

Exceeding the Expectations of AI Performance Through MLPerf

Article

Lenovo AI is building momentum

Achieving success at the highest standards is not simply measured by achieving one single goal and riding off into the sunset on that merit alone. Success is measured by constant improvement and enhancements, which in the world of Artificial Intelligence (AI) means everything to the end user. AI requires significant processing power for its tasks and proving such capabilities requires a respected and establish partner to benchmark.

MLCommons has been a partner of Lenovo AI for 2 years now with a mission to “build fair and useful benchmarks” creating unbiased evaluations on the inference and training of hardware, software, and services. The benchmarking is known as MLPerf, which is becoming an industry standard for performance qualification.

Lenovo AI has continually improved their results quarter after quarter by improving the performance results of their submissions. Through this partnership Lenovo can confidently showcase industry setting standards of their hardware and software, which in return helps customers identify the best solutions for their environments.

MLPerf Inference v3.0 highlights

The highlights for MLPerf Inference v3.0 include the following:

- Support for NVIDIA A100 GPUs on MLPerf Inference v3.0 AI Benchmark
- Optimized Lenovo Server for AI:
 - Lenovo ThinkSystem SR670 V2 with 8x NVIDIA A100 80GB PCIe GPUs

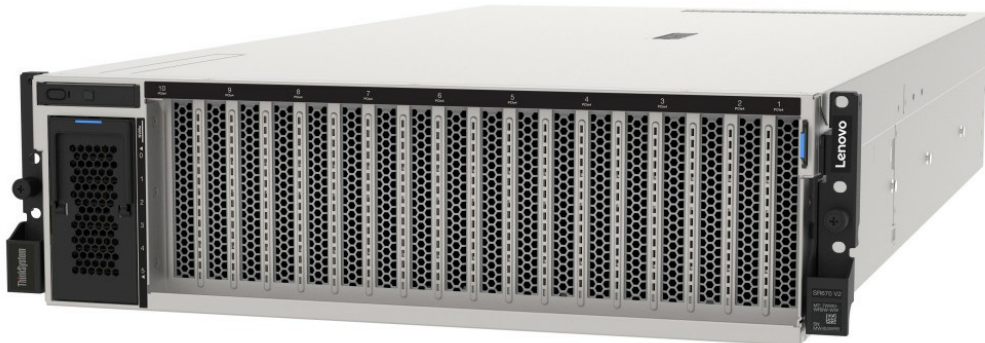


Figure 1. Lenovo ThinkSystem SR670 V2 configured to support eight double-wide GPUs

MLPerf results

Lenovo AI achieved astounding results in the MLPerf Inference v3.0 benchmarking results, winning 7 out of 10 scenarios in that category.

- Resnet50 Server – Queries 275,034.00 (Category Leader)
- 3D-UNet 99 – Offline Samples 28.5854 (Category Leader)
- 3D-UNet 99.9 – Offline Samples 28.5854 (Category Leader)
- RNNT
 - Server Queries - 98,598.90 (Category Leader)
 - Offline Samples – 106,757.00 (Category Leader)
- BERT 99 Server – Queries 24,305.00 (Category Leader)
- BERT 99.9 Server – Queries 11,798.40 (Category Leader)

Lenovo collaborates with NVIDIA

Lenovo demonstrated AI performance across various infrastructure configurations, running on Lenovo ThinkSystem platforms. We showcased the efficiency and performance of our air-cooled systems, providing both PCIe and HGX deployment options in a standard data center platform that enterprises of all sizes can quickly deploy.

Lenovo collaborates extensively with NVIDIA in the AI realm. Through our Lenovo AI Innovation Centers, we're working with NVIDIA to ensure the success of our mutual customers' AI initiatives. This provides customers with access to Lenovo and NVIDIA AI experts to aid with consulting on projects, the proper infrastructure to run a proof of concept, and proof of ROI before deployment. As the AI world continues to evolve, collaborations make coming to market an easier and more effective process.

For more information

For more information, see the following resources:

Explore Lenovo AI solutions:

<https://www.lenovo.com/us/en/servers-storage/solutions/analytics-ai/>

Engage the Lenovo AI Center of Excellence:

<https://lenovoaicodelab.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/en/inference-datacenter-30/>
- MLCommons: <https://mlcommons.org/en/news/mlperf-inference-1q2023/>

Author

David Ellison is the Chief Data Scientist for Lenovo ISG. Through Lenovo's US and European AI Discover Centers, he leads a team that uses cutting-edge AI techniques to deliver solutions for external customers while internally supporting the overall AI strategy for the World Wide Infrastructure Solutions Group. Before joining Lenovo, he ran an international scientific analysis and equipment company and worked as a Data Scientist for the US Postal Service. Previous to that, he received a PhD in Biomedical Engineering from Johns Hopkins University. He has numerous publications in top tier journals including two in the Proceedings of the National Academy of the Sciences.

Related product families

Product families related to this document are the following:

- [Artificial Intelligence](#)
- [ThinkSystem SR670 V2 Server](#)

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 2023. All rights reserved.

This document, LP1714, was created or updated on April 7, 2023.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
<https://lenovopress.lenovo.com/LP1714>
- Send your comments in an e-mail to:
comments@lenovopress.com

This document is available online at <https://lenovopress.lenovo.com/LP1714>.

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.