System z Outperforms Intel in Test of I/O and Virtualization Capabilities

October 2013 Jon Nolting, Roberto Calderon, John J. Thomas,

Superior I/O capabilities translate into more efficient IT infrastructure for handling real-world workloads. Generally people who work with both mainframe and distributed platforms acknowledge that System z hardware has superior I/O capabilities. The question is, what's the quantification? What exactly is the gap in the I/O capabilities between these systems? Are Intel-based systems good enough these days? Many distributed systems experts argue that the servers themselves do not play a big role in I/O capabilities anymore, and that the real distinction is the storage systems attached to those servers. Fans of distributed systems argue that the use of enterprise class storage systems with SSD storage, attached via high-speed fiber will provide equivalent I/O capabilities.

However, one IBM study attempts to quantify the I/O superiority of System z compared to Intel-based servers. IBM just concluded apples-to-apples technical comparisons in which both System z and Intel systems were configured with the same enterprise class back-end storage and were driven with same I/O load.

Unique Design

Efficient transaction processing requires a balance of capabilities—between CPU processing capacity and I/O bandwidth. It's obvious that if a platform is CPU rich but I/O starved, throwing more CPU at a workload won't help. Over many generations, System z engineers have consistently given careful thought to its design—making sure it has sufficient I/O bandwidth to support the CPU capacity on the machine.

System z has a unique, dedicated I/O subsystem that delivers high I/O bandwidth. The zEnterprise EC12 (zEC12) offers up to 16 dedicated system assist processors (SAPs) in addition to the general and specialty processors. I/O requests are handled by the SAPs, leaving the general processors free to do useful work. The 16 SAPs can sustain up to 2.4 million I/O operations per second (IOPS), with a recommended sustained rate of 1.7 million IOPS.

With the latest PCIe Gen2 interconnects, I/O subsystem speed is 8GBytes/second. In addition, the frame can support hundreds of physical FICON cards that provide unmatched bandwidth to back-end storage systems. The channel subsystem itself features capabilities like dynamic channel path management, automatic failover, end-to-end data integrity checking, in-band measurements, etc. System z High Performance FICON (zHPF) technology further improves I/O rates and service times. In addition to the channel system capabilities, z/OS Workload Management (WLM) efficiently manages all system resources from processors to back-end storage. The I/O Priority Manager in DS8000 storage systems works with z/OS WLM. All of these enterprise features work together to provide a truly one-of-a-kind I/O capability.

Configurations

The study takes a look at basic performance comparisons without getting into the advanced I/O capabilities. The goal was to see if there's a difference in I/O performance when both System z and Intel-based systems were configured with the same enterprise class back-end storage and subjected to the same I/O load.

As shown in Figure 1, the study used a 40-core Westmere EX Intel system running Linux and a z/OS LPAR with eight cores on a zEC12. Both servers were connected using four 8Gb paths to identical enterprise storage systems. Both the Intel-based system and System z platform were allocated volumes backed by eight SSDs each.

An application developed in-house was used to drive I/O load on both systems. It consisted of C-language programs that could be ported and compiled on both platforms. The programs ran in a loop, with each pass through the loop creating a 10MB file, making a copy and then reading the new file, resulting in two read-and-write operations per loop. The buffer, block and record size values on both systems were set to ensure the same number of I/O interrupts were being generated on both systems. To drive up load, these programs were run "n-wide," that is, n copies of the application could be run in parallel. Various metrics, such as CPU and disk usage, were collected with the key metric for comparison being elapsed time.

Observations

Resource Messaging Facility (RMF) reports provide comprehensive performance data for System z. RMF reports tracking the activity of CPU, channel paths, I/O queuing, DASD, etc., provide an excellent view of what's going on in the System z I/O subsystem. With the Intel-based system, Linux tools were used to gather performance data. On the Intel-based system, the average CPU percent was captured with a break down of average I/O wait percentage, total read-and-write operations, average time for an I/O to complete, disk drive module average percentage utilization, etc.

On both systems, IBM testers could drive the load up to a 96-wide configuration (96 I/O load applications running in parallel) before the available back-end storage started to become a bottleneck. For both systems, the key metric captured was elapsed time to complete the set of I/O workloads that were launched. Figure 2 shows a comparison of elapsed times between the two systems.

With identical enterprise class storage, identical high-speed connectivity and identical I/O load, the Intel-based system takes 3.6 times longer to complete the I/O workload. It should be noted that at the 96-wide point where this comparison was made (the point to which the tests could scale with the amount of back-end storage available for the study), the System z channel utilization was 27 percent (with four paths configured) and the SAP processors were at 2.5 percent utilization. In addition, the 16 SAPs in a zEC12 can handle up to 2.4 million IOPS, leaving System z with a tremendous amount of I/O capacity left for handling additional load.

The two systems have different numbers of cores and different CPU utilization percentages. As such, to get a true sense of the margin, the elapsed time against CPU was normalized, as shown in Figure 3.

The Impact of Virtualization

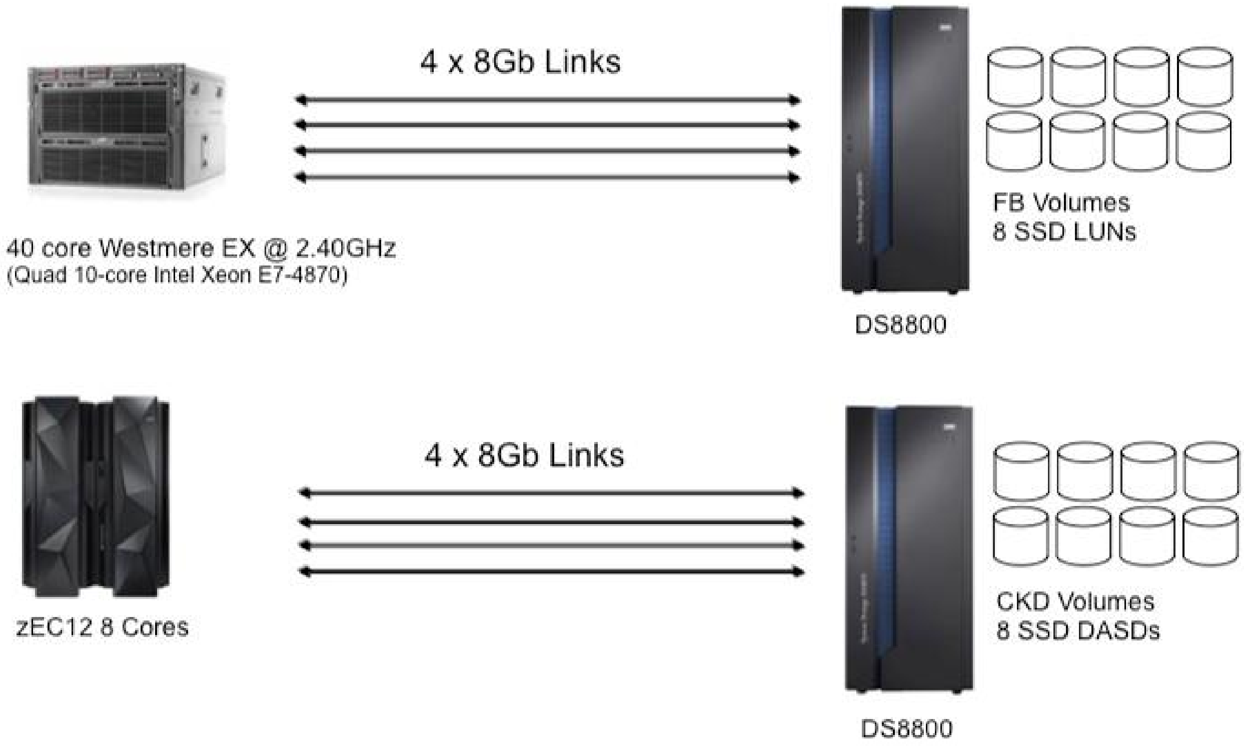
In the next step, the impact of virtualization was examined. The z/OS LPAR is already running virtualized on PR/SM, with shared CPs and shared I/O. For the Intel-based system, a popular x86 hypervisor was installed to create a virtual machine on which the 96-wide load was run. A degradation in workload performance was observed on the virtualized Intel-based system—namely, a 1.38-times increase in elapsed time for the same workload. Figure 4 shows the final result, comparing three deployment options:

- z/OS LPAR
- Bare metal Intel-based system
- Virtualized Intel-base system

Clear Comparison

The study demonstrates that regardless of back-end storage device and connectivity features, the choice of server platform does matter when it comes to handling I/O load. The System z platform demonstrated clear superiority over Intel-based systems in handling heavy I/O load. Further, with sufficient storage devices to support back-end volumes, the subsystem design allows the mainframe to scale and support many more workloads.

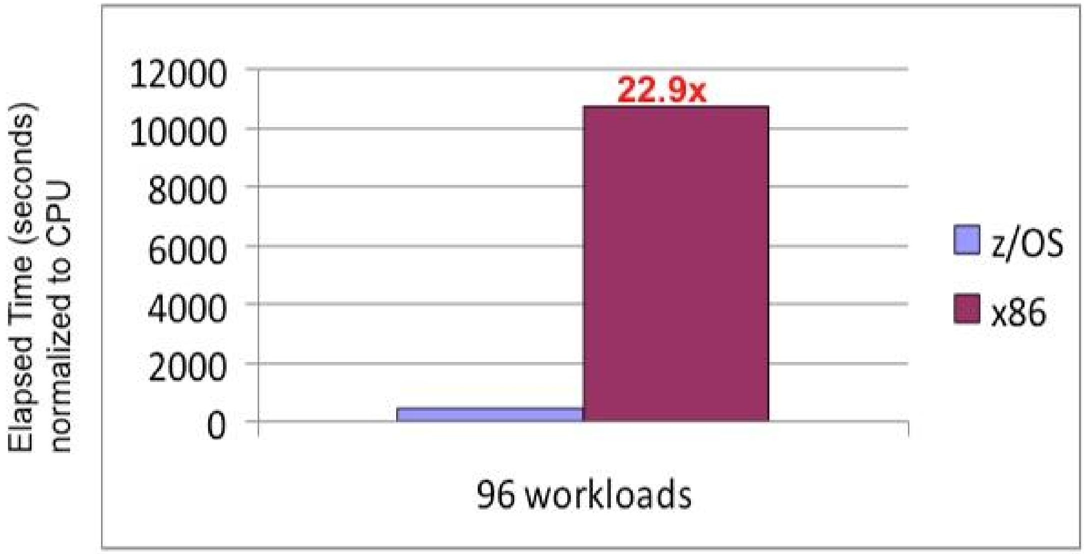
Test Configuration



Elapsed Time Comparison

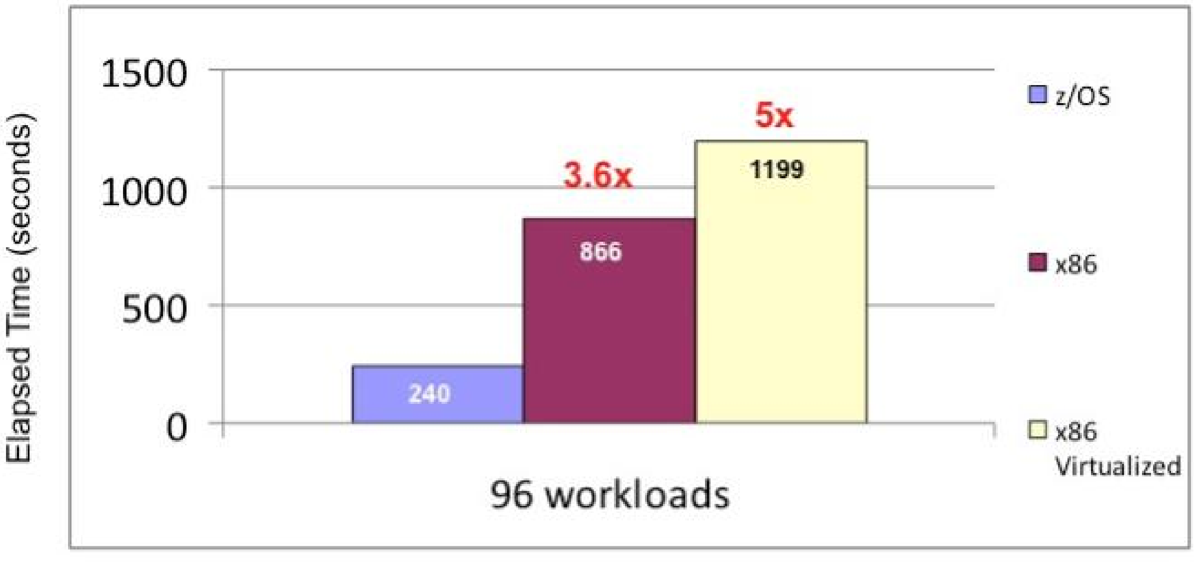
Note: 96-wide was the maximum point to which the test cases could be run before the back-end storage available for this study became a bottleneck. System z channel utilization is at 27 percent and the IOPs are only at 2.5 percent, leaving considerable room to grow.

Elapsed Time Normalized to CPU



Note: 96-wide was the maximum point to which the test cases could be run before the back-end storage available for this study became a bottleneck. System z channel utilization is at 27 percent and the IOPs are only at 2.5 percent, leaving considerable room to grow.

Compare Impact of Virtualization



Note: 96-wide was the maximum point to which the test cases could be run before the back-end storage available for this study became a bottleneck. System z channel utilization is at 27 percent and the IOPs are only at 2.5 percent, leaving considerable room to grow.