

Guide

Director's Guide to Resilient SaaS and PaaS Infrastructure

SaaS and PaaS platforms depend on systems operating continuously and at massive scale. When those systems fail, the effects spread quickly:

- A failed authentication service can interrupt transactions instantly
- API abuse can expose fraud systems before security teams identify the source
- Traffic spikes that once caused slow page loads can now disrupt customer-facing financial workflows entirely

The board does not care whether the problem started in security, infrastructure, APIs, or delivery. They care that customers could not use the platform.

Infrastructure teams are feeling that pressure directly. APIs now sit behind nearly every customer interaction. AI-driven traffic is increasing unpredictability across production environments. Engineering teams continue pushing for faster deployment cycles while security leaders are expected to improve governance and resilience simultaneously.

[Fastly's recent research](#) found that many CISOs are now being held accountable for environments they do not fully control. That gap between responsibility and visibility becomes harder to manage as SaaS/PaaS systems grow more distributed and interconnected.

This pressure is changing how SaaS/PaaS organizations evaluate infrastructure.

The following five considerations are becoming central to platform and security decisions across the industry.

1

Can the platform stay reliable during unpredictable traffic spikes?

For SaaS and PaaS companies, reliability is the product.

Customers expect applications to remain fast and available regardless of how much traffic arrives at once. Product launches, marketing campaigns, AI-driven traffic, global events, and automated requests can all create sudden spikes that overwhelm infrastructure unexpectedly.

When platforms slow down or become unavailable, the impact is immediate:

- Users abandon workflows
- Customer trust declines
- Support tickets increase
- Revenue and retention suffer

Many infrastructure teams are discovering that traditional origin-heavy architectures struggle to absorb modern traffic patterns efficiently. As more applications become API-driven and globally distributed, even small performance bottlenecks can cascade across customer-facing systems.

This is one reason many SaaS and PaaS organizations are moving more logic, caching, and request handling closer to the edge.

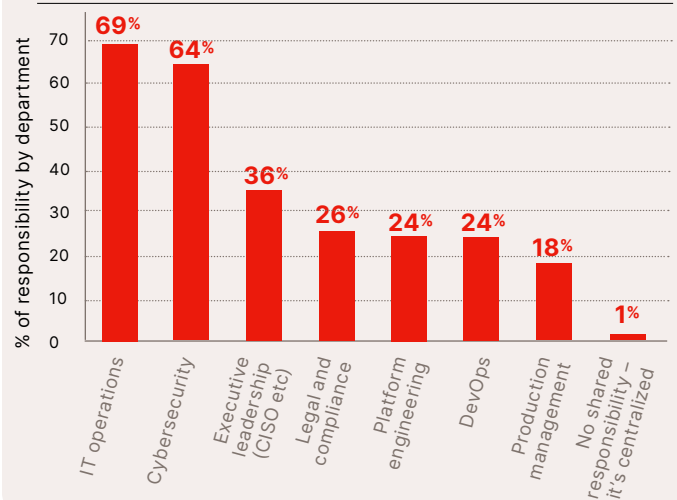
Duolingo, which serves more than 113 million monthly active users globally, faced frequent DDoS attacks that consumed significant engineering time and created operational instability. After deploying Fastly's Next-Gen WAF, the company reduced incident response overhead dramatically while improving resilience during large-scale attack events. According to Duolingo's platform security team, developers could move faster because services were protected automatically at the edge.

Squarespace also relied on Fastly during major traffic surges, including a Super Bowl advertising campaign that drove massive spikes in user demand. The company cited Fastly's stability and edge performance as critical factors in maintaining uninterrupted customer experiences during peak traffic periods.

Questions infrastructure leaders should ask:

- How does the platform perform during sudden traffic spikes?
- How much traffic still depends on origin infrastructure?
- Can traffic be filtered and managed before reaching origin?
- How quickly can configurations be deployed globally?
- Can developers ship safely without creating operational instability?

When it comes to cybersecurity, everyone has a part to play



2

How much infrastructure work is being pushed back to origin?

Modern SaaS and PaaS platforms are increasingly API-heavy, dynamic, and globally distributed. Authentication, personalization, routing, feature flags, content delivery, and application logic often compete for the same origin resources simultaneously.

That workload adds up quickly. As traffic scales, latency increases, infrastructure costs rise, and deployments become harder to manage safely. Small failures create larger downstream effects because so many systems depend on the same centralized infrastructure.

This is pushing many organizations toward edge-native architectures that reduce dependency on origin systems altogether. Instead of routing every request back through centralized infrastructure, teams can execute routing logic, caching policies, personalization, authentication workflows, and request handling closer to users.

Fastly's support for VCL and Compute gives teams significantly more control over how requests are processed globally. Infrastructure teams can deploy new configurations worldwide in seconds, execute custom logic at the edge, and reduce dependency on external professional services or lengthy deployment cycles.

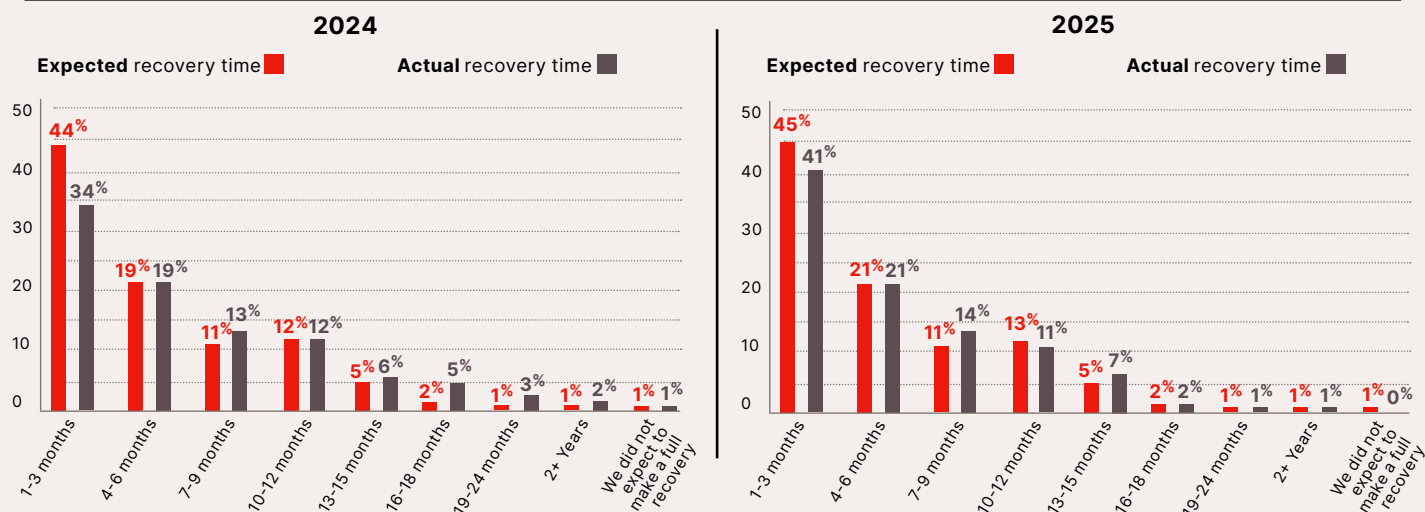
ROLLER, a cloud-based venue management platform operating across more than 30 countries, used Fastly to move complex routing and application logic closer to the edge. The company needed more than traditional geo-routing because customer data often had to be sourced from specific locations regardless of where users were located globally. By running logic at the edge, ROLLER improved performance while reducing operational complexity.

Layercake, which delivers large-scale live streaming workflows for sports and media organizations, also shifted core processing and intelligence away from origin systems using Fastly's edge platform. According to Layercake's engineering team, moving workloads closer to the edge reduced operational costs while improving scalability during major live events where traffic spikes occur instantly.

Questions infrastructure leaders should ask:

- Can API traffic be cached intelligently?
- How quickly can configurations be deployed globally?
- How much dependency remains on origin systems?
- Can infrastructure changes be made without external support involvement?

Recovery expectations are getting more realistic



For many SaaS and PaaS platforms, APIs are the product interface.

Customer applications, integrations, mobile experiences, analytics pipelines, automation tooling, AI agents, and partner ecosystems all depend on APIs operating continuously and securely.

At the same time, APIs have become one of the largest attack surfaces in modern infrastructure. Credential abuse, automated scraping, bot traffic, account takeover attempts, and API reconnaissance are all increasing across SaaS environments. Security teams are expected to improve protection without slowing down developer workflows or creating friction for legitimate users.

This is where many legacy security approaches begin to struggle. Static rules often require continuous tuning as applications evolve. Security teams end up spending large amounts of time managing exclusions, validating alerts, and maintaining policy logic instead of investigating real threats.

False positives create another problem. Blocking legitimate API traffic can directly disrupt customer workflows and damage platform trust. Many SaaS organizations are increasingly prioritizing platforms that integrate security directly into developer workflows instead of treating security as a separate operational layer.

Duo, Cisco's zero trust authentication platform, processes roughly 2 billion authentications every month. After deploying Fastly's Next-Gen WAF, Duo reported significantly fewer false positives while maintaining 99.99% availability standards. The company highlighted Fastly's ability to block malicious traffic without impacting legitimate customer authentication flows.

Questions infrastructure leaders should ask:

- Can API abuse be detected behaviorally?
- How much manual tuning does protection require?
- Can policies be deployed globally in seconds?
- Does security integrate cleanly into CI/CD workflows?
- Can APIs scale securely without adding developer friction?

4

Does the platform simplify infrastructure or create more operational overhead?

Many SaaS and PaaS organizations are trying to reduce infrastructure sprawl.

That becomes harder as security tooling, observability platforms, API gateways, edge services, cloud infrastructure, and deployment tooling continue multiplying across environments.

Infrastructure teams increasingly want:

- Fewer vendors
- Faster deployments
- Less manual policy management
- More self-service control
- Simpler rollback procedures
- Better integration with developer workflows

They do not want infrastructure changes requiring lengthy escalation chains or external professional services involvement.

Operational simplicity is becoming a major competitive advantage. One recurring frustration infrastructure teams describe is how difficult it still is to make small infrastructure changes safely. In many environments, deploying new logic or updating traffic policies still requires long deployment cycles, support tickets, or vendor-side intervention.

Fastly's edge platform allows teams to iterate quickly while maintaining direct control over how infrastructure behaves globally.

QuantCDN specifically cited Fastly's API-first architecture and developer flexibility as major advantages compared to traditional CDN providers. The company emphasized Fastly's openness, programmability, and ability to expose low-level infrastructure primitives that developers could build on directly.

Bending Spoons, which manages a large portfolio of global digital products including Evernote and Remini, also evaluated multiple providers before standardizing on Fastly. The company highlighted Fastly's faster configuration deployment, flexible VCL capabilities, Terraform support, and operational simplicity as key reasons for choosing Fastly over competing platforms.

Questions infrastructure leaders should ask

- How quickly can deployments be automated globally?
- Does infrastructure integrate cleanly into developer workflows?
- Can teams update edge logic without vendor involvement?
- How much ongoing tuning and maintenance is required?
- Does the platform reduce operational complexity over time?

5

Is the platform prepared for AI-driven traffic and automated abuse?

SaaS and PaaS infrastructure teams are already dealing with the effects of AI-driven traffic, even when they are not deploying AI systems directly.

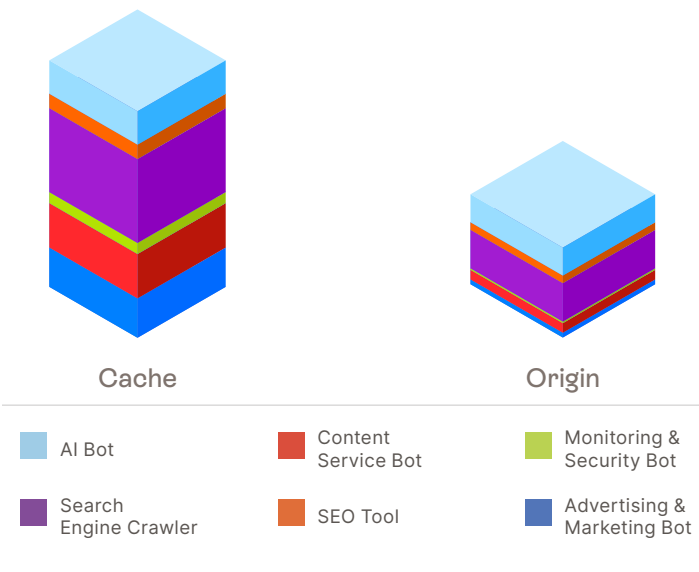
AI-assisted scraping, credential attacks, automated reconnaissance, and bot-driven abuse are increasing request volumes across SaaS platforms. Attackers can now test credentials, probe APIs, generate synthetic traffic, and adapt behaviour faster than many manual security processes can respond.

Wanted Bots by Type

Wanted Bot Traffic



20% of the High Tech industry’s wanted bot traffic is attributable to AI. This is 136% more than the global baseline.



At the same time, AI-assisted commerce and agentic systems are introducing additional APIs, integrations, and services that security teams must monitor and secure. Every new AI-connected system expands the attack surface.

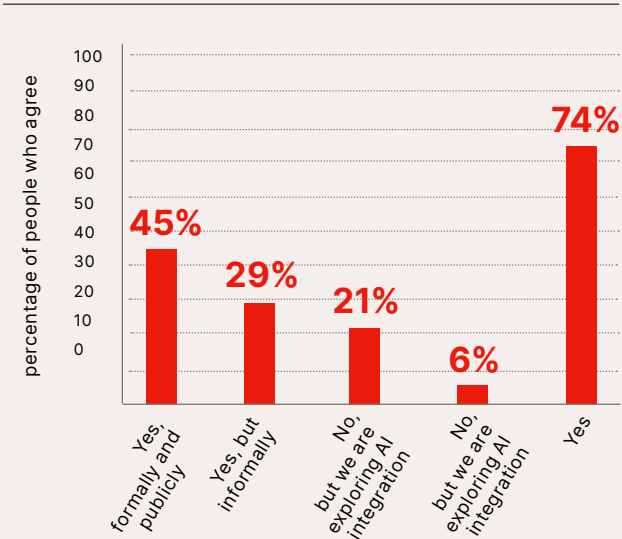
Most executive teams are not asking organizations to slow AI adoption down. They are asking security leaders to maintain governance and visibility while adoption

accelerates. That creates pressure across infrastructure and security teams simultaneously.

Many security teams already feel overloaded. They have not been able to hire their way out of alert fatigue, coordination bottlenecks, and expanding infrastructure complexity. AI-driven automation may reduce some manual work while increasing attack sophistication at the same time. Fastly’s research on AI-first organizations found that many security teams are struggling to maintain visibility and governance as AI adoption moves faster than internal operational processes can adapt.

Traditional security approaches struggle in these environments because they rely heavily on static rules and centralized enforcement models that adapt slowly to changing traffic behaviour..

AI-first organizations took longer to recover from security incidents than non-AI-first ones



Questions infrastructure leaders should ask

- How is automated traffic identified?
- Can bot behaviour be analyzed in real time?
- How much AI-driven traffic is reaching origin infrastructure?
- Can mitigation happen at the edge?
- How quickly can teams respond to changing traffic patterns?

Conclusion

SaaS and PaaS organizations are operating in environments where customer experience, reliability, APIs, security, and platform performance are tightly connected.

The pressure on infrastructure teams is not slowing down.

Many organizations are discovering that adding more tooling does not necessarily improve resilience. In many cases, it increases coordination overhead, operational complexity, and operational friction across already overloaded teams.

At the same time, customer expectations continue rising. Users expect platforms to remain fast, available, and responsive regardless of traffic volume, global demand, or attack activity.

This is changing how organizations think about infrastructure.

Fastly works with SaaS and PaaS organizations to help reduce operational complexity across security, delivery, APIs, and edge infrastructure while improving resiliency, deployment speed, and developer flexibility.

To learn how Fastly can help your team simplify infrastructure operations, strengthen API performance and security, and reduce dependency on origin systems, [contact us today](#).