New Dell I/O Optimized PowerEdge R640 Delivers CDN Performance

The Dell NUMA-Balanced I/O Optimized PowerEdge R640* based on Intel® technologies demonstrates 84% better throughput and consistent latency for I/O-intensive network edge applications such as content delivery networks.¹

A content delivery network (CDN) frequently incorporates a geographically distributed network of caching servers at the network edge that provide fast and efficient delivery of popular content. Consumer trends, such as streaming high-definition video, IoT-related traffic demands, and high-data-intensity use cases will make high-throughput, low-latency content delivery essential. As this content continues to get richer, more localized, personalized, interactive, and immersive, the pressure is on for the CDN providers to deliver much more with less. CDN providers are increasingly looking for ways to optimize platform performance and costs. One of the key platform design considerations to get the most performance efficiency out of a standards-based dual-socket x86 server, is non-uniform memory access (NUMA)-balanced I/O. Intel® Network Builders ecosystem member Dell Technologies worked with Intel to test its I/O Optimized PowerEdge R640 server to demonstrate the performance that can be achieved from a balanced version of the company’s most popular data center server.

Multi-Processor Servers Need Balance

A processor within a system has access to its resources, which include compute, memory, and I/O. Modern systems use non-uniform memory access (NUMA) to scale beyond a single processor. A processor and its resources are considered a NUMA zone. Within a system with multiple Intel® processors, Intel® Ultra Path Interconnect (Intel® UPI), a coherent interconnect for scalable systems, gives processors the ability to access other NUMA zone resources. With modern systems it is critical to architect NUMA zones that are balanced with their network and storage I/O and minimize transactions that cross the Intel UPI. Without a balanced I/O approach, performance may be impacted by added latency from applications going over the Intel UPI for access to network or storage resources from a remote NUMA zone. Utilizing remote NUMA zone resources is associated with network throughput limitations and latency inconsistency.

Dell Technologies has developed an I/O Optimized PowerEdge R640 with a well-engineered approach to a Dual-Processor system with NUMA balance for networking and storage I/O. The new R640 design method has allowed Dell to perform at a level that unbalanced systems cannot come close to achieving.³

Dell Technologies I/O Optimized PowerEdge R640

The I/O Optimized PowerEdge R640 is a new option in the popular PowerEdge R640 server family that enables the outstanding performance with storage and networking resources balanced across two Intel® Xeon® Scalable processors (two models are available, one with Intel Xeon Platinum 8200 series processors and

*NUMA-Balanced I/O Optimized PowerEdge R640 available to select customers
The CPUs are optimized for demanding mainstream data center, multi-cloud compute, and network and storage workloads. With support for higher memory speeds, and enhanced memory capacity, Intel Xeon Scalable processors deliver high performance, advanced reliability, and hardware-enhanced security features.

The I/O Optimized PowerEdge R640 server configuration also includes 24 DIMMs for a maximum memory of 3 TB using RDIMM/LRDIMM. The server features 10 NVMe drives (five per processor), two Intel® Ethernet 800 Series network adapters (one per processor) and two SATA M.2 modules with RAID for boot.

The optimized performance is achieved by applying best-known NUMA optimization methods to assign networking and storage resources to each processor, thus balancing the resources that are directly available to each of the two CPUs. With NUMA optimizations, each CPU can be directly connected to five solid state drives (SSDs). With this configuration, the percentage of remote requests is greatly reduced, resulting in very high overall performance. Balancing the storage on the I/O Optimized PowerEdge R640 allows the server to fully exercise the 200 Gbps networking capacity potential of the 10 NVMe SSDs, for significant performance improvements to the speed of data access at the edge of the network.

The I/O Optimized PowerEdge R640 has been designed to boost acceleration of edge workloads including video streaming, web and mobile content, software downloading, and online retail, in addition to large CDN implementations. The high performance storage access will also be of interest to other storage-dependent edge computing workloads including virtual storage area networks (vSAN), open source CEPH or Dell / EMC’s VxFlex-based block storage service for scale out SANs.

As CDN services expand beyond caching content, performance becomes more important. In fact, by 2020, 60 percent of CDN providers will not charge for optimization of web traffic, but only for value added services. CDN providers are deepening product offerings to include digital brand enhancement, web application delivery, and API protection, streaming media and artificial intelligence services.

With performance being a key overriding competitive distinction for CDN, delivering lower latency and faster time to first bit (TTTFB) metrics are differentiators for CDN optimization. Additional bandwidth capability may also be used to support higher quality services or more simultaneous users.

The I/O Optimized PowerEdge R640 is designed with the small footprint and high storage density to provide advanced CDN functions that require more performance at the edge. The I/O Optimized PowerEdge R640 is a natural fit in a CDN architecture made up of edge, core, and cloud components.

The I/O Optimized PowerEdge R640 setup used for performance testing was designed to mimic CDN environments. The server was outfitted with NGINX Application Platform, a web server software suite that features content caching services. NGINX is used both in the cache node and in the origin server. WRK was used to generate the traffic for these experiments. For these experiments, 5 MB objects were used to mimic chunks of 1080P video being streamed to the clients. In order to achieve the necessary bandwidth to exercise dual 100 GbE, two systems were used to run WRK and test execution was controlled with parallel-ssh. The test setup requires four nodes (systems), each executing specific tasks:

- **Edge Cache Node**: I/O Optimized PowerEdge R640 is the content server which stores data supporting web operations.
- **Content Origin**: This node acts as a web server, which grants the test generator servers access to the edge cache server.
- **Test Generator Servers 1 and 2**: These nodes are configured as load generator servers.

![Figure 1. Only three steps are needed for an operation performed locally with balanced storage and network.](image)
Figure 2. Servers used in CDN performance tests of I/O Optimized PowerEdge R640.

The edge cache nodes were outfitted with ten 4 TB Intel® Solid State Drives Data Center P4510 in each of the drive bays, two 100 GbE Intel® Ethernet Controller E810 cards as well as two SATA M.2 modules with RAID for boot. This is an extension over the standard PowerEdge R640, which supports one low profile network card with 10 NVMe drives and two SATA M.2 drives for boot.

Figure 3. Servers used in CDN performance tests of Standard PowerEdge R640.

The performance reported in tables 1 and 2 is the average of six runs for latency and bandwidth. However, the 99th percentile (P99) latency is the worst case of the six runs on each client.

The network bandwidth stated in tables 1 and 2 was captured using Prometheus and graphed via Grafana. The WRK bandwidth reported is the sum of the bandwidths reported by each traffic generator. Similarly, the average latency reported is the average across the two traffic generators.
The tests were performed on the I/O Optimized PowerEdge R640 configuration and the standard PowerEdge R640, which has known performance and is used as a benchmark. It is not optimized for storage I/O and so the tests are not intended to be a direct comparison for all use cases.

The test process utilized both a bare metal server configuration and a Docker containerized configuration. The hardware test setup process is common to both test methodologies; however, there are differences in the software configuration.

In the tests, the container-based applications were much smaller in size than the bare metal application, and changes were able to be made to each module without having to rebuild the entire application. The containerized test forces the application to only use specific CPU, memory, network, and disk resources. For the Standard R640, these restrictions to only use cores, memory and disks on a socket end up reducing performance. However, these same restrictions are used on the I/O Optimized PowerEdge R640 to guarantee NUMA local CPU, memory, network, and disk resources and provide significant increases to bandwidth and a large reduction in average latency.

**Test Results**

The tables below show the results of the bare metal and containerized tests conducted on both the I/O Optimized PowerEdge R640 and the standard PowerEdge R640.

In the non-containerized testing, the I/O Optimized PowerEdge R640 showed up to 24 percent increase in average network throughput, and up to 24 percent increase in WRK bandwidth as seen in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>AVERAGE NETWORK THROUGHPUT (Gb/S)</th>
<th>WRK BANDWIDTH (GB/S)</th>
<th>AVERAGE LATENCY* (MS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard R640</td>
<td>99.40</td>
<td>11.47</td>
<td>54.30</td>
</tr>
<tr>
<td>I/O Optimized</td>
<td>123.93</td>
<td>14.33</td>
<td>66.48</td>
</tr>
</tbody>
</table>

*Table 1. Bare metal NGINX CDN Performance*

In the containerized testing, the I/O Optimized PowerEdge R640 demonstrated more than twice the network throughput, more than twice the WRK bandwidth and almost half of the latency compared to this configuration of PowerEdge R640.

<table>
<thead>
<tr>
<th></th>
<th>AVERAGE NETWORK THROUGHPUT (Gb/S)</th>
<th>WRK BANDWIDTH (GB/S)</th>
<th>AVERAGE LATENCY* (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard R640</td>
<td>87.93</td>
<td>10.08</td>
<td>54.85</td>
</tr>
<tr>
<td>I/O Optimized</td>
<td>182.90</td>
<td>21.06</td>
<td>30.77</td>
</tr>
</tbody>
</table>

*Table 2. Containerized NGINX CDN Performance*

As configured for these tests, one socket of the Standard R640 will be able to serve 20 percent of the traffic without crossing Intel UPI, while the other socket will have 80 percent of the drives, but will have to cross Intel UPI to access the NIC. On the other hand, with the I/O Optimized R640, each socket operates largely independent from the other, and is able to serve 100 percent of the traffic it receives without crossing Intel UPI. This leads to a significant improvement in bandwidth and a reduction in latency.

The results of the containerized tests show the performance potential of an I/O balanced system combined with an application that has been made NUMA aware, in this case by using containers instead of making code changes to NGINX.

For the I/O Optimized R640, reducing cross socket traffic with containerization leads to up to a 47 percent improvement in bandwidth, 53 percent reduction in average latency.

If a comparison is conducted on the most effective way to use each system, bare metal for the Standard R640 and containerized for the I/O Optimized R640, the I/O Optimized R640 with containerization offers up to an 83 percent increase in bandwidth, a 43 percent reduction in average latency, and a 26 percent reduction in worst case P99 latency over the bare metal Standard R640.

*Latency = Time to last byte, as seen from the client side*
Conclusion

With growth in streaming media, CDN providers are challenged to deliver more streams at reduced cost. With the industry transitioning to 100 Gb networks and higher throughput, having a fully balanced CDN server platform is critical to maximizing performance and achieving QoS. In server configurations that do not feature evenly balanced storage access between CPUs, the time it takes for the CPUs to communicate with each other and pass information can slow performance, impacting the CDN quality of service (QoS) and the customer’s quality of experience (QoE). Dell Technologies is dedicated to accelerating the delivery of innovative technologies to large scale customers. With this configuration of the I/O Optimized PowerEdge R640, the percentage of communication between the CPUs is greatly reduced, resulting in higher overall performance.

<table>
<thead>
<tr>
<th></th>
<th>AVG. NETWORK THROUGHPUT (Gb/S)</th>
<th>WRK BANDWIDTH (GB/S)</th>
<th>AVERAGE LATENCY* (ms)</th>
<th>WORST CASE P99 LATENCY* (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard R640 Config</td>
<td>99.40</td>
<td>11.47</td>
<td>54.30</td>
<td>456.61</td>
</tr>
<tr>
<td>(bare metal software configuration)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I/O Optimized R640</td>
<td>182.90</td>
<td>21.06</td>
<td>30.77</td>
<td>334.17</td>
</tr>
<tr>
<td>(containerized)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Optimal Software Configuration NGINX CDN Performance

Learn More

For sales inquiries, contact Dell: www.delltechnologies.com/en-us/contactus.htm
For R640 information: www.dell.com/en-us/work/shop/povw/poweredge-r640
Dell is a member of the Intel® Network Builders program: networkbuilders.intel.com
Intel® Xeon® Scalable Processors: www.intel.com/xescalable
Intel® Ethernet Controller E810: www.intel.com/ethernet
Intel® SSD Data Center Family: www.intel.com/datacenterssd

Notices & Disclaimers

1 Tests conducted by Intel in April 2020: The dual-socket Dell I/O Optimized PowerEdge R640 system under test configuration featured two 28-core Intel Xeon Gold 6258R CPUs (microcode: 0x5000029) with Intel® Hyper-Threading Technology and Intel® Turbo Boost Technology both turned on. System memory totaled 768 GB comprising 24 32 GB DDR modules operating at 2933 MHz. Storage was 10 4.0 TB Intel® Solid State Drives Data Center P4510. Two 100 GbE Intel® Ethernet Controllers E810 provided network connections and the Intel® C621 Chipset was used as a platform hub. The dual-socket Dell R640 system under test configuration was configured with the same features, except it utilized a single 100 GbE Intel® Ethernet Controller E810 and the microcode for its two 28-core Intel Xeon Gold 6258R CPUs was 0x500002c.

The client servers and origin server configuration for the tests featured a 28-core Intel Xeon Platinum 8176 CPU (microcode: 0x2000050) with Intel Hyper-Threading Technology and Intel Turbo Boost Technology both turned on. System memory totaled 384 GB, comprising 12 32GB DDR modules operating at 2666 MHz. Storage was one 400 GB Intel® Solid State Drive. One 100 GbE Mellanox MCX516A-CCAT provided network connectivity.

Software stack: Operating system was Ubuntu 19.04 (kernel: 5.3.7-050307-generic); Workload & version: Wrk 4.1.0-4-gd2efada-dirty; Compiler: gcc version 8.3.0 (Ubuntu 8.3.0-6ubuntu1); Libraries: Glibc 2.29-0ubuntu2, Libssl-dev 1.1.1b-1ubuntu2.4, libnuma 2.0.12-1. Containerized testing used Docker version 19.03.3 build a872fc2f66. Edge cache utilized NGINX 1.16.1 (cache size: 30TB) and the origin server utilized NGINX 1.15.6. Data generation tool was Web ATS_gen.py (modified for 5MiB)

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit www.intel.com/benchmarks.

Performance results are based on testing as of dates shown in configurations (April 2020) and may not reflect all publicly available updates. See backup for configuration details. No product or component can be absolutely secure.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.