Komodor 2025 Enterprise Kubernetes Report

September 2025

komodor

Executive Summary

This report outlines the key trends and challenges observed across Komodor's customer base, providing a snapshot of the current state of cloud-native and Kubernetes adoption. Our data reveals a landscape defined by increasing complexity, a rapid evolution of tooling, and a significant shift in organizational structures. Key findings indicate that while modern practices like GitOps and platform engineering are becoming standard, challenges in change management, cost control, and a persistent knowledge gap remain primary obstacles. The rise of AI/ML workloads and the resurgence of AIOps signal the next frontier for Kubernetes, while the growth in cluster and hybrid cloud adoption confirms its central role in the enterprise.

Key Highlights

Kubernetes maturity:

- ~80% of orgs run Kubernetes in production
- 93% are using/piloting/evaluating
- App separation methods: namespaces 88%, separate clusters 65%, labels 31%
- GitOps adoption: 77% (some/much/nearly all)

Footprint scale & complexity:

- A “typical adopter” now runs 20+ clusters
- 37% of orgs report & 100 clusters
- 12% of orgs report & 1000 clusters
- 48% operate clusters across four or more environments (on-prem, public cloud(s), edge, etc.)

Key Highlights

Role of AI/ML in expansion:

- 48% aren’t yet running AI/ML on K8s
- Among those that are common workloads include batch AI/ML pipelines (~11%), model experimentation (~10%), real-time inference (~10%), data preprocessing (~9%), batch inference (~8%) (share of total respondents)

Balancing isolation vs. efficiency:

- Most teams use namespaces (88%) for isolation, with 65% also segmenting by separate clusters.
- This is a practical mix that limits blast radius without multiplying ops overhead.

komodor

Change Management: The Root of Instability

The core challenge in modern infrastructure remains the management of change. Our data shows that **79% of production incidents originate from a recent system change.** The dynamic nature of cloud-native environments, coupled with the widespread use of feature flags, has made tracking and understanding the impact of changes exceptionally difficult. This complexity is a leading contributor to system instability and remains the most significant pain point for operations teams.

Other common incident root causes:

- Network failures (7%)
- Third-party/cloud provider failures (14%)
- Human change to environment (42%)
- Deploying app changes (37%)

MTTD / MTTR levels:

- Cross-org MTTD for high-impact outages: 37 min; 29% take ≥1 hour to detect
- Cross-org MTTR for high-impact outages: 51 min; 39% take ≥1 hour to resolve
- Annualized, teams spend ~134 hours detecting and ~141 hours resolving interruptions (medians across impact levels).

Change Management: The Root of Instability

Share of incidents that are customer-impacting:

- 38% experience high-business-impact outages at least weekly
- 62% estimate $1M+/hour for high-impact downtime - clear evidence that a material slice of incidents directly affects end users.
- Median annual downtime: 177 hours
- Average number of engineers involved in typical incident response: 5

Measuring business impact:

- Organizations with business observability (correlating telemetry with revenue/customer metrics) see 40% less annual downtime and 24% lower hourly outage costs. It’s becoming the standard for tying SLOs to dollars.

MTTR → outcomes:

- Reductions in MTTD/MTTR via unified telemetry + automation are associated with fewer outages and lower cost/hour
- Single-platform users also report 50% less engineering time spent on disruptions, labor that can be redirected to customer-facing work.

Tooling and Deployment: The GitOps and Helm Duopoly

The Dominance of GitOps

The adoption of modern deployment methodologies is widespread, with **84% of customers utilizing a GitOps approach.** This signifies a fundamental shift towards declarative, version-controlled infrastructure management.

- ArgoCD is the clear market leader, serving as the most popular GitOps tool.
- Flux holds a strong second position, with particularly high adoption among GitLab users.

Helm as the Standard

Helm is the de facto standard for deploying applications to Kubernetes, with 95% usage across our customer base. Its templating and packaging capabilities have made it an indispensable part of the cloud-native toolkit, often used in conjunction with GitOps controllers like ArgoCD and Flux to manage application delivery.

komodor

Infrastructure Trends: Scale, Hybridity, and the Edge

Explosive Cluster Growth

The number of Kubernetes clusters is growing at an accelerated rate. We've observed a **35% year-over-year increase in the number of clusters** managed by our customers. This growth is driven by several factors:

- Migration of legacy applications to modern infrastructure
- Adoption of a multi-cluster strategy for security and tenant isolation
- Cost management strategies that involve dedicated clusters for specific workloads

Hybrid Cloud is the Enterprise Reality

For large enterprises, hybrid cloud is not just a trend—it's the standard operating model. Nearly all of our large enterprise customers use **Kubernetes as a bridge between on-premises data centers and public cloud providers.** This allows them to leverage the cost-effectiveness of their existing infrastructure while gaining the agility and scalability of the cloud.

- **Cloud Adoption:** AWS remains the dominant cloud provider, followed by Azure in second place and Google Kubernetes Engine (GKE) in third.

The Rise of Kubernetes on the Edge

Thanks to lightweight distributions like **K3s**, we are seeing a growing number of companies deploying Kubernetes to manage workloads at the edge. This trend enables more efficient processing of data closer to its source, as exemplified by Komodor customers like KFC, who are leveraging this for their in-store operations.

Prevalence:

- **Multi-cluster** is the norm; **>50%** of orgs run **>20 clusters**, and **~48%** span **≥4 environments** (on-prem + ≥1 public cloud + edge)

Consistency challenges:

- Policy/config consistency across clusters/environments ranks among top concerns; measured reliability gains come from **unified policy + GitOps + single-platform observability.**

Failover/DR across environments:

- Higher performers pair **multi-cluster topologies** with **automated releases and GitOps** to enable predictable failover and rollbacks.

Tooling consolidation impact:

- Consolidation correlates with **fewer outages (-71% to -77%), less downtime (-18% to -79%)**, and **lower cost/hour** across multi-env estates

komodor

Operational Challenges: Sprawl, Cost, and Knowledge Gaps

Monitoring Sprawl and the Quest for Observability

Despite significant investment in monitoring, true observability remains elusive for many. **Over 50% of customers use more than one APM vendor**. However, this tool sprawl often leads to fragmented data and alert fatigue rather than a clear understanding of system health. Almost all organizations report a need for better, more integrated observability.

Organizational & Cultural Factors:

- **Centralized/platform-led ops vs. distributed:**
- Benchmarks show tool and data consolidation (a hallmark of platform-led models) is associated with fewer outages (-71% to -77%) and less downtime when coupled with full-stack observability/best practices.
- **Velocity with guardrails:**
- Teams employing policy-as-code + GitOps preserve speed while improving compliance These same practices appear in the profiles associated with lower MTTD/MTTR
- **Standardized incident workflows:**
- Where workflows and tools are standardized (single platform, unified telemetry), teams report less time addressing disruptions (-25% to -50%) and lower outage cost.
- **“Ops Powerhouses” vs. “Legacy Anchors”:**
- Powerhouses typically align to full-stack observability, unified telemetry, policy automation, GitOps, and exhibit 79% less annual downtime and 48% lower hourly outage costs than their peers.

Operational Challenges: Sprawl, Cost, and Knowledge Gaps

Cost Management is a Top Priority

Inefficient resource utilization is a major financial drain. Our data indicates that **almost 90% of customers are overspending on cloud resources**, with capacity utilization often falling below 80%. For organizations without a dedicated cost management solution, these numbers are even starker. Manually setting and maintaining resource limits and requests is rarely sufficient to control costs effectively.

◦ **Autoscaling:**

- HPA is mainstream (50% adoption; ~80% of adopters use it on a majority of clusters). Expect continued movement toward event-driven autoscaling as telemetry integration deepens.

◦ **Resource utilization today:**

- **65%** of Kubernetes workloads consume **50%** of their **requested CPU and memory** → widespread over-provisioning. **VPA** usage remains **1%; HPA** is **50%** adopted, with **~80%** of HPA users enabling it on most clusters.

◦ **Top drivers of K8s overspend:**

- **Over-provisioned requests/limits** and **rightsizing gaps** (**37%** of orgs have **≥50%** of workloads needing rightsizing)
- **Fragmented autoscaling (low VPA adoption)**
- **Idle/orphaned resources and outdated images/charts** that block consolidation.

Operational Challenges: Sprawl, Cost, and Knowledge Gaps

- **Tying cost to reliability/risk:**
- Teams consolidating observability tools (single-platform users) report **18% less annual downtime, 45% lower hourly outage cost,** and **50% less engineering time on disruptions.** A pattern where cost control and stability improve together.

- **Policies that remove waste:**
- **Rightsizing policies** (enforce CPU requests/limits), **HPA coverage, lifecycle cleanup** (PVs/LBs/idle nodes), and **image currency/SBOM policies** (to unblock smaller base images) are the highest-leverage levers.

komodor

The Persistent Knowledge Gap

The Persistent Knowledge Gap

The lack of skilled and knowledgeable Kubernetes professionals has become the **number one bottleneck for organizations**. This skills gap often leads to failed or stalled migrations and prevents companies from fully utilizing the power and potential of Kubernetes.

Best Practices & Policy

Adherence snapshot:

- Liveness/readiness probes are missing in ~65% of orgs.
- Replicas: 55% of orgs have 21% of workloads missing replicas.
- CPU requests missing: 67% of orgs have &gt;11% of workloads missing CPU requests (down from 78%).
- Security hardening gaps: 28% of orgs have 90% of workloads running with insecure capabilities; 70% have ≥11% of workloads on outdated Helm charts.
- Rightsizing need: 37% of orgs have 50% or more of their workloads in need of rightsizing.

Enforcement methods with best compliance/velocity balance:

- Org-wide policy-as-code (OPA/Gatekeeper/Kyverno or PSA levels) + admission controls + CI checks is the dominant approach in the field
- The best outcomes, however, come from proactive policy automation built directly into the delivery pipeline, which is far more effective than reactive, after-the-fact audits.

Configuration drift impact & detection:

- Drift is a frequent precursor to incidents in multi-cluster estates
- GitOps adoption at 77% (some/much/nearly all) + automated releases
- has become the baseline to detect/rollback drift quickly

The Future: Platform Engineering and the AI Revolution

The Rise of Platform Engineering

In response to rising complexity, 68% of Komodor customers have established a dedicated platform team. These teams are responsible for building and maintaining the tools, infrastructure, and "paved roads" that enable internal development teams to ship software quickly and reliably. The adoption of Internal Developer Platforms (IDPs) with tools like Backstage, Port, and OpsLevel is growing, although some organizations report challenges with the maintenance overhead and ROI of these solutions.

AI/ML Workloads on Kubernetes

We are observing a significant trend of AI/ML workloads, particularly for inference, running on Kubernetes. The scalability and resource management capabilities of Kubernetes make it an ideal platform for these demanding applications, with technologies like vLLM becoming the new standard for efficient model serving.

The Future: Platform Engineering and the AI Revolution

- **How widespread (2025):**
- **~52%** of respondents report **some** form of AI/ML on K8s (implying **48%** not yet), with use spread across **batch pipelines (~11%), experimentation (~10%), real-time inference (~10%), etc.**

- **Utilization & scheduling reality:**
- Industry surveys flag **under-utilized GPUs as a major 2024–2025 concern**
- **~40%** plan to adopt/expand **scheduling & orchestration** to maximize existing GPU capacity.

- **Operational challenges vs. CPU-only:**
- Hotspots include **scheduling fairness/queueing, driver (i.e CUDA) fragility, and bin-packing fragmentation;**
- Orgs address this via **K8s-native GPU operators/schedulers** (e.g., NVIDIA GPU/NIM operators, Kueue) plus **GitOps** for repeatability.

- **What boosts GPU efficiency:**
- **Queue-based scheduling + quotas, pre-caching models for faster scale-up,** and **policy controls** that prevent idle reservations are the cited levers.

The Future: Platform Engineering and the AI Revolution

AIOps is Back

After several years on the sidelines, AIOps is experiencing a resurgence. **35% of organizations are already using an AIOps solution,** and another **40% are planning to explore AIOps capabilities by 2026.** This indicates a growing belief that AI-driven analysis is necessary to manage the complexity of modern systems.

- **AI-assisted operations:** Already present in observability stacks (AI monitoring 42%, ML model monitoring 29%, AIOps 24%).
- Anticipated to expand across detection, triage, RCA, and auto-runbooks, where data is unified.

Looking Ahead

The data reveals a maturing Kubernetes ecosystem where initial adoption challenges are giving way to operational sophistication. Organizations are moving beyond basic containerization to focus on reliability, cost efficiency, and developer experience. Success increasingly depends on platform engineering capabilities, comprehensive observability, and - critically - building internal expertise to navigate the complexity of modern cloud-native systems.

The rise of AI/ML workloads and the return of AIOps suggest that 2025-2026 will be defined by organizations seeking intelligent automation to manage the operational complexity they've created through rapid cloud-native adoption.

Methodology

The Komodor 2025 Enterprise Kubernetes Report is based on aggregated, anonymized data from hundreds of production environments, covering thousands of Kubernetes incidents. It combines large-scale telemetry with AI-driven user insights to benchmark reliability, troubleshooting effort, cost efficiency, and emerging practices in AI-assisted operations.