

Unveiling the Power of the Lenovo ThinkSystem SR675 V3 Server Through MLPerf Benchmarking

Article

MLPerf benchmarks are the gold standard for evaluating the performance of machine learning models across diverse tasks. The Lenovo ThinkSystem SR675 V3 server, equipped with cutting-edge hardware, has been put through rigorous testing across various ML workloads, showcasing exceptional results that highlight its efficiency and capabilities.

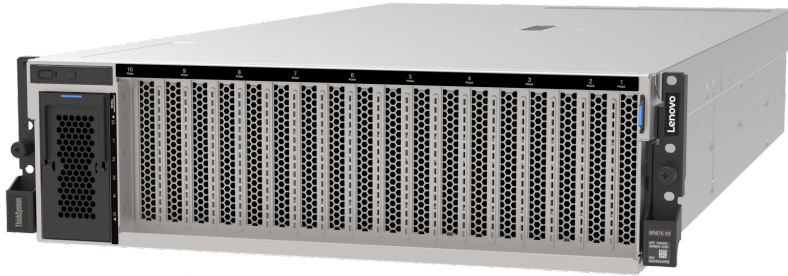


Figure 1. Lenovo ThinkSystem SR675 V3 configured with 8x 80GB PCIe H100 GPUs

Technical breakdown of MLPerf results

The results across different workloads on the SR675 V3 server and its configuration with 4x 96GB SXM5 H100 (AMD EPYC 9654 96-Core) and 8x 80GB PCIe H100 (AMD EPYC 9554 64-Core) to certify their performance.

Table 1. Benchmark results comparison (green indicates a higher result)

Server	SR675 V3	SR675 V3
GPU	4x 96GB SXM5 H100 GPUs	8x 80GB PCIe H100 GPUs
Processor	AMD EPYC 96C 9654	AMD EPYC 64C 9554
BERT	13.195	18.513
Mask RCNN	50.513	38.422
ResNet50	35.063	22.659
SSD	90.078	58.579
RNNT	31.503	26.398
3D-Unet	24.064	19.386

These results showcase the significant performance gains observed when utilizing 4x 80GB SXM5 H100 GPUs and the AMD EPYC 9654 96-Core processors, emphasizing the server's scalability and enhanced processing capabilities across various tasks.

Industry applications of SR675 V3's performance

The improved performance across these workloads has a myriad of real-world applications in diverse industries:

- **NLP in Customer Service and Financial Sectors**
Faster BERT inference speeds enable real-time sentiment analysis and document summarization for enhanced customer experiences and streamlined financial analyses.
- **Computer Vision in Retail and Manufacturing**
Quicker Mask RCNN and ResNet50 inference times facilitate faster object detection, aiding inventory management, defect detection, and quality control in manufacturing and retail.
- **Autonomous Vehicles and Surveillance**
SSD workload enhancements support faster object detection, crucial for real-time decision-making in traffic control and surveillance applications.
- **Speech Recognition in Healthcare and Telecommunications**
Improved RNNT performance can revolutionize speech recognition applications, benefiting healthcare transcription and voice-activated systems in telecommunications.
- **Healthcare Imaging and Research**
Speedier 3D-UNet inference can advance medical imaging for disease diagnosis, drug discovery, and treatment planning, potentially saving lives through timely diagnoses.

Conclusion

The MLPerf results for the SR675 V3 server and its performance with 80GB SXM5 H100 (AMD EPYC 9654 96-Core) unveil not only its technical capabilities but also its potential to drive innovation and efficiency across multiple industries. The SR675 V3 stands as a beacon of high-performance computing solutions, promising to power transformative applications and services in AI and machine learning.

From healthcare to retail, finance to autonomous systems, the SR675 V3 server's exceptional performance across various ML workloads underscores its significance in shaping the future of AI-driven applications. As organizations seek advanced technologies, the SR675 V3 server proves to be a leading solution in driving this technological evolution.

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://lenovoaiodelab.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/training-normal-20/>
- 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:

- [AI Servers](#)
- [Artificial Intelligence](#)
- [High Performance Computing](#)
- [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 2025. All rights reserved.

This document, LP1915, was created or updated on May 1, 2024.

Send us your comments in one of the following ways:

- Use the online Contact us review form found at:
<https://lenovopress.lenovo.com/LP1915>
- Send your comments in an e-mail to:
comments@lenovopress.com

This document is available online at <https://lenovopress.lenovo.com/LP1915>.

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:

AMD and AMD EPYC™ are trademarks of Advanced Micro Devices, Inc.

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