A Deep-Dive Into the Aembit Workload IAM Platform

This action-oriented guide covers all the essentials on securing and streamlining your workload and non-human identities with Aembit.

aembit

Gartner COOL VENDOR 2022

CSO Cybersecurity startups to watch for in 2023

2024 WINNER CYBER SECURITY EXCELLENCE AWARDS

RSAC Innovation Sandbox 2024 FINALIST

2024 SINET16 INNOVATOR AWARD

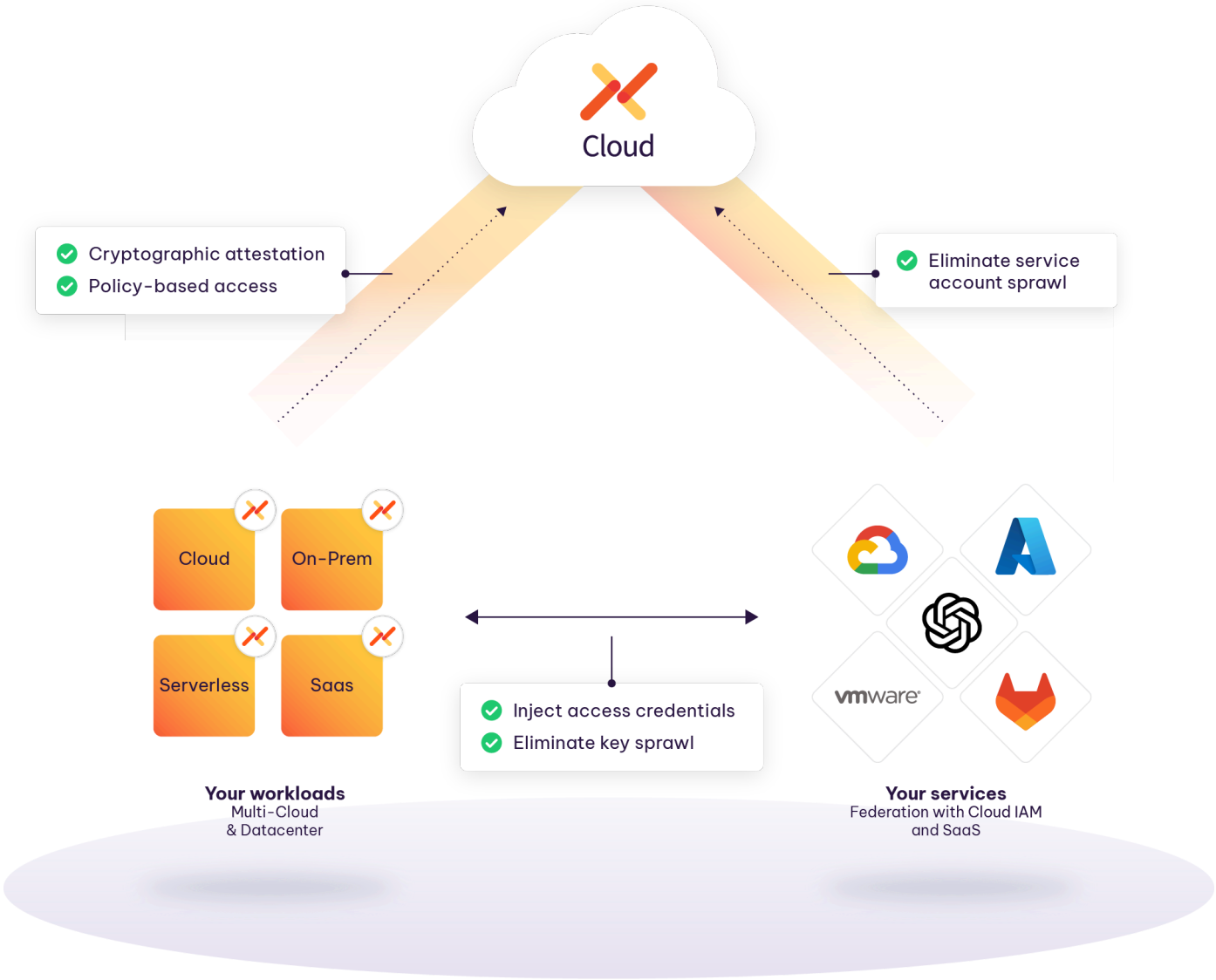
Introduction: A Unified Approach to Non-Human Identity and Access Management

Aembit has built the industry's first workload IAM platform to safely allow your non-human identities (NHIs) to connect to any services, applications, or APIs they need. Based on cryptographically verifiable identities instead of simply possessing long-lived secrets, we give you a simple, secure, and central way to provide policy-based access, instead of needing to manage low-level secrets.

Enterprises, industry analysts, and standards organizations are focusing on non-human identity access as the number of these identities, relative to human identities, continues to rise. The adoption of cloud, microservices, and SaaS has created the trend, and the coming wave of Agentic AI is set to make it explode. And, while zero trust focused on users at its introduction, this trend is making it essential for non-human identities as well.

As we will cover, Aembit not only provides IAM for non-human identities, Aembit also extends zero trust principles to NHIs by providing conditional access both natively and using 3rd party integrations, along with enabling MFA-like capabilities even when the services used don't support them. This allows enterprises to have consistency between their human and non-human identity and access policies to better meet compliance requirements.

Essentially, you can think of us as Okta, but for workloads, machine and non-human identities. Instead of user-to-workload access, Aembit manages workload-to-workload access.



With Aembit, organizations can implement identity-first security for their workloads in the most complex of situations: where applications and services need to reach across boundaries and establish trust through identity. Moreover, security professionals do not simply want another "dashboard report." They instead want something that proactively automates a solution at scale. With Aembit, enterprises help their organizations in three key ways:

1

Stop Workload Attacks

Defend against common workload communication threats, including credential exposure, unauthorized and over-privileged access, lack of key rotation, service account sprawl, and weak or misconfigured authentication, to ensure robust security in your distributed application infrastructure.

2

Build Great Products, Make Developers Happier

Aembit saves hundreds to thousands of developer hours by automating the heavy lifting of workload IAM and credential management, while ensuring a consistent auth implementation across your entire environment.

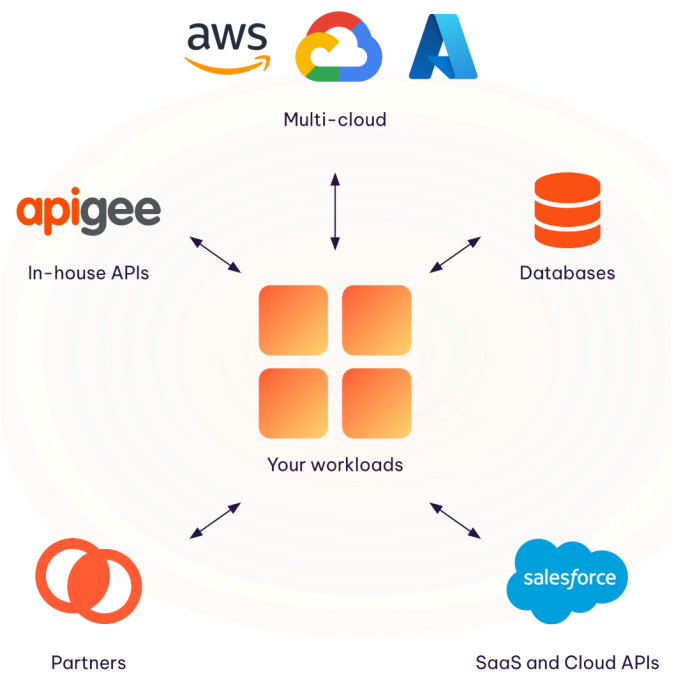
3

Scale DevSecOps

Save your DevSecOps teams 50% or more of the time they currently spend on workload-to-workload security. Automate and optimize activities, such as credential rotation, compliance, and audits. Aembit gives you a single place to operate, control, and analyze workload-to-workload access using UI, APIs, and Terraform.

Enterprise Context and Today's Approaches

Modern applications are highly distributed, with applications or microservices leveraging many different resources to source data and delegate operations. Modern application access is also distributed between human users and non-human identities. Some estimates put the ratio of NHI to human users at 45:1 at most enterprises.



For example, you might be using services from multiple clouds, databases that are distributed in multiple locations, APIs from your own applications or other services, SaaS applications (like Salesforce or Snowflake), or even partner applications.

Against this backdrop, you can see an increasing attack surface with every new connection your application makes. Operational complexity is increasing for your teams to manage it all – conducting a regular audit or performing credential rotation, for example, can become a major project.

And secrets, the method used to connect applications today, are sprawled across applications, whether they are hard-coded or managed on a team-by-team basis.

Enterprises that have spent many years upleveling their local and remote access for users by implementing zero trust architectures and principles are now concerned about the security and access gaps for non-human identities and access. Multi-factor authentication, conditional access, and device posture checks have long been available for users. Creating a consistent identity and access policy for your human and non-human identities is possible, not just helping reduce the attack surface but also puts your organization in a better position to meet compliance and audit requirements.

We generally see three methods of managing access today:

Cloud Provider IAM

This works well for controlling access to native services at a single cloud provider, but breaks down once you start crossing boundaries into other clouds, SaaS services, or even your own on-prem software. Many of these Cloud Provider IAM solutions have legacy vaults that store static, long-lived keys. Some have Identity Federation services which have benefits in their own cloud but aren't built for hybrid and multi-cloud environments that enterprises are using today.

Vaults

They are good for storing secrets, but using them implicitly assumes that you understand the identity of the client requesting access. Further, they typically require effort and maintenance from your dev team to integrate into your applications and your DevOps team to manage and rotate keys.

DIY

Do-it-yourself approaches generally include a set of manual configurations, best practices, and 'allow' lists that don't scale well and can easily fall apart during an incident. The DIY solutions are also primarily focused on a single application while other solutions are purchased, configured, and deployed for the other SaaS services and software used.

The challenges for enterprises typically are seen in two ways:

1

Providing an identity and access layer that works across environments is particularly cumbersome.

If you are operating in one environment (for example, wholly in AWS or using identities within a confined Kubernetes cluster), there are native options that might allow you to achieve your goals. If, however, you are trying to cross boundaries (i.e. AWS to GCP, your cloud to SaaS, your on-prem environment to a partner's cloud), access management tools break down.

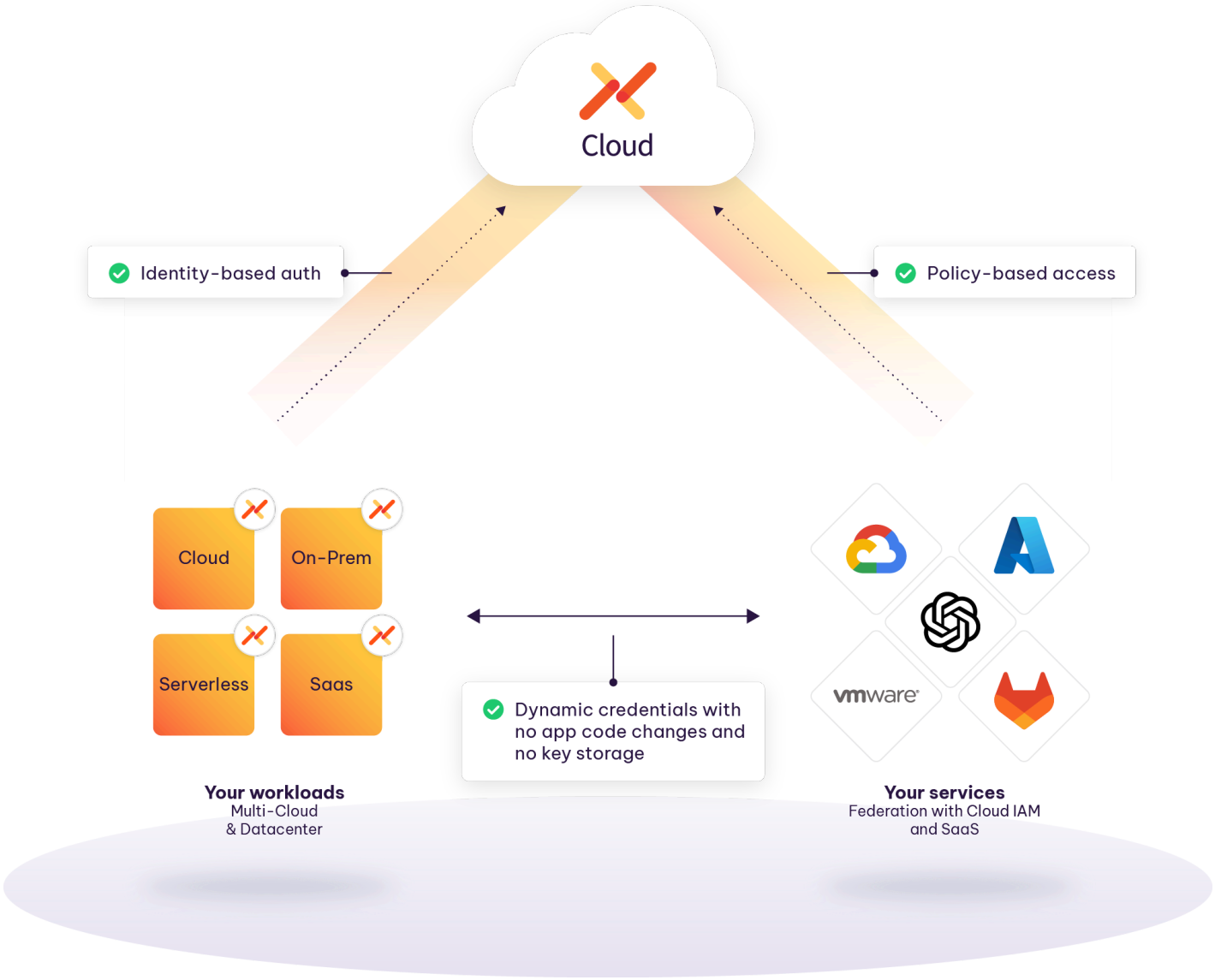
2

Enterprises are looking for ways to easily automate and scale workload identity and access management.

Even if you have clear policies for workload identity access, trying to do it at scale is not easy when you are using different methods and tools. Most of this is done manually today, which is a burden on your IT teams and still requires code changes, which is a burden on your developers. Validating access rights, storing and rotating keys, enforcing conditional access, and performing auditing are examples of tasks that aren't going away and should be automated.

Targeted Benefits of Workload IAM

At a broad level, the workload IAM architecture looks somewhat similar to your user IAM strategy: a third-party, independent broker that can validate identities and evaluate access policies.



Naturally, with workload IAM the workflows need to be different to work with what machines understand and can do to validate themselves. We'll get into that in more detail.

Workload IAM sits in the area between security, DevOps, and developers, and as a result provides, benefits to each group:

Security

- Enforce identity-based, conditional access policies centrally.
- Move applications to short-tokens or short-lived credentials.
- Perform identity-based logging for faster incident response, audit, compliance.

DevOps

- Limit service account exposure and eliminate secret sprawl.
- Automate credential rotation.
- Automate deployment, configuration, and management using APIs and Terraform.

Developers

- Avoid the need for code auth and storing secrets in config files.
- Maintain consistent auth implementation everywhere.
- Move to dynamic credentials with no app changes.

Aembit Architecture and Approach

We will begin with a high-level overview of the architecture. Following this, we will examine a detailed flowchart that illustrates the interaction between services during a request. Together, these elements provide a comprehensive understanding of how Aembit functions within your infrastructure.

Aembit is designed to be your workload identity provider (IdP). We broker between workload identity and service credentials, using policies to control access. We do this across trust boundaries, giving you a consistent implementation everywhere you operate.

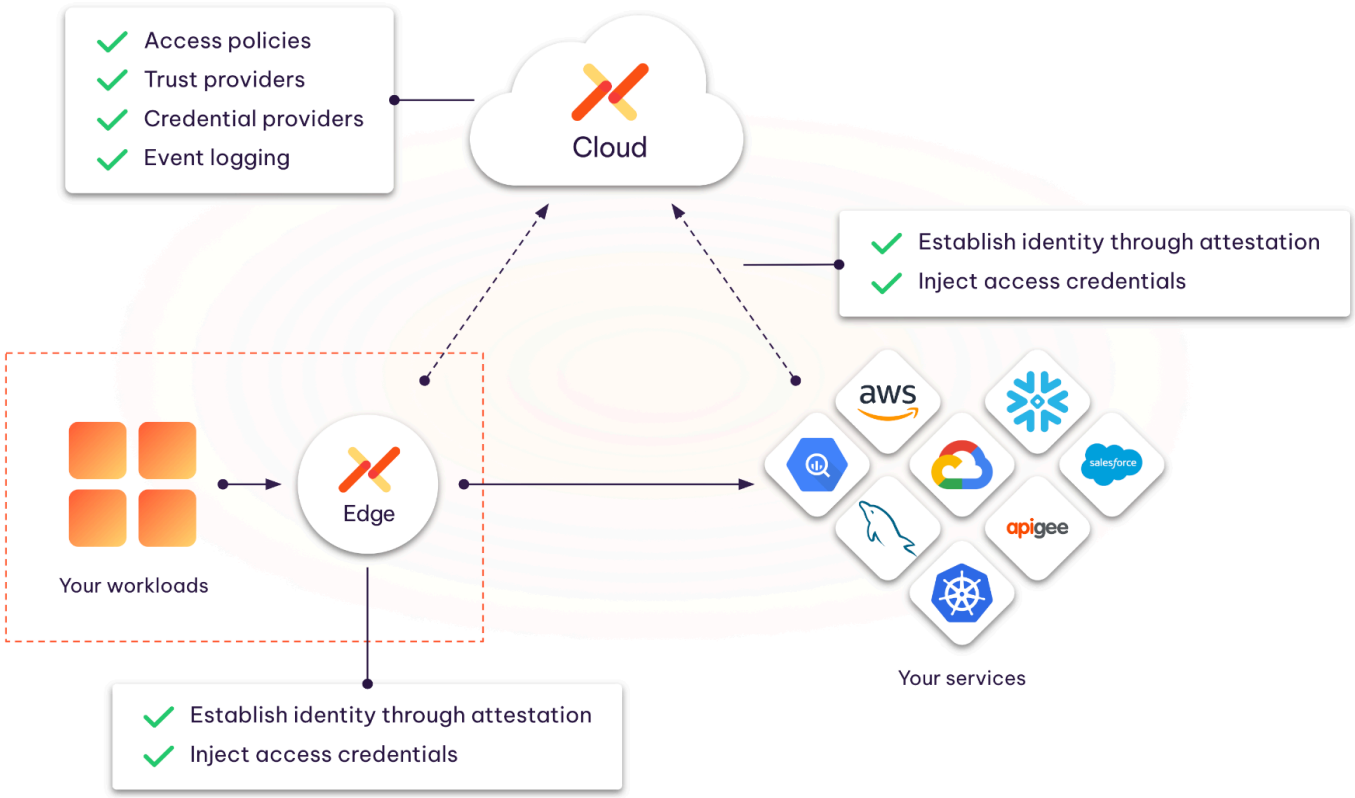
To enable deployments at scale, Aembit supports management by UI, APIs, and Terraform. Moreover, agent-based discovery quickly provides the building blocks, such as an inventory of server workloads, service accounts and more, needed to quickly and accurately create policies.

Essentially, we federate with both the workload and the service environments to make this happen. This works across the array of environments and services your business operates in today, with flexibility as you adapt and change.

> **“Workload IAM sits in the area between security, DevOps, and developers, and as a result provides benefits to each group.**

Aembit Architecture

- Centralized auth, policy-based access, and auditing
- 2-sided federation with short-lived credentials
- Cloud-based Control-plane (not data plane) architecture
- Aembit Edge deployed as a sidecar, agent, or serverless extension

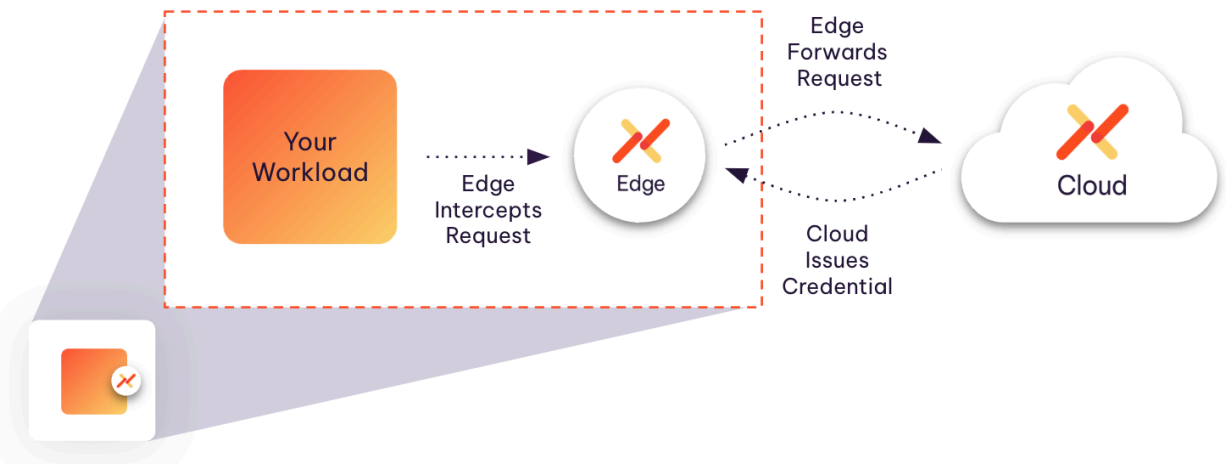


Imagine your workloads on the left – custom code that you write, or packaged applications that you run. They can be in the cloud or on-prem, running in Kubernetes, serverless environments like AWS Lambda or in VMs for example. On the right are services you access – APIs, SaaS applications, apps run by your business partners, or additional cloud services.

Up top, you see Aembit Cloud. Aembit Cloud is our SaaS service that provides a centralized approach to workload identity management, workload access policies, visibility, and logging. It is operated by Aembit as a SaaS platform. We built this with a control-plane architecture in mind: Aembit manages access, but doesn't see your sensitive application data.

On the left, Aembit Cloud federates with your workload environment through trust providers, which allow us to cryptographically validate the workload's identity using single or multi-factors.

We do this by leveraging the Aembit Edge. Aembit Edge can be deployed as a sidecar, agent, or serverless extension. Our host-based proxy is a transparent proxy that you deploy alongside each of your applications for which you'd like to manage workload identity. In Kubernetes environments, Aembit Edge is deployed as a sidecar, whereas in VM setups, it functions as an agent. Aembit Edge serverless extension is primarily used in serverless environments such as AWS Lambda or with SaaS applications such as GitLab. Aembit Edge both works to validate the identity of the workload and inject credentials on behalf of it.



Aembit Edge is what makes no-code auth possible. Instead of requiring your developers to modify code in their application, Aembit Edge intercepts authorization requests and injects credentials into validated requests. That means we'll work well for greenfield but also existing apps. We just slot in and can manage credentials without app changes.

The other big benefit here is that you reduce key sprawl. Workloads don't even need to store keys as Aembit Edge takes care of this dynamically for the app.

On the service side, Aembit Cloud federates with services through credential providers, which allow us to be trusted by the services you rely on to issue credentials that the client can use to access the service. Credential providers can store a range of credential types and can also use the latest 3-legged OAuth to generate short-lived dynamic tokens.

Aembit Cloud also enforces conditional access either natively or through integrations with vendors like CrowdStrike and Wiz which provide real-time security posture info on your workloads. This helps ensure consistency with your zero trust policies for users.

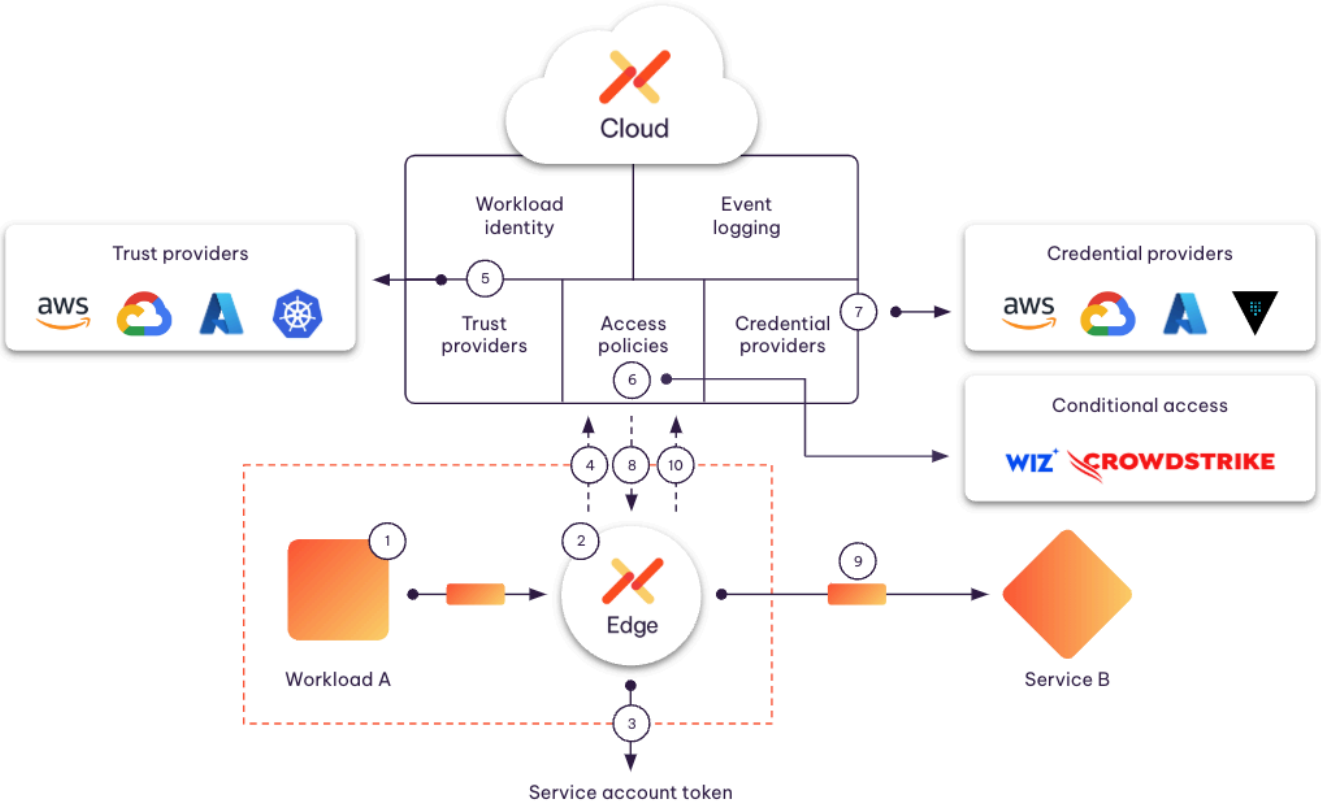
Finally, Aembit Cloud provides a structured logging approach that allows you to see and analyze access based on identity. This can be viewed in our console or sent to your SIEM for further correlation or alerting.

This architecture gives you a single workload IAM platform across your environment today – and in the future as you add clouds, services, or workloads.

> “
> **Aembit Edge makes no-code auth possible.**

Detailed Flow of a Workload Access Request

Let's walk through an example of a workload making a request using the Aembit Edge host-based proxy.



1. Client workload makes a request to service.

Custom or packaged software that you operate makes a request to a SaaS service, API, database, or another workload. The request is not aware of Aembit and it doesn't need to be.

All traffic can be sent to the Aembit Edge or explicit steering can be enabled to send only specific traffic and requests.

2. Aembit Edge intercepts client request.

The Aembit Edge is a transparent forward proxy, ensuring that you can continue best practices and will work for the range of your applications.

As a reminder, Aembit Edge is a multi-protocol proxy, deployed as a sidecar or agent, which injects service credentials without workload code changes. It performs TLS decryption for HTTPS and more, so it will work for much more than HTTP-based apps.

3. Aembit Edge retrieves service account token for secretless auth.

Aembit Edge extracts signed metadata directly from the workload's environment, serving as a robust attestation to the workload's identity. This approach eliminates the requirement for a stored identity secret on the workload by substituting it with secure dynamic identity metadata provided by the trusted workload environment.

4. Aembit Edge requests access a credential on behalf of the client.

Aembit Edge sends the request and the identity metadata to Aembit Cloud.

5. Aembit Cloud authenticates client using attestation.

Using a trust provider that was configured to trust the workload's environment, Aembit Cloud cryptographically validates the workload's identity. Trust providers can range from your local Kubernetes instance to AWS to GitLab. Multiple rules can be configured for more granular attestation.

6. Aembit Cloud checks access policy.

Policies are defined in the cloud by your team to manage access between workloads. Optional but recommended, conditional access policies such as geo-location, day and time, along with device posture conditions from integration partners such as CrowdStrike and Wiz are evaluated.

7. Aembit Cloud requests access credential from credential provider

A credential provider may be the service itself (i.e., we communicate directly with a SaaS service to provide a credential) or it may be a trusted provider who provides a credential on behalf of the application (e.g., vault, or another IAM system).

Depending on the application and use case, Aembit Cloud can act as the credential provider storing static long lived credentials such as API keys or username and password combinations, further simplifying the environment.

Aembit also uses credential providers to generate credentials (short or long lived) for many different services using OAuth 2.0 as well integrating with workload identity federation solutions from AWS, GCP, and Azure. Each server workload can have its own unique credential provider.

8. Aembit Cloud responds with the access credential.

Aembit passes the shortest-lived credentials the service supports back to Aembit Edge.

This is particularly valuable because you can move the security posture of the communication to a shorter-lived token without asking developers to make any code changes.

> **“ Injecting credentials via Aembit Edge reduces work related to credential rotation up to 95%.**

9. Aembit Edge injects credentials into client request and forwards it to the service. The workload itself never sees or stores the credential. This significantly reduces key sprawl, as workloads no longer need to know about the credential. Additionally, credential rotation is dramatically simplified. While any service that depends on long-lived credentials still needs to be manually rotated, no downstream activity must happen in workloads requesting access.

Applications that make API calls that use out-of-date, hard-coded keys also benefit from Aembit since the old, hard-coded keys are automatically replaced by Aembit Edge before the request is forwarded.

This reduces work related to credential rotation up to 95%.

After this, communications continue between workload and service as they normally would.

10. Aembit Edge sends access event log to Aembit Cloud.
The access request and access metadata are logged for analytics, auditing, and compliance. You can easily send these into a data warehouse or SIEM.

Extending Aembit to Provide Zero Trust for Workloads

Enterprises are increasingly looking to implement a zero trust architecture for workload-to-workload access, much like they have already moved or are moving to zero trust network access (ZTNA) for user access to applications.

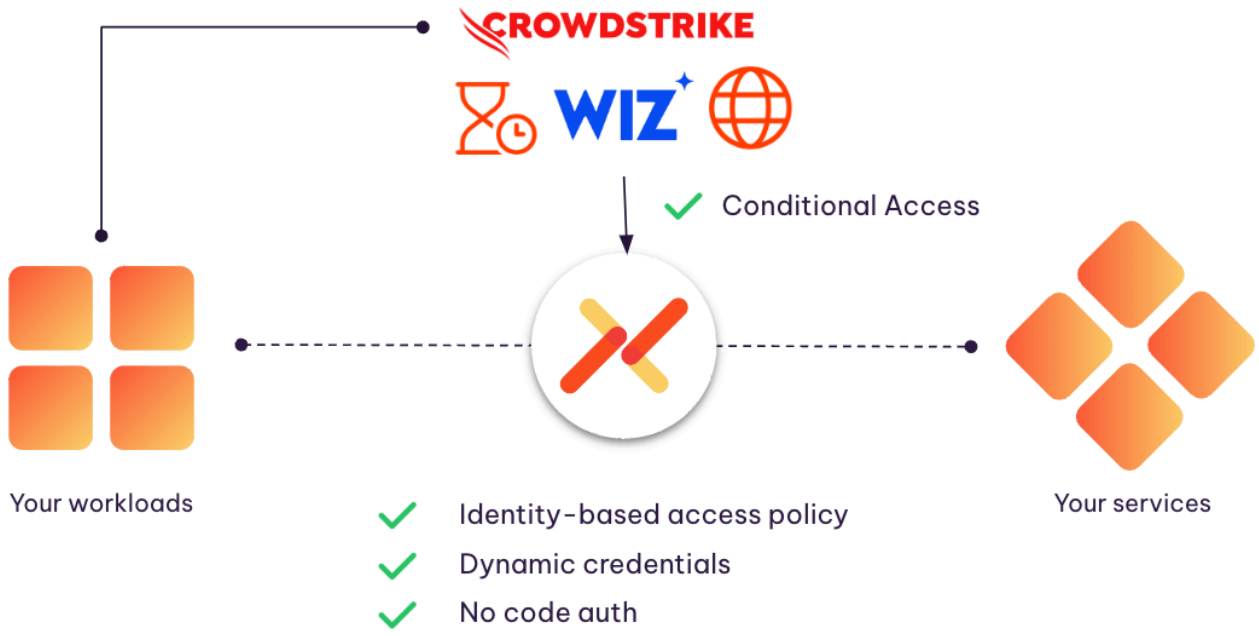
Aembit provides the most foundational elements for zero trust for workloads:

Identity + policy-based access to services. In addition, Aembit enables conditional access based on workload posture.

Not only can you use Aembit to validate the identity of an application, you can also use contextual information to determine the current state of the machine and provide conditional access based on a range of conditions. This is possible even when the end service itself doesn't support this.

These conditions are easily defined within an access policy and can be paired with Aembit's native conditional checks such as geolocation, day, and time.

The diagram below shows how Aembit implements zero trust with CrowdStrike and Wiz, as an example.



Most Common Use Cases

Enterprises generally are taking a step-wise approach to implementing workload IAM, focusing on particular environments where they believe the need for the most amount of control immediately – and then expanding to include more resources to create a unified management environment.

We most frequently see these drivers:

SaaS & API Access and Governance

Control access to third-party SaaS services in a policy-based manner. (e.g., Salesforce, MSFT Graph, Stripe and more)

Secure Data Lake and Database Access

Policy-driven access based on workload identity, consistently applied across legacy and modern environments. Applies across multiple clouds.

Protect Critical Infrastructure

Secure access to critical components like CI/CD, secrets vaults, and message buses or queues.

Zero Trust for Workloads

Allow for conditional access by checking workloads for posture conditions before access. (i.e., CrowdStrike installed with checks passing)

Enable and Protect AI

Control access to AI tools, such as ChatGPT and Gemini/Bard, either through APIs or local agents.

Federate across Environments

Create federated access using workload identity federation between different clouds, along with on-premises resources.

Typical Deployment, Packaging and Benefits

Aembit is designed as a product that you could easily self-deploy (**request a demo**).

You or your colleagues need to be able to deploy Aembit Edge within your infrastructure or lab and then set up policies in Aembit via UI, API, and/or Terraform. With our quickstart guide, you can complete your initial deployment and be testing your first server workload in 15 minutes. This can be done on your laptop, using free software, and without needing to make any network or environment changes.

As a way to begin rollout, we also encourage customers to focus on a particular application or set of applications that have sensitive data.

Typical examples include:

- Sensitive data such as a database or data warehouse.
- A particular SaaS service you'd like to begin to control, such as Salesforce.
- Strategic infrastructure, such as HashiCorp Vault or a message bus.

Our Model Is Designed to Allow You Fast, Cost Efficient Deployment

- Use our quickstart kit. It's designed to get your lab up and running in an hour.
- Make it frictionless to start. Our free tier offers up to 10 workloads at no cost, with production-class performance.
- The price is based on workloads so that costs are visible and predictable without complex per-request charges.

Based on this, we are targeting a six-month return on investment for paying customers. We know it will vary depending on your environment and use cases.

We See Our Customers Deriving Benefits in Three Areas

1. More robust security.

2. Simpler operations, which include less manual labor and reduced toolset.

3. Reduced operational load on your dev team, freeing them up to focus more on shipping your revenue-generating products than worrying about your security infrastructure.

Built on Trust

Aembit is dedicated to being a highly secure company that provides a solution that has enterprise-grade security, scalability, and reliability. Here's an overview of how this is achieved.

Security	Reliability	Scalability
ISO 27001 and SOC 2 Type 2 certified processes and procedures along with continuous monitoring of five pillars and five vendors (see trust.aembit.io)	Multi-region Aembit Cloud in leading cloud service provider (CSP) with out-of-band management and control plane monitoring and notifications.	Active monitoring and daily testing for scalability. Metrics tracked and "health-based" routing with automatic failover between regions enabled.
Segmented control and data plane with no customer traffic going to Aembit. Multi-tenant application with segmented databases and per-tenant encryption keys. Hardened application and virtual appliances.	Automatic failover enabled across all Aembit Cloud components across various availability zones and across regions.	Aembit Cloud auto scaling of all front end and back end components across various availability zones within a region.
Periodic 3rd party pentesting again Aembit Cloud infrastructure and our Aembit Edge components and virtual machines.	Hardened Aembit Edge components support multiple controller and multiple proxy deployments that can run in your primary and disaster recovery (DR) environments simultaneously.	Aembit Edge high-availability and automatic scaling across most deployment models such as Kubernetes and virtual appliances.

Case Study: Snowflake

Snowflake is a cloud-based data platform that enables organizations to store, analyze, and share large-scale data and has robust built-in security features, like data encryption and user authentication, that are designed to protect stored data. However, when it comes to workload-to-workload communication you will often discover identity and access problems. The built-in security features don't cover securing cross-boundary non-human access between Snowflake and another software or service provider.

Before partnering with Aembit, Snowflake specifically called these issues to address:

1. Inconsistent security access methods for Snowflake.
2. Widespread secrets sprawl and persistent secrets risks, with no identity-based access control.
3. Manual monitoring and key rotation, lacking dynamic control.

Working with Aembit, Snowflake was able to achieve the following:

Automate Credentials Issuance, Just-in-Time

Workload access requests are intercepted and evaluated by Aembit on an ongoing basis, injecting credentials when a policy is met. This means workloads no longer need to store long-lived credentials, and humans don't need to touch them either. This has eliminated the need for the manual credential rotation process in workloads and the potential of breaking critical application flows as a result.

Go Secretless

Since Snowflake (and a growing number of other services) support dynamic access credentials, Aembit can issue short lived credentials instead of long-lived keys. The Aembit Edge can inject these credentials into a request, even when the workload thinks it is using an API key. This enables the flexibility to improve security without burdening developers, or without asking our vendors to change how their product works.

Enable Zero Trust Conditional Access

Not only can Snowflake provide access based on an identity, it can assess characteristics of the workload before providing access. For example, is the workload being actively managed by the corporate Cloud Security tool? Or is it performing its operation during the time of day we would expect? There are a range of access conditions that can be used prior to allowing access.

Provide a Highly Automated, Compliant System of Record

Snowflake can now centrally see every access request, the policy it met, and what credential was issued. No more chasing down application owners, and no need to undergo complex reporting tasks for auditors.

Summary

Workload IAM is a powerful way to secure your application-to-application communications, especially where applications need to communicate across trust domains to accomplish their tasks.

Leveraging the Aembit Workload IAM Platform can simultaneously help you prevent credential loss, accelerate your development cycles, and eliminate the manual burdens on DevSecOps teams, including credential rotation, credential tracking, audits, and compliance.

We encourage you to **start today with our forever-free tier**, enabling you to secure up to 10 workloads at any scale.