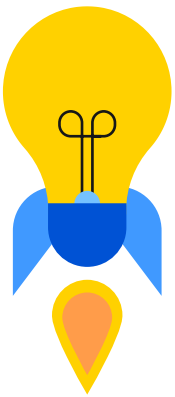
Progress® Kemp® LoadMaster®

AI-Ready Application Delivery

SOLUTION BRIEF



High-Performance Application Delivery for Inference-Driven Enterprises

AI training, inference, and RAG pipelines stall the moment the data path does. As datasets grow into the petabyte range and span clouds and regions, the access layer in front of object storage and AI APIs becomes the single biggest determinant of whether expensive GPUs stay fed or sit idle.

Artificial intelligence workloads place unprecedented demands on application delivery infrastructure. AI inference APIs, model-driven microservices, and data-intensive pipelines require ultra-low latency, massive concurrency, and continuous availability, all without introducing security risk or runaway cost.

Progress Kemp LoadMaster delivers application delivery purpose-built for AI. It ensures AI applications and data pipelines remain fast, resilient, and secure across hybrid, multi-cloud, and containerized environments, without the cost, complexity, or rigidity of legacy ADC platforms.

The Challenge: AI Breaks Traditional App Delivery

AI platforms expose critical weaknesses in legacy load balancing and ADC architectures:

- **Data-path bottlenecks** that leave expensive GPUs idle when object storage is slow, unreachable, or inconsistent across regions
- **Unpredictable traffic patterns** driven by inference bursts, real-time requests, and bursty RAG lookups
- **Latency sensitivity** where milliseconds directly impact outcomes and user experience
- **Highly distributed architectures** spanning clouds, on-premises, and edge locations
- **Expanding attack surfaces** from exposed AI APIs, model endpoints, and object-storage front doors
- **Escalating infrastructure costs** driven by inflexible licensing and hardware dependencies

Traditional ADCs were designed for static, monolithic applications. They were not built for dynamic, inference-driven AI systems that depend on distributed data and API access.

High-Performance AI Traffic and Data Delivery

- Ultra-low-latency L4/L7 load balancing for inference APIs and data endpoints
- Optimized handling of bursty, high-concurrency inference and RAG traffic
- Front-doors object-storage, feature store, and vector-database APIs for training, inference, and retrieval pipelines
- Advanced health checks and storage-aware scheduling for distributed AI data sources
- Predictable performance under extreme demand

Built for Modern AI Architectures

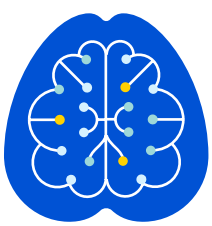
- Consistent delivery layer across on-premises, multi-cloud, and hybrid environments
- Storage-aware traffic steering and health-based routing for distributed object-storage endpoints
- REST APIs for CI/CD automation and DevOps workflows
- Stateless and stateful service optimization for diverse AI workload patterns

Intelligent Traffic Control

- Context-aware traffic steering across services and sites
- Backend isolation to protect model availability and data-service SLAs
- GSLB-based multi-site resiliency with staged cutover and migration support
- Support for canary and blue-green deployments
- Intelligent failover to maintain SLAs during appliance, site, or backend failures

Security Without Performance Penalties

- Integrated, OWASP-aligned Web Application Firewall (WAF) protecting AI APIs and model endpoints
- Pre-authentication, SSO, and MFA integrations enabling Zero Trust-style access
- DDoS mitigation for exposed AI endpoints and object-storage front doors
- TLS inspection and encryption at scale with modern cipher and protocol enforcement
- API-level protection built into the data path
- Centralized certificate lifecycle management via LoadMaster 360 to reduce outage risk from expiring certificates



Observability and Day 2 Control with LoadMaster 360

LoadMaster 360 consolidates AI front doors into a single control plane where data and API flows can be monitored, governed, and analyzed for performance and scalability:

- Telemetry and dashboards for throughput, latency, and error rates across AI data paths and APIs
- Incident management that surfaces issues early with rich context across services and sites
- Centralized certificate lifecycle management for AI endpoints, APIs, and published storage services
- Policy management, templates, backup/restore, and API-driven automation for consistent governance

Instead of treating AI traffic as "just another app," LoadMaster 360 lets platform teams see AI data flows as a first-class object in their fleet, with the same operational simplicity Progress Kemp LoadMaster already delivers.

Stop letting the data path bottleneck your AI investment.
GPUs are expensive. Downtime is more expensive. Progress Kemp LoadMaster keeps your AI pipelines fast, resilient, and secure, from inference APIs to object storage, across every cloud.

Start your evaluation today at www.progress.com/kemp

About Progress Software

Progress Software (Nasdaq: PRGS) empowers organizations to achieve transformational success in the face of disruptive change. Our software enables our customers to develop, deploy and manage responsible AI-powered applications and personalized digital experiences with agility and ease. Businesses of all sizes get a trusted provider in Progress, with the products, expertise and vision they need to turn AI disruption into a competitive advantage. Millions of developers and technologists at hundreds of thousands of organizations depend on Progress every day. Learn more at www.progress.com

/progresssw
/progresssw
/progresssw
/progress-software
/progress_sw_

Progress