NEWS RELEASE

Penguin Solutions Expands ClusterWareAI Capabilities to Accelerate Deployment and Scaling of NVIDIA Vera Rubin Platform

2026-11-17

Latest release enhances cluster management and IT integrations to streamline deployment and optimize operations for large-scale AI factories

FREMONT, Calif.--(BUSINESS WIRE)-- **Penguin Solutions, Inc.** (Nasdaq: **PENG**), the AI Factory Platform Company, today announced the latest release of its ClusterWareAI™ AI factory platform operating system software. With rapid provisioning and scaling capabilities for the NVIDIA Vera Rubin platform, this update enables enterprises and AI cloud providers to quickly deploy the latest NVIDIA POD-scale systems for agentic AI and reasoning.

Penguin Solutions ClusterWareAI software provides intelligent cluster management and a unified control plane for seamless operations, monitoring, and lifecycle management of rack-scale data center AI infrastructure. The ClusterWareAI 13.3 release introduces specialized capabilities, such as chain boot, that accelerates node provisioning and creates immutable system configurations for stable, reliable operations for teams standing up the Vera Rubin platform, including NVIDIA DGX Vera Rubin NVL72 systems. By incorporating support for Arm-based architectures, including cross-architecture imaging, and emerging classes of specialized silicon such as DPUs, this release gives customers greater flexibility as they adopt new AI infrastructure technologies.

“AI infrastructure is evolving quickly. Organizations need to bring new infrastructure like Vera Rubin online fast while confidently operating and scaling these environments," said Ian Colle, SVP and General Manager, Advanced Computing at Penguin Solutions. "ClusterWareAI accelerates time to production through rapid deployment, streamlined cluster management, and the operational capabilities needed to run large-scale AI factories reliably.”

Enhancements for Operational Efficiency and IT Integration

ClusterWareAI 13.3 improves day-to-day AI factory operations with streamlined operational intelligence, expedited human-in-the-loop remediation, and seamless integration with standard IT tools and workflows.

- Updated AI Factory Operations Agent – The AI Factory Operations Agent is a built-in conversational assistant for cluster diagnostics and operational insight. The updated agent in ClusterWareAI 13.3 introduces retrieval-augmented generation (RAG)-based knowledge access, drawing on existing documentation, automated remediation system (ARS) history, and switch inventory data. The agent delivers rapid responses based on operational context and a more efficient LLM.
- Prescriptive automated remediation – Version 13.3 expedites human-in-the-loop decision making through enhanced automated remediation capabilities. Prescriptive automated remediation provides administrators with detailed insight into system health and data-driven recommendations for automated repair action to optimize cluster performance.
- IT alerting and workflows – As cluster operations (ClusterOps) mature and increasingly align with the IT domain, organizations require AI infrastructure that integrates with IT standards. ClusterWareAI 13.3 adds routing of cluster events into industry standard IT tools, such as ServiceNow , PagerDuty, and Microsoft Teams, to support notifications, dashboards, and workflows that IT teams already use to manage operations.

Adaptable AI Factory Operations

ClusterWareAI serves as the operating system software at the center of Penguin Solutions' Full-Stack AI Factory Platform, unifying compute, memory, networking, storage, and software across large-scale data center AI infrastructure. By combining infrastructure lifecycle management, operational intelligence, and AI-powered assistance, ClusterWareAI helps organizations deploy, operate, and scale AI factories with confidence.

Learn more about ClusterWareAI AI Factory Platform Operating System Software at **penguinsolutions.com**.

ClusterWareAI and Penguin Solutions are trademarks or registered trademarks of Penguin Solutions, Inc. or its affiliates. All other trademarks are the property of their respective owners.

About Penguin Solutions

Penguin Solutions is the AI Factory Platform Company. We design, build, and manage next-generation data centers for enterprises, sovereign AI initiatives, and AI cloud providers.

With deep design expertise at the intersection of data center AI infrastructure and memory solutions, our Full-Stack AI Factory Platform combines differentiated infrastructure software, advanced memory, compute systems, end-to-

end services, and industry-leading partner technologies to help customers accelerate deployment, optimize token economics, and maximize the return on their AI investments.

Learn more at **PenguinSolutions.com**.

PR Contact

Lerin ONeill
Penguin Solutions
Corporate Communications
415-954-2800
pr@penguinsolutions.com

Source: Penguin Solutions, Inc.