TEST REPORT

Tolly.

#224134

June 2024

Commissioned by

Dell Inc.

Dell PowerSwitch Z9664F-ON 400GbE Ethernet AI Fabric Performance

EXECUTIVE SUMMARY

AI training involves massive amounts of data being processed in parallel by GPUs across a collection of high-end servers. Network designers need to have confidence that their Ethernet networking fabric can handle the demands of AI processing and perform at or near wire speed. The Dell PowerSwitch line offers high-performance and high-port density that can serve as a network fabric for AI.

Dell Inc. commissioned Tolly to benchmark a 400GbE network fabric consisting of 10 Dell PowerSwitch Z9664F-ON switches with Enterprise SONiC Distribution by Dell Technologies version 4.3.0 in a RAIL optimized topology. Eight Dell PowerEdge servers, each outfitted with eight GPUs and eight 400GbE network interfaces, were connected to the switch fabric and were tasked with running various AI workloads across the network fabric.

The Dell PowerSwitch Ethernet Fabric provided zero-loss handling of AI traffic from the 64 connected NICs/GPUs delivering ~391Gbps of Perftest RDMA throughput and ~390GBps of NCCL benchmark inter-node throughput. See Figure 1.

THE BOTTOM LINE

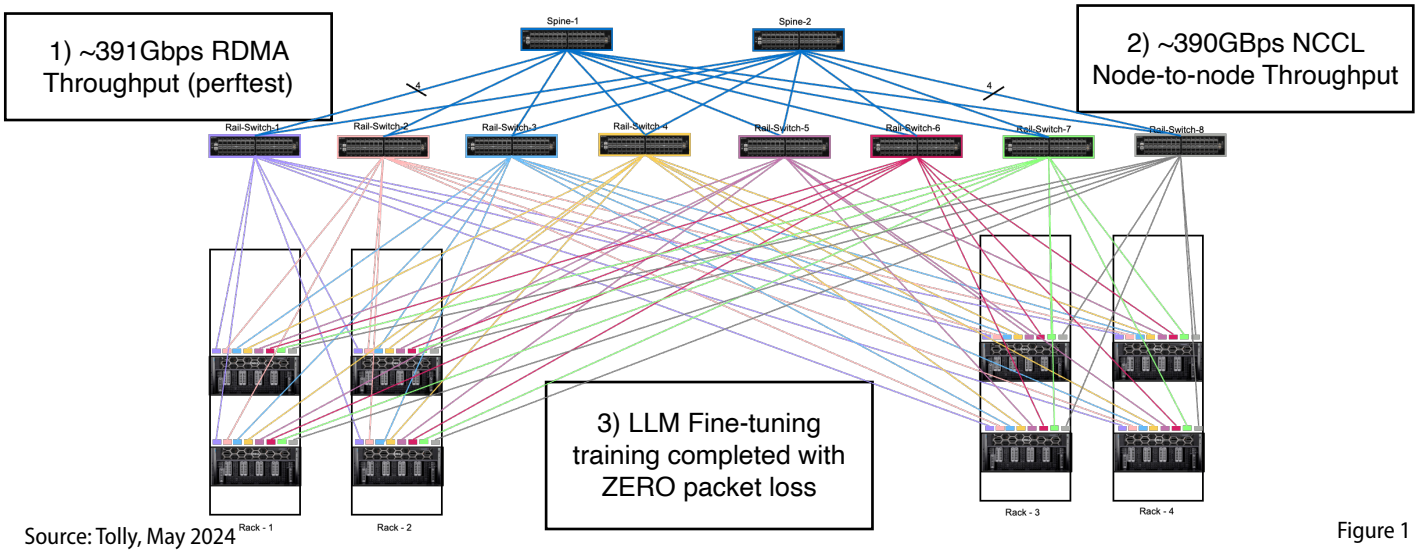
The Dell Z9664F-ON PowerSwitch Ethernet Fabric demonstrated support for carrying:

1 ~391Gbps RDMA perftest throughput

2 ~390GBps inter-node bandwidth for xCCL test (8 x 400GbE NICs)

3 AI training via LLM fine-tuning with zero packet loss

Dell PowerSwitch Z9664F-ON 400GbE AI Fabric Test Environment



Source: Tolly, May 2024

Figure 1

Goal

While the test configuration was complex, the test goal was simple - to demonstrate that the Dell PowerSwitch 400G Ethernet network fabric could transport RDMA and NCCL traffic at or near line rate across the 400GbE network fabric and without packet loss. As noted briefly above, the test was successful.

Configuration

While many test environments are small and represent a microcosm of a real-world environment, this test bed was a real-world environment, running actual AI micro benchmarks and applications, one of the most extensive ever used in a Tolly test. The scale was chosen to provide a realistic load and traffic flow and testers used recognized AI micro benchmarks to generate the workload on the systems and, thus, the traffic across the network fabric. Configuration details are found in Tables 1 through 3 later in the report.

Switch Fabric

All LAN switches were Dell PowerSwitch Z9664F-ON devices with all ports set to 400Gbps. The fabric topology consisted of two spine switches and eight leaf switches with BGP using RAIL optimized architecture. The leaf switches used adaptive routing and switching to implement the fabric.

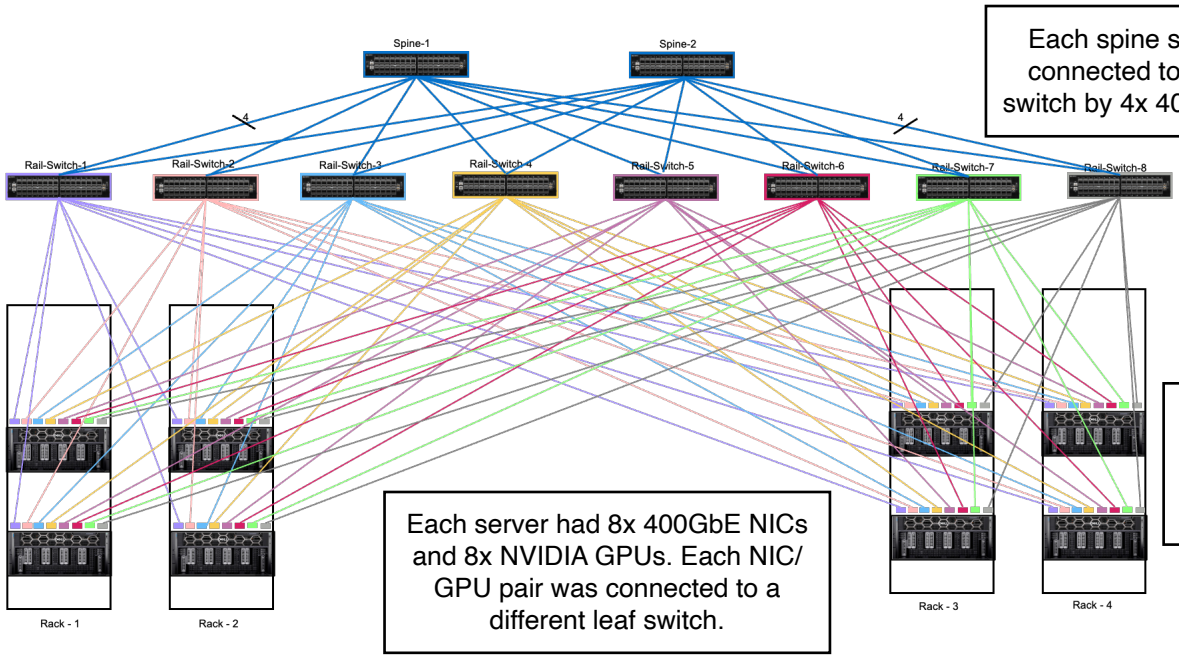
There were four, 400GbE links between each of the two spine switches and each of the eight leaf switches totaling 32x 400GbE uplinks for each spine switch. (See Figures 1 and 2, same diagram with different notations.) Dell has developed a "roce enable" command that configures its switches to provide optimized settings for RDMA over converged Ethernet (RoCE) traffic.

AI Server Nodes

The test used eight Dell PowerEdge XE9680 servers shown, two to a rack, in Figures 1 and 2. Each server was outfitted with 8x 400GbE NICs and 8 GPUs. To optimize traffic flow for GPU-to-GPU parallel processing, each GPU used a dedicated NIC. NIC 1 in each server node was connected to switch 1, NIC 2 to switch 2, and so forth. This is known as a RAIL optimized configuration. This configuration provides for GPUs in the same server node to communicate either internally or via the spine switches.

Dell PowerSwitch Z9664F-ON Switch: AI Fabric Test

RAIL Optimized Topology: 10 Switches, 8 Servers, 64 GPU Cluster



Source: Tolly, May 2024

Figure 2

Tests & Results

RDMA Throughput

RDMA throughput was demonstrated using the perftest package. Perftest is an open source micro benchmark, architected as a client-server model, that drives RDMA reads/writes.

For this test, one node was configured as the target server and another was configured as clients. RDMA "reads" were issued from the clients to the server using the ib_read_bandwidth command. Test was run on all eight NICs simultaneously on both client and server nodes.

The test was run for two minutes and the network throughput for each of the eight NICs averaged 390.89Gbps, close to the 400Gbps maximum. Tolly engineers confirmed that the switch links being traversed by the pretest traffic were showing utilization of 99% or 100%.

NVIDIA Collective Communication Library (NCCL) Test

These tests are provided by the GPU vendor, in this case NVIDIA and consist of various test routine. For this test, the "All reduce" routine was used. The algorithm, chosen by the benchmark, was "ring" for the eight nodes involved in the test. RoCE v2 was used.

The NCCL test was run using an open source "Pytorch" container and the test data size was 16GBytes. Test run time was unlimited.

The test showed node-to-node (i.e., eight NICs/ports to eight NICs/ports) bus bandwidth to be 389.7Gbytes. This means that each of the eight, 400GbE NICs and associated switch ports were running at wire speed.

LLM Fine-Tuning

This test also used all eight nodes. Key elements:

- Model: Llama-2 70B
- Precision: bf16
- Fine-tuning Technique: PEFT - LoRA
- Data Set: Databricks-Dolly-15K
- Steps: 1010, Global Batch Size: 128, Micro Batch Size =1, TP=8, PP=1
- LLM Training Framework: Nemo

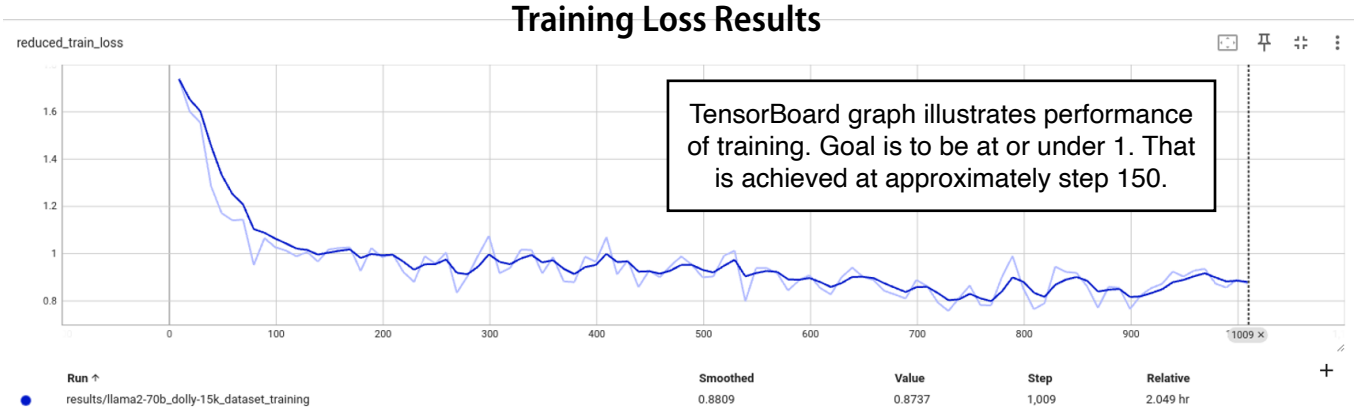
The purpose of this test was to illustrate that AI training could take place across the Ethernet fabric without packet loss caused by congestion.

The fine-tuning test ran for 2 hours and 6 minutes and completed successfully. The test consisted of multiple functions as the model was loaded and the test was run. Thus, the load on the network fabric varied during the test. The Dell Ethernet fabric was more than sufficient for fine-tuning workloads

After the test completed, Tolly engineers reviewed the counters on the Dell PowerSwitch fabric and confirmed that there was zero packet loss.

With any AI model, the training process or training loss is another key parameter that requires attention when benchmarking any AI model relative to the infrastructure. Training loss is a measure of how well a model is performing on the training data. The end goal is to decrease over time and be at 1 or under.

Training Loss Results



Source: Tolly, May 2024

Figure 3

Figure 3 shows how the Dell Z9664F-ON switch fabric performed optimally. At the beginning of the model training, one can see how training loss is high, but eventually, as the training continues, the training loss decreases as expected and is equal to or less than 1. If the fabric infrastructure does not perform well, if for example, there is packet loss, congestion, or low throughput, training loss and overall completion time would be adversely affected.

Test Setup & Methodology

The test methodology and setup for each test was explained, as appropriate, in the "Test Results" section of the report. Switch hardware and software as well as server hardware and software components and versions levels can be found in Tables 1 through 3.

Dell customers interested in additional details of server BIOS settings and specific switch configuration parameters should contact their Dell technical team.

Dell Inc.

Dell PowerSwitch Z9664F-ON 100/400GbE Aggregation Switch

AI Ethernet Fabric Performance Evaluation

Tested May 2024

Dell Ethernet AI Fabric Test Bed Details

Switch Fabric (Same for All 10 Switches)

Vendor	Switch	OS	Version	Notes
Dell	Z9664F-ON	Enterprise SONiC Distribution by Dell Technologies	4.3.0 GA	RAIL Optimized switch-server configuration: Two spine, eight leaf switches. All ports configured for 400Gbps.

Table 1

Server Configuration (Same for All 8 Servers)

Vendor/ System	Dell PowerEdge XE9680 (48c, 350W)
CPU	2x Intel Xeon 8468 (48c, 350W)
Memory	16x 64GB 4800 GHz RDIMMS
GPU	8x NVIDIA H100 module
BOSS M.2	2x 480GB
Drives	4x 1.92TB Samsung U.2 NVMe SSD
PCIe cards	8x NVIDIA ConnectX-7 VPI 400G 2x CX-6Dx 100GbE
Onboard LOM	Broadcom BCM5720 2x 1GbE

Table 2

Server Software

Node Operating System	Ubuntu Linux 22.04.3 LTS
NVIDIA CUDA	12.4
NVIDIA GPU Driver	550.54.15
NVIDIA ConnectX-7 Firmware	28.39.1002
OFED Driver	MLNX_OFED_LINUX-24.01-0 .3.3.1
NCCL	2.21.5
NVIDIA Fabric Manger	550.54.15
NGC Pytorch Container	24.03

Table 3

Source: Tolly, May 2024

Note: Each server used eight 400GbE NICS and eight GPUs for running the tests.

About Tolly

The Tolly Group companies have been delivering world-class IT services for more than 35 years. Tolly is a leading global provider of third-party validation services for vendors of IT products, components and services.

You can reach the company by E-mail at sales@tolly.com, or by telephone at +1 561.391.5610.

Visit Tolly on the Internet at:
http://www.tolly.com

Terms of Usage

ef-30-wt-2024-06-28-Verl