Elevating AI Performance: Insights from MLPerf Training 4.0 Benchmarks

Article

Highlights from MLPerf Training 4.0

Lenovo proudly participated in the MLPerf Training 4.0 benchmark suite, achieving outstanding results with the state-of-the-art Lenovo ThinkSystem SR685a V3 with 8x NVIDIA H100 Tensor Core GPUs. Here are the key outcomes:

- **ResNet-50 (Training)**: Training Time: 13.4 minutes; Hardware: 8x
- **SSD-Large (Training)**: Training Time: 36.3 minutes
- **3D-Unet-99.0 (Training)**: Training Time: 12.2 minutes

Transformative Industry Applications

The performance improvements demonstrated in the MLPerf Training 4.0 benchmarks have far-reaching implications across various industries. Let's explore how these results translate into real-world benefits:

- **Healthcare Innovation**: Model: 3D-Unet-99.0; Application: Medical imaging for accurate organ and tissue segmentation in MRI and CT scans.

  Impact: Accelerated training times enable quicker development of diagnostic tools, enhancing patient care through precise and timely medical interventions.

- **Autonomous Driving**: Model: SSD-Large; Application: Object detection in autonomous vehicles, critical for safe navigation and environment interaction.

  Impact: Reduced training times allow for rapid iteration and deployment of object detection systems, improving the safety and reliability of self-driving cars.

- **Retail Automation**: Model: ResNet-50; Application: Image classification for product recognition and inventory management in retail.

  Impact: Efficient model training supports the automation of retail operations, leading to improved accuracy in inventory tracking and enhanced customer experience through faster service.

The Lenovo Value Proposition

The Lenovo ThinkSystem SR685a V3 is an NVIDIA Certified system, equipped with 8x NVIDIA H100 Tensor Core GPUs, demonstrated its exceptional capability to handle demanding AI workloads in the MLPerf Training 4.0 benchmarks. Lenovo's commitment to innovation and performance means that our systems are not only powerful but also reliable and efficient, providing a robust foundation for AI applications across industries.

Scalability, Reliability, and Efficiency

- **Scalability**: The ThinkSystem SR685a V3 offers advanced scalability, accommodating the increasing demands of AI models as they grow in complexity and size.
- **Reliability**: Built with enterprise-grade components, our systems deliver maximum uptime and reliability, crucial for mission-critical AI applications.
- **Efficiency**: Optimized for power efficiency, the ThinkSystem SR685a V3 helps organizations reduce operational costs while maintaining top-tier performance.

Figure 1. Lenovo ThinkSystem SR685a V3 with 8x NVIDIA H100 Tensor Core GPUs

The Future of AI Performance

The MLPerf Training 4.0 benchmarks underscore the rapid evolution of AI hardware and software, highlighting the potential for significant improvements in performance and efficiency. Our contributions to these benchmarks reflect our commitment to pushing the boundaries of what AI can achieve.

As we continue to innovate, the insights gained from MLPerf Training 4.0 will guide us in developing cutting-edge AI solutions that address the most pressing challenges across industries. From healthcare to automotive and retail, the transformative power of AI is becoming increasingly evident, and we are excited to be at the forefront of this revolution.

Conclusion

The insights from the latest MLPerf benchmarks are critical for stakeholders in the machine learning ecosystem, from system architects to application developers. They provide a quantitative foundation for hardware selection and optimization, crucial for deploying scalable and efficient ML systems. Future developments in hardware and software are anticipated to further influence these benchmarks, continuing the cycle of innovation and evaluation in the field of machine learning.

Professionals in the field are encouraged to consider these results in their future hardware procurement and system design strategies.

For further discussion or consultation on leveraging these insights in specific use cases, engage with our expert team at aidiscover@lenovo.com.

For more information

For more information, see the following resources:

- Explore Lenovo AI solutions:
  https://www.lenovo.com/us/en/servers-storage/solutions/ai/
- Engage the Lenovo AI Center of Excellence:
  https://lenovo-ai-discover.atlassian.net/servicedesk/customer/portal/3
- MLCommons®, the open engineering consortium and leading force behind MLPerf, has now released new results for MLPerf benchmark suites:
  - Benchmark results: https://mlcommons.org/benchmarks/training/
  - Latest news about MLCommons: https://mlcommons.org/news-blog

Author

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.

Related product families

Product families related to this document are the following:

- Artificial Intelligence
- MLPerf Benchmark

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, LP2003, was created or updated on August 20, 2024.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
  https://lenovopress.lenovo.com/LP2003
- Send your comments in an e-mail to:
  comments@lenovopress.com

This document is available online at https://lenovopress.lenovo.com/LP2003.

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®

The following terms are trademarks of other companies:

NVIDIA® is a trademark of NVIDIA Corporation.

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