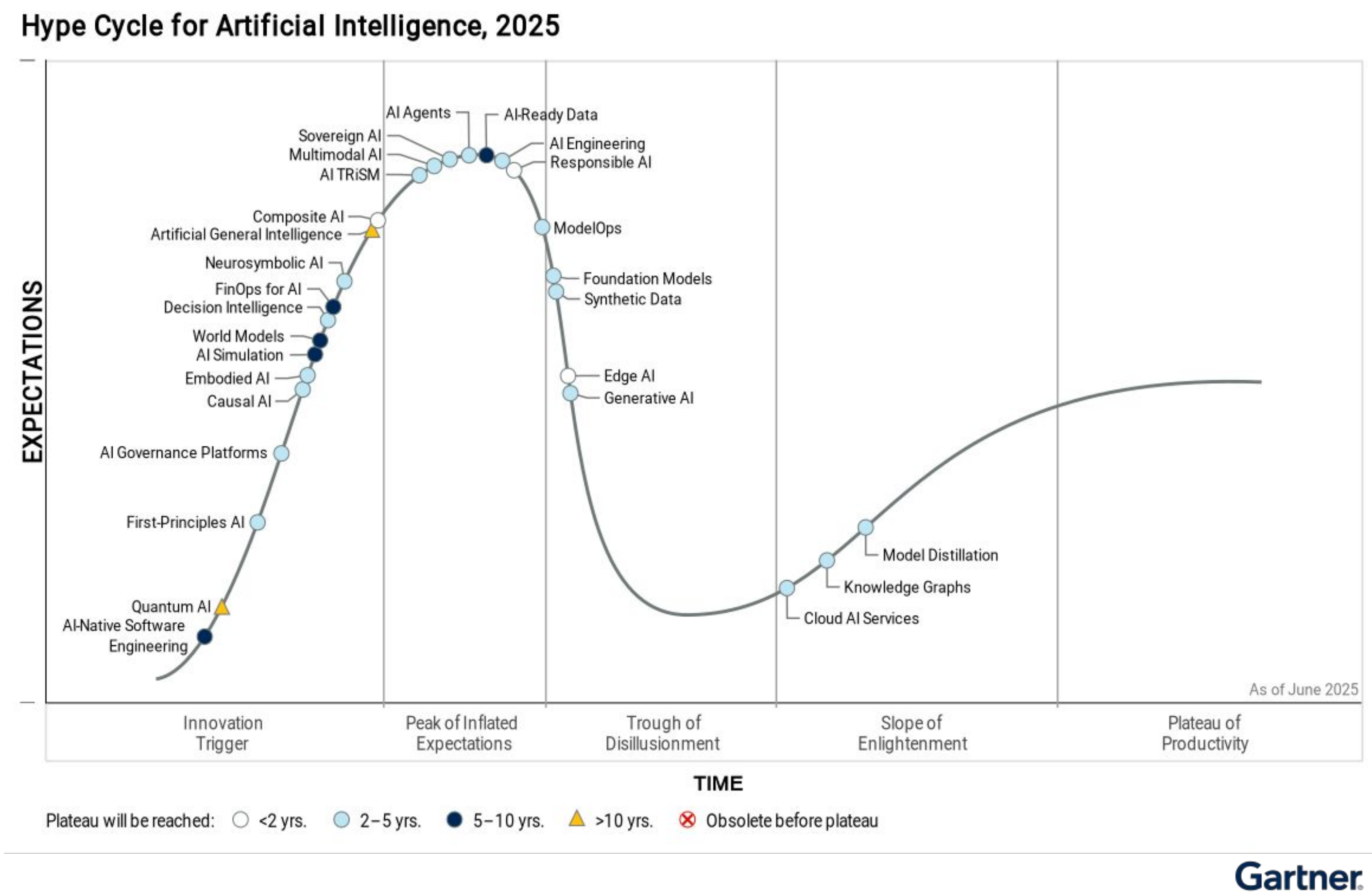
- **Forrester's Zero Trust eXtended (ZTX) Ecosystem:** This framework is described as a security architecture and operations playbook https://www.forrester.com/report/the-zero-trust-extended-ztx-ecosystem/RES137210 (behind paywall)

- **Gartner's Secure Access Service Edge (SASE):** SASE is a new package of technologies including SD-WAN, SWG, CASB, ZTNA and FWaaS as core abilities. https://blogs.gartner.com/andrew-lerner/2019/12/23/say-hello-sase-secure-access-service-edge/

- CISCO, Microsoft ,Google, Amazon, IBM, Oracle, vmware, and many more.

AI and Zero Trust

Gartner: AI Hyper Cycle 2025



Enterprise: AI Security and Zero Trust

- Currently, AI in the enterprise is primarily focused on Generative AI, particularly large language models (LLMs).
- Copilots and chatbots are built on LLMs.
- AI agents enable semi- or fully automated workflows, though enterprise adoption is still in the early stages.
- Fear of Missing Out (FOMO) is real in AI. Many CxOs are asking: *Can AI agents deliver 10x productivity?*

AI as an enabler for Zero Trust security

- AI powered device profiling
- AI powered behavioral analytics
- AI powered anomaly detection
- AI powered policy optimization

Fundamental challenge with LLM based Agents

- LLMs are inherently **non-deterministic**.
- Traditional software systems produce **predictable outputs** from defined inputs.
- LLM responses and decisions are shaped by probabilistic models, prompts, and internal state, making the **same input** capable of producing **different outputs** over time.
- This is **by-design**.
- It’s a **feature** not a **bug**.

OWASP Top 10 for LLM Applications 2025

OWASP | GENAI SECURITY PROJECT - 2025 TOP 10 LIST FOR LLMs AND GEN AI

genai.owasp.org/llm-top-10/

2025 OWASP Top 10 List for LLM and Gen AI

LLM01:25

Prompt Injection

This manipulates a large language model (LLM) through crafty inputs, causing unintended actions by the LLM. Direct injections overwrite system prompts, while indirect ones manipulate inputs from external sources.

LLM02:25

Sensitive Information Disclosure

Sensitive info in LLMs includes PII, financial, health, business, security, and legal data. Proprietary models face risks with unique training methods and source code, critical in closed or foundation models.

LLM03:25

Supply Chain

LLM supply chains face risks in training data, models, and platforms, causing bias, breaches, or failures. Unlike traditional software, ML risks include third-party pre-trained models and data vulnerabilities.

LLM04:25

Data and Model Poisoning

Data poisoning manipulates pre-training, fine-tuning, or embedding data, causing vulnerabilities, biases, or backdoors. Risks include degraded performance, harmful outputs, toxic content, and compromised downstream systems.

LLM05:25

Improper Output Handling

Improper Output Handling involves inadequate validation of LLM outputs before downstream use. Exploits include XSS, CSRF, SSRF, privilege escalation, or remote code execution, which differs from Overreliance.

LLM06:25

Excessive Agency

LLM systems gain agency via extensions, tools, or plugins to act on prompts. Agents dynamically choose extensions and make repeated LLM calls, using prior outputs to guide subsequent actions for dynamic task execution.

LLM07:25

System Prompt Leakage

System prompt leakage occurs when sensitive info in LLM prompts is unintentionally exposed, enabling attackers to exploit secrets. These prompts guide model behavior but can unintentionally reveal critical data.

LLM08:25

Vector and Embedding Weaknesses

Vectors and embeddings vulnerabilities in RAG with LLMs allow exploits via weak generation, storage, or retrieval. These can inject harmful content, manipulate outputs, or expose sensitive data, posing significant security risks.

LLM09:25

Misinformation

LLM misinformation occurs when false but credible outputs mislead users, risking security breaches, reputational harm, and legal liability, making it a critical vulnerability for reliant applications.

LLM10:25

Unbounded Consumption

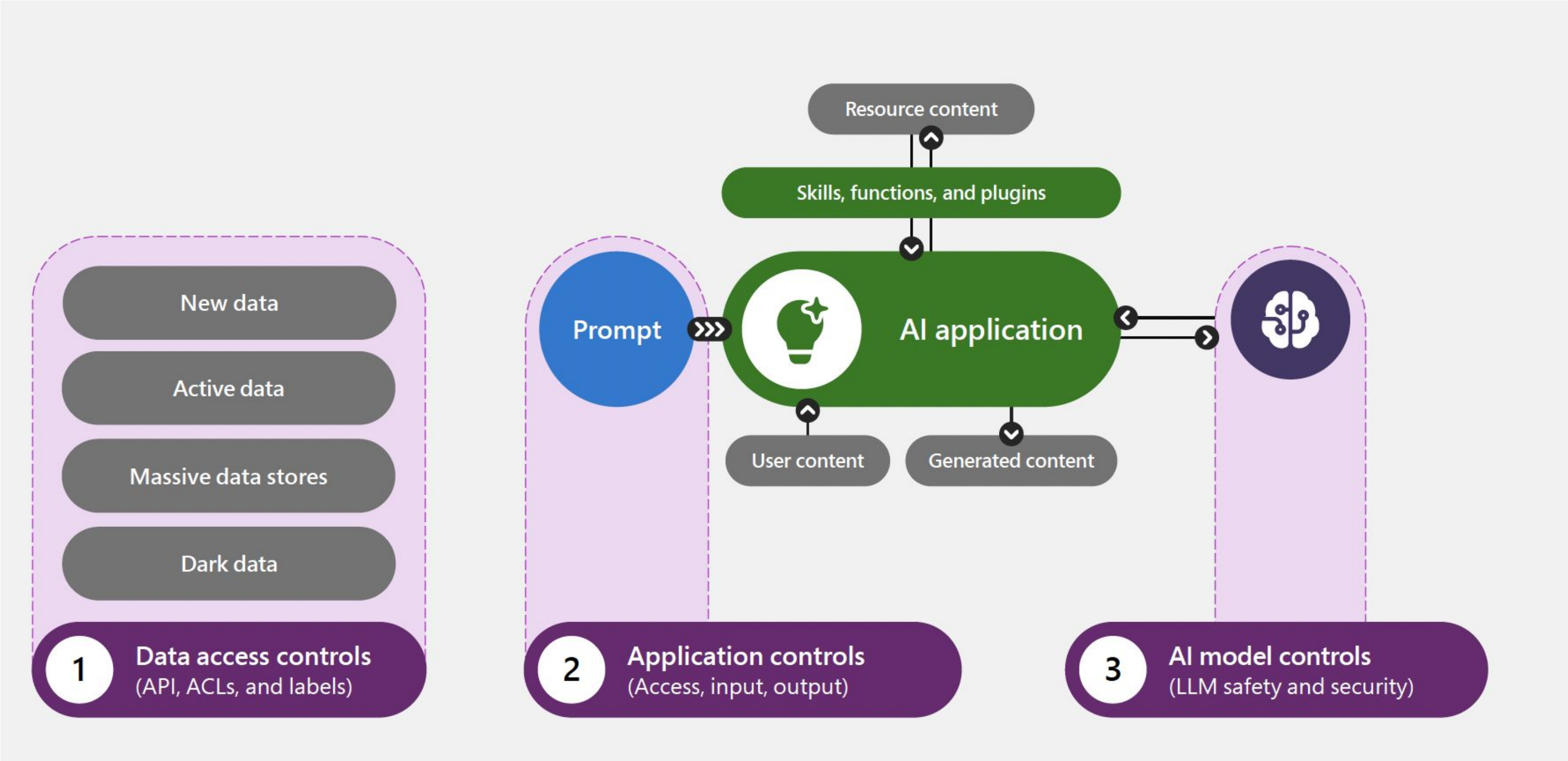
Unbounded Consumption occurs when LLMs generate outputs from inputs, relying on inference to apply learned patterns and knowledge for relevant responses or predictions, making it a key function of LLMs.

genai.owasp.org

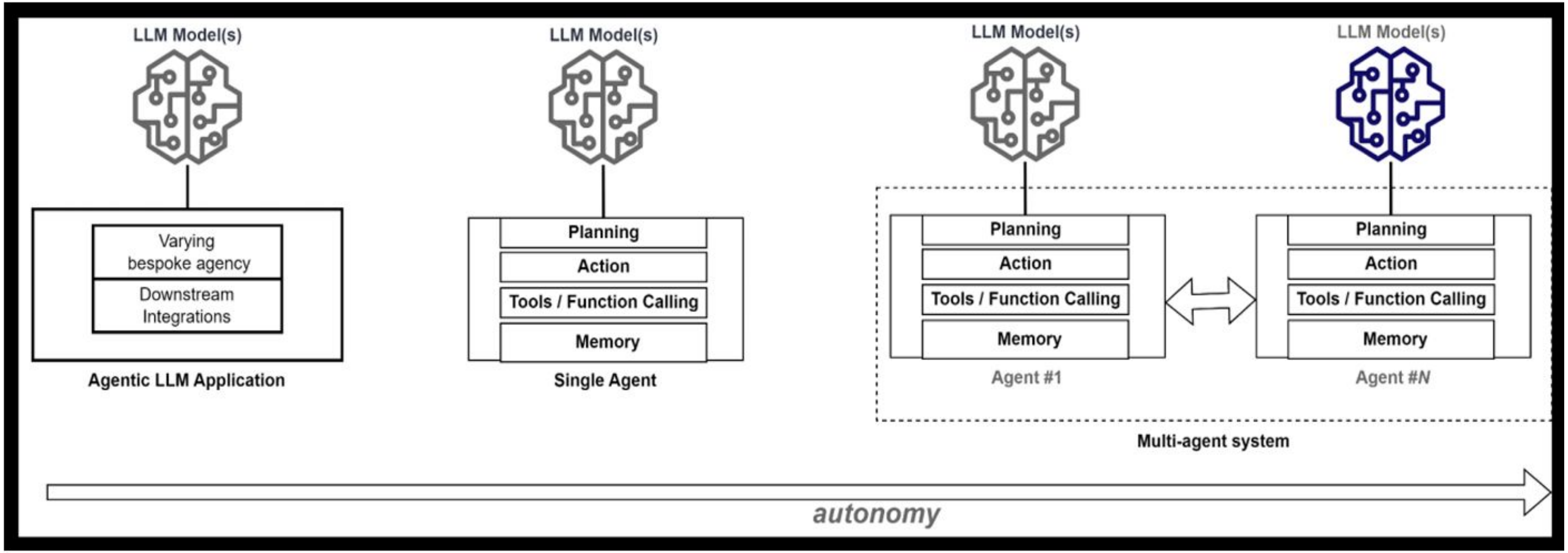
MITRE ATLAS

Reconnaissance 5 techniques	Resource Development 7 techniques	Initial Access 6 techniques	ML Model Access 4 techniques	Execution 3 techniques	Persistence 3 techniques	Privilege Escalation 3 techniques	Defense Evasion 3 techniques	Credential Access 1 technique	Discovery 4 techniques	Collection 3 techniques	ML Attack Staging 4 techniques
Search for Victim's Publicly Available Research Materials	Acquire Public ML Artifacts	ML Supply Chain Compromise	ML Model Inference API Access	User Execution	Poison Training Data	LLM Prompt Injection	Evade ML Model	Unsecured Credentials	Discover ML Model Ontology	ML Artifact Collection	Create Proxy ML Model
Search for Publicly Available Adversarial Vulnerability Analysis	Obtain Capabilities	Valid Accounts	ML-Enabled Product or Service	Command and Scripting Interpreter	Backdoor ML Model	LLM Plugin Compromise	LLM Prompt Injection		Discover ML Model Family	Data from Information Repositories	Backdoor ML Model
Search Victim-Owned Websites	Develop Capabilities	Evade ML Model	Physical Environment Access	LLM Plugin Compromise	LLM Prompt Injection	LLM Jailbreak	LLM Jailbreak		Discover ML Artifacts	Data from Local System	Verify Attack
Search Application Repositories	Acquire Infrastructure	Exploit Public-Facing Application	Full ML Model Access						LLM Meta Prompt Extraction		Craft Adversarial Data
Active Scanning	Publish Poisoned Datasets	LLM Prompt Injection									
	Poison Training Data	Phishing									
	Establish Accounts										

AI Security and Zero Trust



CSA : AI Threats and Mitigations



Poll

What’s the biggest impact of AI on Zero Trust security in your view? Choose the one that best aligns with your thinking.

Choices:

- AI as an enabler that strengthens and accelerates Zero Trust
- AI as a challenge that adds new risks and complexity
- AI with minimal impact on core Zero Trust principles
- AI as both a catalyst for progress and a source of new threats

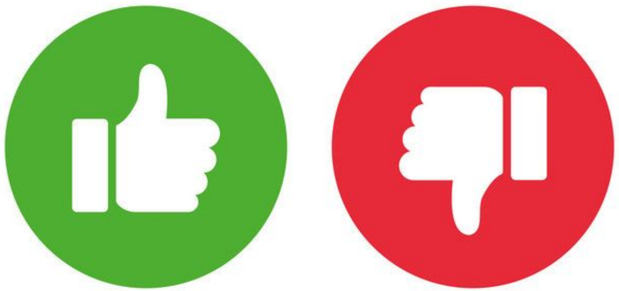
Q&A

Break

Exercise: Mapping AI Threats to Zero Trust Pillars

Pulse Check

Do you believe Zero Trust will benefit from AI-driven security controls?

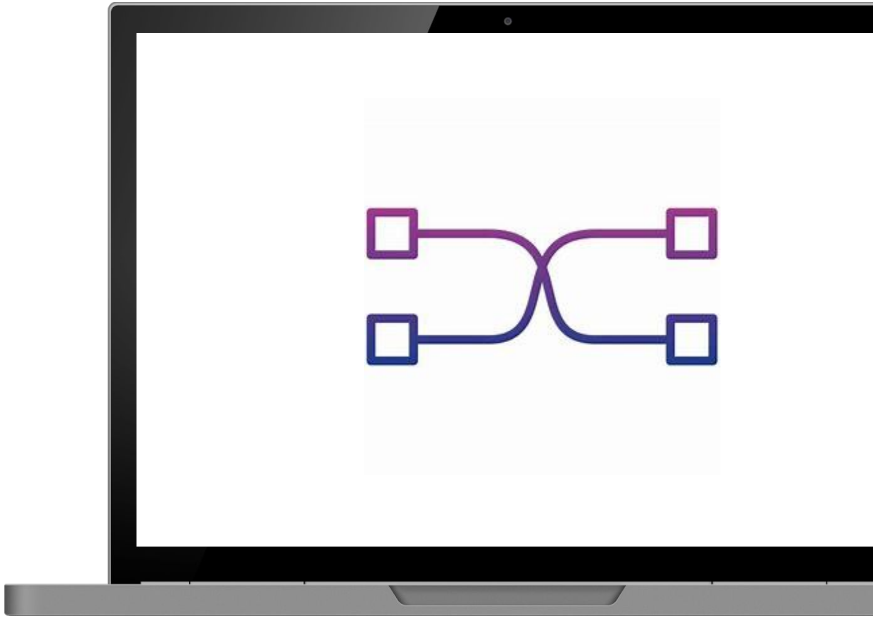


Exercise

Zero Trust is about never trust, always verify across identity, devices, applications, data, and networks. AI brings both opportunities and new risks.

In this exercise, we'll map key AI threats to Zero Trust pillars and see where controls need to evolve.

Let's get started!



AI Threats

Deepfake identity spoofing: AI-generated voices/faces bypassing MFA/biometrics

Poisoned training data: Malicious data inserted into ML pipelines

Prompt injection/jailbreaks: Adversarial inputs manipulating LLM behavior

Model inversion/extraction: Attackers leaking sensitive training data or stealing models

AI-driven lateral movement: Automated reconnaissance and pivoting across systems

Shadow AI services: Unapproved AI SaaS apps & un-managed endpoints outside IT

Zero Trust Pillars

Identity	Device	Application/ Workload	Data	Network/ Environment

Mapping: AI Threats x Zero Trust Pillars

AI Threat	Primary Pillar(s)	Why Primary	Secondary Pillars	Key Zero Trust Controls
Deepfake identity spoofing	**Identity**	Direct attack on authentication and proofing systems (MFA, biometrics, liveness).	Device (if bound to hardware tokens); Application (if app-specific auth flows).	Phishing-resistant MFA (FIDO2/WebAuthn), behavioral analytics, continuous authentication, liveness detection.
Shadow AI (local endpoints & unapproved SaaS)	**Device + Application/Workload**	Endpoint-level shadow AI (desktop wrappers, browser plug-ins) → Device issue. Use of external SaaS AI without approval → Application/Workload issue.	Data (sensitive uploads to AI); Network (egress paths to AI).	Device attestation, extension allow-lists, SaaS discovery, CASB, tenant allow-lists, DLP/egress monitoring.
Prompt injection / jailbreaks	**Application/Workload**	Exploits application logic (LLM context injection, tool-use boundaries).	Data (output leakage); Network (tool invocation isolation).	Context isolation, signed prompts, I/O filtering, workload guardrails, least-privilege tool scoping.

Mapping: AI Threats x Zero Trust Pillars (Cont.)

AI Threat	Primary Pillar(s)	Why Primary	Secondary Pillars	Key Zero Trust Controls
Poisoned training data	Application/Workload	Compromises model integrity in ML training pipelines (supply chain).	Data (dataset governance, quality control).	Dataset provenance, model signing, reproducible training, SBOMs for models, CI/CD policy checks.
Model inversion / extraction	Data	Goal is exfiltration of sensitive data or IP embedded in models.	Application (attack surface = APIs), Network (rate limiting, geo fencing), Identity (API authN/Z).	Differential privacy, encryption, strict API authZ, query throttling, anomaly detection, model artifact protection.
AI-driven lateral movement	Network	AI automates reconnaissance and pivoting; blast radius defined by segmentation.	Identity (stolen tokens abused), Device (unmanaged footholds).	Identity-based micro-segmentation, mTLS/service identity, east–west traffic analytics, adaptive policy updates.

Poll

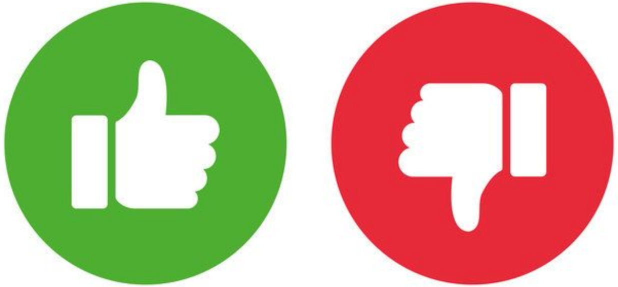
What is your organization’s biggest concern with AI? Select the one that aligns best with your view.

Choices:

- Security and privacy risks (data leakage, model theft, prompt injection)
- Reliability and safety (hallucinations, lack of explainability, control)
- Compliance and governance (regulations, auditability, accountability)
- Scaling and operations (integration, cost, performance, vendor lock-in)

Pulse Check

Do you believe static Zero Trust controls able to stop AI threats?



Case Study: AI real world use cases that enhance Zero Trust security

Athena: AI-Driven Financial Insights Platform

Multinational fintech is about to launch Athena, an AI-driven financial insights platform.

The stack spans hybrid cloud, SaaS, and internal systems; teams include global developers, data scientists, and executives. Sensitive IP, customer PII, model weights, embeddings which live across managed cloud, object stores, and modern developer tooling.

Let’s take a closer look!

Athena

- **Table stakes:** The platform contains proprietary machine learning models, customer PII, regulatory-sensitive financial data, and intellectual property.

- **Challenge:** To accelerate productivity, the company has rolled out **GenAI copilots** and **internal RAG chatbots**. Attackers know the value of this data and have launched **AI-powered phishing, deepfakes, adversarial ML, and insider recruitment campaigns**. Adversaries respond with **AI-enhanced social engineering (multi-language spear-phishing, deepfakes), model-aware malware, and data poisoning** moving at machine speed and exploiting the mismatch between **static controls** and **dynamic AI traffic.**

- **Goal:** Implement AI powered Zero Trust controls across all NIST pillars (Identity, Device, Network, Applications, and Data), with cross-pillar governance and resilience against AI-based threats by using AI itself.

Identity: Risk Based Adaptive Access

- Identity verification must move from one-time checks to **continuous, behavior-aware verification**. Every session and micro-action is evaluated against learned baselines (“never trust, always verify”), and **least-privilege/JIT** is adjusted in real time when risk changes.

- **Zero Trust Security Controls without AI:** A spear-phish with a deepfake CEO video tricks a developer into disclosing creds. Because the attacker uses the victim’s enrolled device and valid MFA, the 2 AM login from a new ASN looks “compliant.” Static policy (MFA + posture = allow) grants full repo access. Source code is exfiltrated quietly—no rule explicitly captured the **time/geo/sequence** abnormality.

- **AI-enhanced Zero Trust Security Controls:** UEBA compares the login and subsequent repo sequence to the developer’s baseline (time, location, resource mix) in real time to calculate risk in real time. The system **steps up to phishing-resistant MFA and downgrades to read-only**, generating a contextual SOC alert (“off-hours, new ASN, rare repo chain”). The exfiltration path is shut before data leaves. Programmatically, this aligns with verify-explicitly + least-privilege at **per-request** granularity.

Device: Continuous Device Trust & Compromise Prediction

- AI continuously recalculates **device trust** using EDR/posture drift, process baselines, and outbound patterns, so access decisions reflect **current** risk rather than a morning compliance snapshot.

- **Zero Trust Security Controls without AI:** A data scientist's laptop passes the 9 AM posture check (encrypted, patched). At noon, a poisoned Python package spawns an odd process tree and new egress domains. Because posture is only re-checked periodically, the device retains **write** rights to data lakes and model stores, enabling stealthy siphoning.

- **AI-enhanced Zero Trust Security Controls:** Real-time analytics flag rare child processes and anomalous egress. The device trust score drops; **access is downgraded to read-only** and the EDR quarantines. Subsequent requests inherit lower trust, blocking dataset pulls and halting persistence. This is the device-plane execution of **assume breach** + adaptive access.

Network: Micro-Segmentation & Lateral Movement Defense

- Enterprise RAG integrates Zero Trust by enforcing document-level ACLs, citations, and up-to-date retrieval; RAFT improves grounding. Pure LLMs can hallucinate and ignore permissions, but RAG adds source-bound answers with doc security.

- **Zero Trust Security Controls without AI:** An employee asks the chatbot, “What are executive salaries?” The LLM pulls HR sheets from a vector DB and answers, because retrieval isn’t permission-checked and the model doesn’t know about RBAC/ ACLs or provenance.

- **AI-enhanced Zero Trust Security Controls:** The RAG layer filters by document ACL and labels before retrieval; the generator includes citations. If unauthorized, the bot returns a policy-aware denial (“no authorized source for your identity”). With RAFT, grounded results improve while minimizing hallucinations—document-level security is preserved.

Data — Intelligent Discovery, Classification & Protection

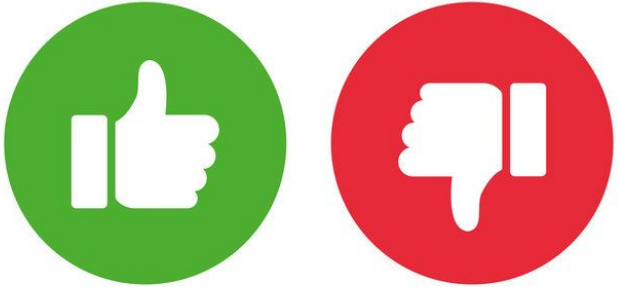
- AI enables **data-centric Zero Trust**: it discovers and classifies sensitive content (structured/unstructured) across SaaS and cloud, then **enforces labels** (encryption, DLP, conditional sharing) that **follow the data**, essential because **AI traffic is encrypted/dynamic** and eludes perimeter tools.

- **Zero Trust Security Controls without AI:** An engineer copies a confidential valuation model to a personal folder and shares it externally. Pattern/regex DLP misses it, and the model escapes governance no lineage, no audit trail, no containment.

- **AI-enhanced Zero Trust Security Controls:** Semantic classification tags the file as **Confidential IP**, auto-enables encryption, and **blocks external sharing**. The system emits a full lineage timeline (creator → movements → attempted exfil). This reflects the Microsoft guidance to make ZT **data-centric** for AI age risks.

Pulse Check

Do you now see how AI-enhanced Zero Trust closes the gaps in static defenses?



Q&A

Break

Takeaways

- AI is simultaneously a headwind and a tailwind for Zero Trust
- Upcoming AI Courses:
  - Live Training Courses (AI + Cybersecurity)
- Don't forget to check GitHub for resources: https://github.com/razi-rais/Zero-Trust-for-AI-Systems
- Feel free to connect and continue the conversation:
  - LinkedIn: https://www.linkedin.com/in/razirais
  - Meet: https://calendly.com/razi-rais/30min

Q&A

O'REILLY®