# PowerEdge R760 ResNet50 Testing Overview and Results

**Authors**:
Todd Mottershead, Jay Engh, Charan Soppadandi, Nagesh DN (Intel), Patryk Wolsza (Intel), Esther Baldwin (Intel)

## Summary

The testing outlined in this paper was conducted in conjunction with Intel and Solidigm. Server hardware was provided by Dell, processors and network devices were provided by Intel, and storage technology was provided by Solidigm. All tests were conducted in Dell Labs with contributions from Intel Performance Engineers and Dell System Performance Analysis Engineers.

With the introduction of the 4th Gen Intel® Xeon® Scalable processors, the new Dell PowerEdge R760 can benefit from important new features such as Advanced Matrix Extensions (AMX) to improve deep learning performance. To evaluate this, we recently tested the R760 using the TensorFlow framework with the ResNet50 (residual network) CNN model to determine the performance of these new features compared to previous generations of servers. This testing demonstrated more than 3x improvement in performance in the BF16 compared to FP32 precision and more than 2x improvement in performance compared to the previous generation R750 in INT8 precision.

## Configurations tested

- BASELINE: Intel® Xeon Platinum 8380 (ICX Config): 4 Nodes, Each Node with 2x Intel® Xeon® Platinum 8380 Processor, 1x PowerEdge R750, Total Memory 1536 GB (16x 32GB + 16X64GB , DDR4 3200MHz), HyperThreading: Enabled, Turbo: Enabled, NUMA noSNC,, BIOS:Dell1.6.5 (ucode:0xd000375),Storage (boot): 1x 480 GB Micron SSD, Storage (cache): 2x 800 GB Intel® Optane™ DC SSD P5800X Series, Storage (capacity): 6x 3.2 TB SolidigmDC P5600 Series PCIe NVMe, Network devices: 1x Intel® Ethernet E810CQDA2 E810-CQDA2,at 100GbERoCEv2,Network speed: 100GbE, OS/Software: VMware 8.0, 20513097, Test by Dell & Intel as of 21/12/2022using Ubuntu Server 22.04 VM (vHW=20, vmxnet3), vSAN default policy (RAID-1, 2DG), Kernel 5.19.17, intel-optimized-tensorflow:2.11.0, ResNet50v1.5, Batch size=128, VM=80vCPU+96GBRAM

- SPRPlus: Intel® Xeon® Platinum 44 core Pre-Production Processors. 4 Nodes, Each Node with 2x Intel® Xeon® Platinum Pre-Production Processors, 1x PowerEdge R760, Total Memory 2048 GB (16x 128GB DDR5 4800MHz), HyperThreading: Enable, Turbo: Enabled, NUMA noSNC, BIOS: Dell 0.2.3.1(ucode:0x2b000081), Storage (boot):1x600GB Seagate Enterprise drive, Storage (cache): 2x 800 GB Intel® Optane™ DC SSD P5800X Series, Storage (capacity): 6x 3.2 TB Solidigm SSD DC P5600 Series PCIe NVMe, Network devices: 1x Intel® Ethernet E810CQDA2 E810-CQDA2,at 100GbERoCEv2,Network speed: 100GbE, OS/Software: VMware 8.0, 20513097, Test by Dell & Intel as of 11/21/2022using Ubuntu Server 22.04 (vHW=20, vmxnet3), vSANdefault policy (RAID-1, 2DG), Kernel 5.19.17, intel-optimized-tensorflow:2.11.0, ResNet50v1.5, Batch size=128, VM=88vCPU+96GBRAM

## Security mitigations

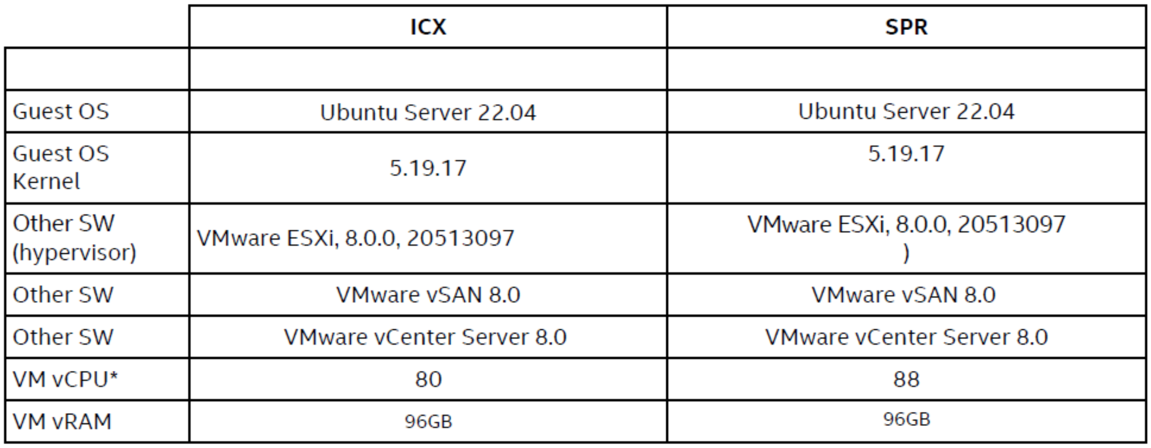
The following security mitigations were evaluated and passed:

CVE-2017-5753, CVE-2017-5715, CVE-2017-5754, CVE-2018-3640, CVE-2018-3639, CVE-2018-3615, CVE-2018-3620, CVE-2018-3646, CVE-2018-12126, CVE-2018-12130, CVE-2018-12127, CVE-2018-11091, CVE-2018-11135, CVE-2018-12207, CVE-2020-0543, CVE-2022-0001, CVE-2022-0002

## Systems architecture

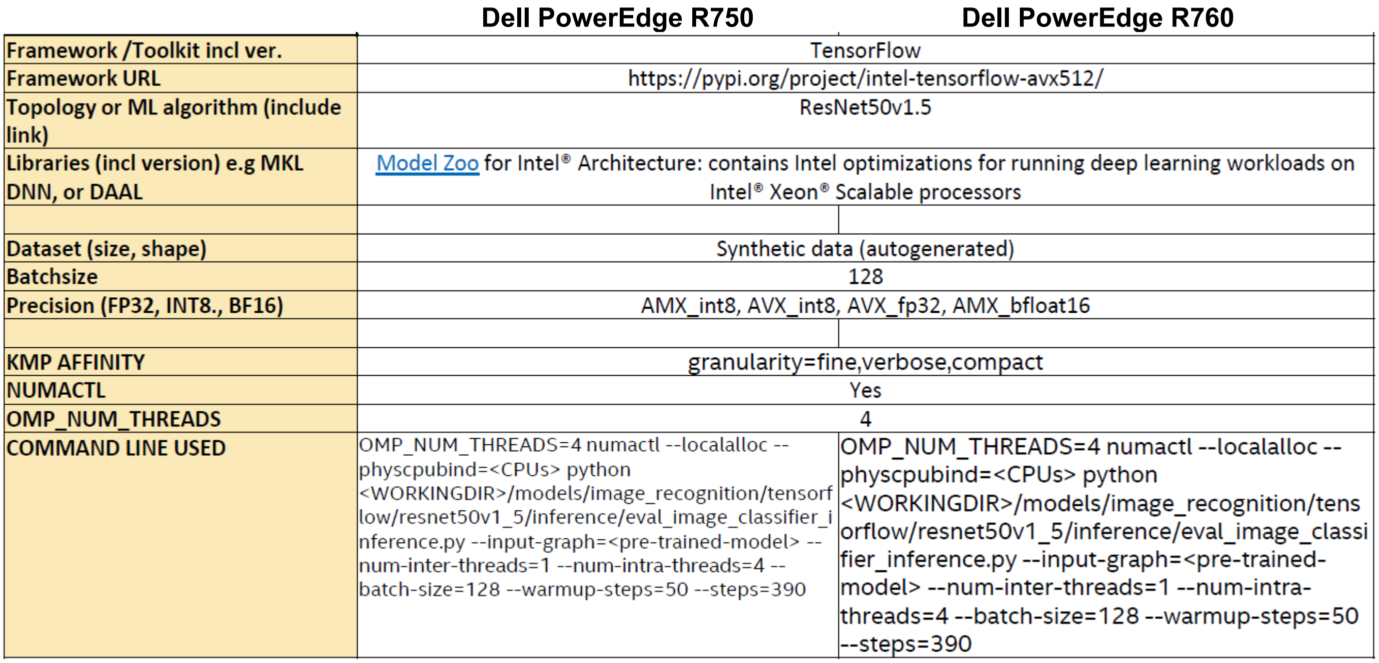
Deep learning environments both process and generate large amounts of data. To facilitate this in our testing, we used a VMware vSAN 8 cluster to store all data.

**Hypervisor, VM, and guest OS configuration**



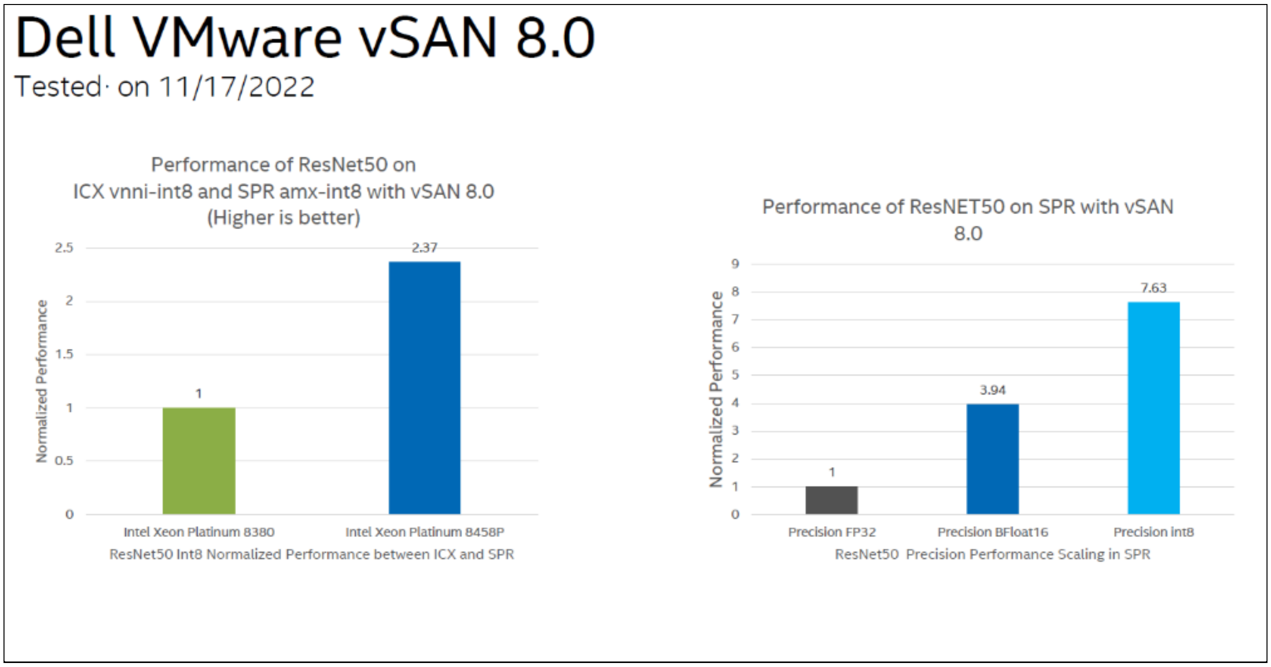
| | ICX | SPR |
|---|---|---|
| | | |
| Guest OS | Ubuntu Server 22.04 | Ubuntu Server 22.04 |
| Guest OS Kernel | 5.19.17 | 5.19.17 |
| Other SW (hypervisor) | VMware ESXi, 8.0.0, 20513097 | VMware ESXi, 8.0.0, 20513097 ) |
| Other SW | VMware vSAN 8.0 | VMware vSAN 8.0 |
| Other SW | VMware vCenter Server 8.0 | VMware vCenter Server 8.0 |
| VM vCPU* | 80 | 88 |
| VM vRAM | 96GB | 96GB |

## Benchmark configuration



| | Dell PowerEdge R750 | Dell PowerEdge R760 |
|---|---|---|
| **Framework /Toolkit incl ver.** | TensorFlow | |
| **Framework URL** | https://pypi.org/project/intel-tensorflow-avx512/ | |
| **Topology or ML algorithm (include link)** | ResNet50v1.5 | |
| **Libraries (incl version) e.g MKL DNN, or DAAL** | Model Zoo for Intel® Architecture: contains Intel optimizations for running deep learning workloads on Intel® Xeon® Scalable processors | |
| | | |
| **Dataset (size, shape)** | Synthetic data (autogenerated) | |
| **Batchsize** | 128 | |
| **Precision (FP32, INT8., BF16)** | AMX_int8, AVX_int8, AVX_fp32, AMX_bfloat16 | |
| | | |
| **KMP AFFINITY** | granularity=fine,verbose,compact | |
| **NUMACTL** | Yes | |
| **OMP_NUM_THREADS** | 4 | |
| **COMMAND LINE USED** | OMP_NUM_THREADS=4 numactl --localalloc --physcpubind=<CPUs> python <WORKINGDIR>/models/image_recognition/tensorflow/resnet50v1_5/inference/eval_image_classifier_inference.py --input-graph=<pre-trained-model> --num-inter-threads=1 --num-intra-threads=4 --batch-size=128 --warmup-steps=50 --steps=390 | OMP_NUM_THREADS=4 numactl --localalloc --physcpubind=<CPUs> python <WORKINGDIR>/models/image_recognition/tensorflow/resnet50v1_5/inference/eval_image_classifier_inference.py --input-graph=<pre-trained-model> --num-inter-threads=1 --num-intra-threads=4 --batch-size=128 --warmup-steps=50 --steps=390 |

## Test results



Dell VMware vSAN 8.0

Tested· on 11/17/2022

ICX – 3rd Gen Intel® Xeon® processors used in the R750

SPR – 4th Gen Intel® Xeon® processors used in the R760

## Conclusion

The new Dell PowerEdge R760 with 4th Gen Intel® Xeon® processors delivers outstanding machine learning (ML) performance. Using the Intel® AMX features and AVX-512 instruction set delivers performance levels up to 2.37x better than previous generations. As customers look to expand their deployments of ML workloads, the combination of 4th Gen Intel® Xeon® processors and the innovative Dell PowerEdge R760 provide a cost-effective solution that does not require the addition of expensive GPU technologies.

For more info, visit the Servers Info Hub

Contact us for feedback and requests

Follow us for PowerEdge news

DELL Technologies