



NEWS RELEASE

Penguin Solutions Expands OriginAI Solution to Accelerate AI Factory Deployment and Optimize Performance

6/18/2024

Penguin's OriginAI solution includes validated architectures backed by intelligent cluster management software and expert services that enable companies to rapidly deploy AI infrastructure at scale

FREMONT, Calif.--(BUSINESS WIRE)-- **Penguin Solutions™**, an **SGH™** brand (Nasdaq: **SGH**) that designs, builds, deploys, and manages AI and accelerated computing infrastructures at scale, today announced the expansion of its **OriginAI® solution** to include validated, pre-defined AI architectures incorporating NVIDIA technology. Backed by Penguin's intelligent cluster management software and expert services, OriginAI infrastructure streamlines AI implementation and management, and enables predictable AI cluster performance in support of customer ROI from clusters that can range in size from hundreds to thousands of GPUs.

Penguin Solutions announces the expansion of its OriginAI solution to include validated, pre-defined AI architectures backed by Penguin's intelligent cluster management software and expert services. OriginAI infrastructure streamlines AI implementation and management, enabling predictable AI cluster performance. (Graphic: Penguin Solutions)

"Designing, deploying, and operating AI factories is an incredibly complex endeavor. Our OriginAI solution builds on Penguin's extensive AI

infrastructure expertise to reduce this complexity and accelerate return on investment," said Penguin Solutions President Pete Manca. "Our OriginAI solution is a major step forward in providing CEOs and CIOs the essential and reliable infrastructure they need to deploy and manage demanding AI workloads at scale."

The OriginAI solution provides assured infrastructure for critical, demanding workloads by combining proven architectures, latest generation hardware, advanced cluster management software, and expert professional

services. These architectures are based on 1-pod, 4-pod, and 16-pod configurations which can scale from 256 to more than 16,000 GPUs. The OriginAI solution incorporates NVIDIA H100 GPUs, Penguin's Scyld ClusterWare® 12.2 software, and industry-leading networking and storage options – and is supported by Penguin's full range of managed services.

“Better GPU performance and controlled costs are top of mind for customers today. Validated, scalable OriginAI architectures developed with Penguin Solutions' hands-on expertise in designing, integrating, installing and provisioning AI infrastructure deliver both,” noted Matt Eastwood, SVP of enterprise infrastructure research at IDC. “Penguin continues to enable AI workloads for the most demanding environments through its innovative high-performance solutions and services.”

Based on proven architectures, the OriginAI solution utilizes Penguin's innovative in-factory burn-in and integration environment to validate AI cluster performance and confirm production readiness prior to shipment, ensuring that customers get the expected performance and return on investment from the moment of deployment. By combining these architectures with Penguin's Scyld ClusterWare software and its expert managed services that predictively monitor cluster health and manage solution throughput, OriginAI delivers greater than 95% overall cluster efficiency while driving higher GPU throughput than traditional approaches.

Penguin Solutions is an NVIDIA-certified Elite OEM and DGX AI Compute Systems Solution Provider and **DGX-Ready Managed Services partner** that has been delivering AI factories at scale since 2017. With 25+ years of HPC experience – and more than 75,000 GPUs deployed and managed to date – Penguin is a trusted strategic partner for AI and HPC solutions and services for leading organizations such as Georgia Tech, Meta, Sandia Labs, and the U.S. Navy.

For more information, visit Penguin OriginAI infrastructure solutions at <https://www.penguinsolutions.com/solutions/ai/originai>.

Penguin Solutions, OriginAI, and Scyld ClusterWare are trademarks of Penguin Computing, Inc., a subsidiary of SMART Global Holdings, Inc. All other trademarks are the property of their respective owners.

About Penguin Solutions

The **Penguin Solutions™ portfolio**, which includes **Penguin Computing™** and **Penguin Edge™** branded products, accelerates customers' digital transformation with the power of emerging technologies in HPC, AI, and IoT with solutions and services that span the continuum of edge, core, and cloud. By designing highly-advanced infrastructure, machines, and networked systems we enable the world's most innovative enterprises and government institutions to build the autonomous future, drive discovery and amplify human potential.

Follow Penguin Solutions on **LinkedIn**, **Twitter**, **YouTube**, and **Facebook**.

Penguin Solutions is an **SGH** Brand.

SGH Corporate Communications

Maureen O'Leary

Director, Communications, Penguin Solutions/SGH

pr@sghcorp.com

Source: SMART Global Holdings, Inc.