



NVIDIA Run:ai + Lenovo Infrastructure

Article

AI is reshaping industries—improving decision-making in finance, accelerating research in healthcare, and enhancing customer experience in retail. However, as enterprises move from small-scale pilots to real-world AI deployments, they often face critical bottlenecks in infrastructure management:

- AI practitioners experience inconsistent access to compute resources.
- IT teams struggle to allocate GPU workloads efficiently across diverse projects.
- Leadership lacks visibility into resource usage and cost implications.

Traditional infrastructure—built for static workloads—wasn’t designed to handle the dynamic, GPU-intensive demands of modern AI pipelines.

The result? **Slow AI adoption, inflated costs, and missed business opportunities.**

Overview: Introducing NVIDIA Run:ai on Lenovo Infrastructure

NVIDIA Run:ai is a Kubernetes-native platform that abstracts, virtualizes, and orchestrates GPU resources across AI workloads. It allows teams to run AI workloads as if they had an elastic GPU cloud, all while maintaining control, compliance, and visibility.

Table 1. Key features

Component	Capabilities
AI Lifecycle Integration	Build, train, serve within a unified platform
Resource Management	Pool infrastructure, enforce policies, support fractional GPU usage
Workload Orchestration	Intelligent scheduling, GPU orchestration, multi-user fairness
Open Architecture	Seamless integration with existing data science and MLOps tools

Lenovo Hybrid AI 285 Platform

NVIDIA Run:ai is a great companion for the Lenovo Hybrid Ai 285, a validated designed for both Lenovo and NVIDIA to deliver best in class performance for AI applications.

NVIDIA Run:ai also is compatible with:

- Lenovo ThinkSystem SR780a V3, SR680a V3, SR685a V3 servers
- Integration with 200Gbps Ethernet & InfiniBand, HGX H200 GPU platforms

This enables organizations to deploy Run:ai with confidence on scalable, high-performance Lenovo hardware—ensuring both operational continuity and enterprise-grade support.

For more information, see the [Lenovo Hybrid AI 285 Platform Guide](#).

Role-Based Benefits

The benefits of Run:ai include the following:

- **For IT Managers:** Centralized control, capacity planning, policy enforcement, audit-ready compliance
- **For AI Practitioners:** Reliable scheduling, elastic GPU access, lifecycle automation
- **For Platform Admins:** Efficient GPU allocation, user access control, performance monitoring

Technical Architecture

NVIDIA Run:ai consists of:

- A **Control Plane** for resource governance, monitoring, and workload submission
- Multiple **Run:ai Clusters** for distributed scheduling and orchestration
- All deployed on a **Kubernetes cluster** with secure HTTPS communication, enabling end-to-end visibility and control across AI operations.

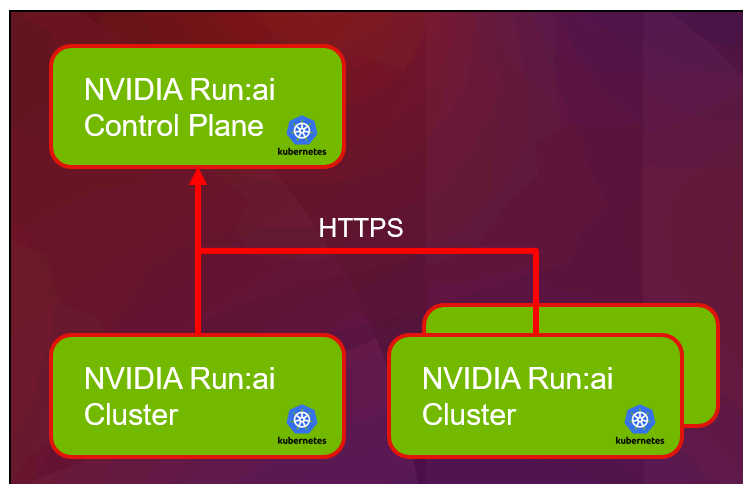


Figure 1. Technical Architecture

Conclusion

AI success depends not just on algorithms, but on infrastructure. Without modern orchestration tools, organizations face underutilized GPUs, project delays, and rising costs.

NVIDIA Run:ai on Lenovo AI Infrastructure solves these challenges by:

- Delivering dynamic, efficient GPU orchestration
- Supporting end-to-end AI lifecycle management
- Unifying infrastructure into a single, scalable platform

For IT Managers and AI teams ready to move AI from pilot to production, this solution represents a **critical leap forward in operationalizing AI at scale**.

For more information see the following resources:

- [NVIDIA Run:ai on ThinkSystem Solution Brief](#)
- [NVIDIA Run:ai Documentation](#)

Authors

Carlos Huescas is the Worldwide Product Manager for NVIDIA software at Lenovo. He specializes in High Performance Computing and AI solutions. He has more than 15 years of experience as an IT architect and in product management positions across several high-tech companies.

Sandeep Brahmarouthu is the Head of Partnerships at NVIDIA for the Run:ai product. With over 15 years of experience in sales and business development, he leads the strategy and execution of identifying and sourcing new business opportunities, acquiring new customers and partners, and negotiating complex enterprise deals in the AI and Data space.

Related product families

Product families related to this document are the following:

- [AI Servers](#)
- [Artificial Intelligence](#)

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service. Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary. Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk. Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

© Copyright Lenovo 2026. All rights reserved.

This document, LP2255, was created or updated on July 10, 2025.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
<https://lenovopress.lenovo.com/LP2255>
- Send your comments in an e-mail to:
comments@lenovopress.com

This document is available online at <https://lenovopress.lenovo.com/LP2255>.

Trademarks

Lenovo and the Lenovo logo are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. A current list of Lenovo trademarks is available on the Web at <https://www.lenovo.com/us/en/legal/copytrade/>.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Lenovo®

ThinkSystem®

Other company, product, or service names may be trademarks or service marks of others.