Skylake memory study

Authors: Joseph Stanfield, Garima Kochhar and Donald Russell, Bruce Wagner.
HPC Engineering and System Performance Analysis Teams, HPC Innovation Lab, January 2018

To function efficiently in an HPC environment, a cluster of compute nodes must work in tandem to compile complex data and achieve desired results. The user expects each node to function at peak performance as an individual system, as well as a part of an intricate group of nodes processing data in parallel. To enable efficient cluster performance, we first need good single system performance. With that in mind, we evaluated the impact of different memory configurations on single node memory bandwidth performance using the STREAM benchmark. The servers used here support the latest Intel Skylake processor (Intel Scalable Processor Family) and are from the Dell EMC 14th generation (14G) server product line.

Less than 6 DIMMS per socket

The Skylake processor has a built-in memory controller similar to previous generation Xeons but now supports *six* memory channels per socket. This is an increase from the four memory channels found in previous generation Xeon E5-2600 v3 and E5-2600 v4 processors. Different Dell EMC server models offer a different number of memory slots based on server density, but all servers offer at least one memory module slot on each of the six memory channels per socket.

For applications that are sensitive to memory bandwidth and require predictable performance, configuring memory for the underlying architecture is an important consideration. For optimal memory performance, all six memory channels of a CPU should be populated with memory modules (DIMMs), and populated identically. This is a called a balanced memory configuration. In a balanced configuration all DIMMs are accessed uniformly and the full complement of memory channels are available to the application. An unbalanced memory configuration will lead to lower memory performance as some channels will be unused or used unequally. Even worse, an unbalanced memory configuration can lead to **_unpredictable memory performance_** based on how the system fractures the memory space into multiple regions and how Linux maps out these memory domains.
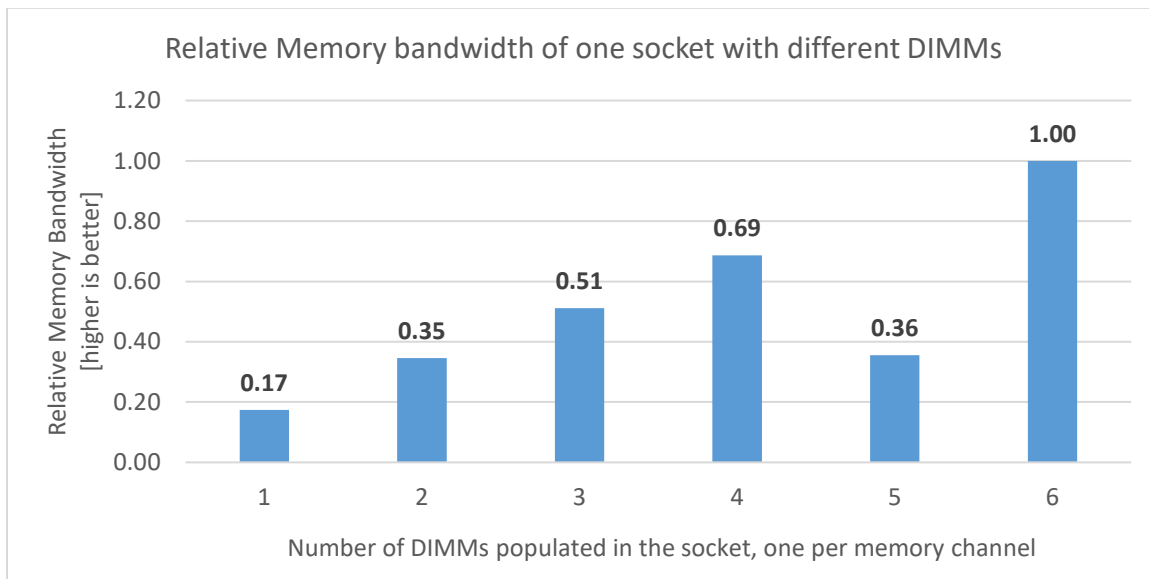

Figure 1. **Relative memory bandwidth with different number of DIMMs on one socket. PowerEdge C6420. Platinum 8176. 32 GB 2666 MT/s DIMMs.**

Figure 1 shows the drop in performance when all six memory channels of a 14G server are not populated. Using all six memory channels per socket is the best configuration, and will give the most predictable performance. This data was collected using the Intel

Xeon Platinum 8176 processor. While the exact memory performance of a system depends on the CPU model, the general trends and conclusions presented here apply across all CPU models.

Balanced memory configurations

Focusing now on the recommended configurations that use all 12 memory channels in a two socket 14G system, there are different memory module options that allow different total system memory capacities. Memory performance will also vary depending on whether the DIMMs used are single ranked, double ranked, RDIMMS or LR-DIMMs. These variations are, however, significantly lower than any unbalanced memory configuration as shown in Figure 2.

- 8GB 2666 MT/s memory is single ranked and have lower memory bandwidth than the dual ranked 16GB and 32GB memory modules.

- 16GB and 32GB are both dual ranked and have similar memory bandwidth with 16G DIMMs demonstrating higher memory bandwidth.

- 64GB memory modules are LR-DIMMS are have slightly lower memory bandwidth than the dual ranked RDIMMS.

128GB memory modules are also LR-DIMMS but are lower performing than the 64GB modules and their prime attraction is the additional memory capacity. Note that LR-DIMMs also have higher latency and higher power draw. Here is an older study on previous generation 13G servers that describes these characteristics in detail.

Memory Bandwidth with diff. system memory capacities - 12 R- and LR-DIMMs

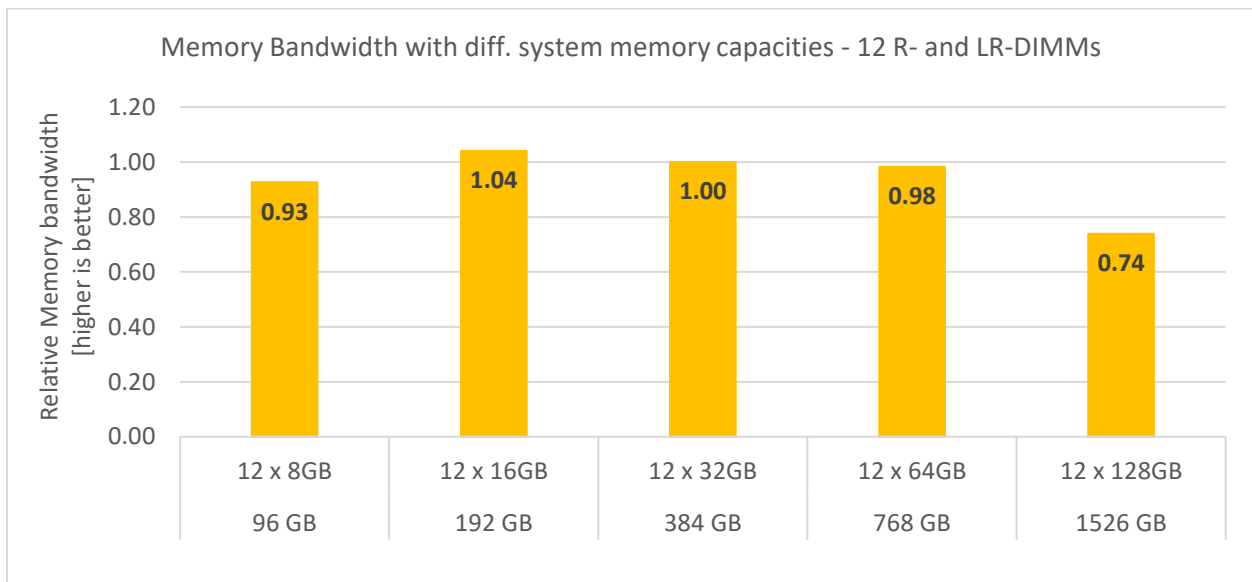|	12 x 8GB	12 x 16GB	12 x 32GB	12 x 64GB	12 x 128GB
Relative Memory bandwidth	0.93	1.04	1.00	0.98	0.74
	96 GB	192 GB	384 GB	768 GB	1526 GB

Figure 2.   Relative memory bandwidth for different system capacities (12D balanced configs). PowerEdge R740. Platinum 8180. DIMM configuration as noted, all 2666 MT/s memory.

Data for Figure 2 was collected on the Intel Xeon Platinum 8180 processor. As mentioned above, the memory subsystem performance depends on the CPU model since the memory controller is integrated with the processor, and the speed of the processor and number of cores also influence memory performance. The trends presented here will apply across the Skylake CPUs, though the actual percentage differences across configurations may vary. For example, here we see the 96 GB configuration has 7% lower memory bandwidth than the 384 GB configuration. With a different CPU model, that difference could be 9-10%.

Figure 3 shows another example of balanced configurations, this time using 12 or 24 identical DIMMs in the 2S system where one DIMM per channel is populated (1DPC with 12DIMMs) or two DIMMs per channel are populated (2DPC using 24 DIMMs). The information plotted in Figure 3 was collected across two CPU models and shows the same patterns as Figure 2. Additionally, the following observations can be made:

- With two 8GB single ranked DIMMs giving two ranks on each channel, some of the memory bandwidth lost with 1DPC SR DIMMs can be recovered with the 2DPC configuration.

- 16GB dual ranked DIMMS perform better than 32GB DIMMs in 2DPC configurations too.

We also measured the impact of this memory population when 2400 MT/s memory is used, and the conclusions were identical to those for 2666 MT/s memory. For brevity, the 2400 MT/s results are not presented here.
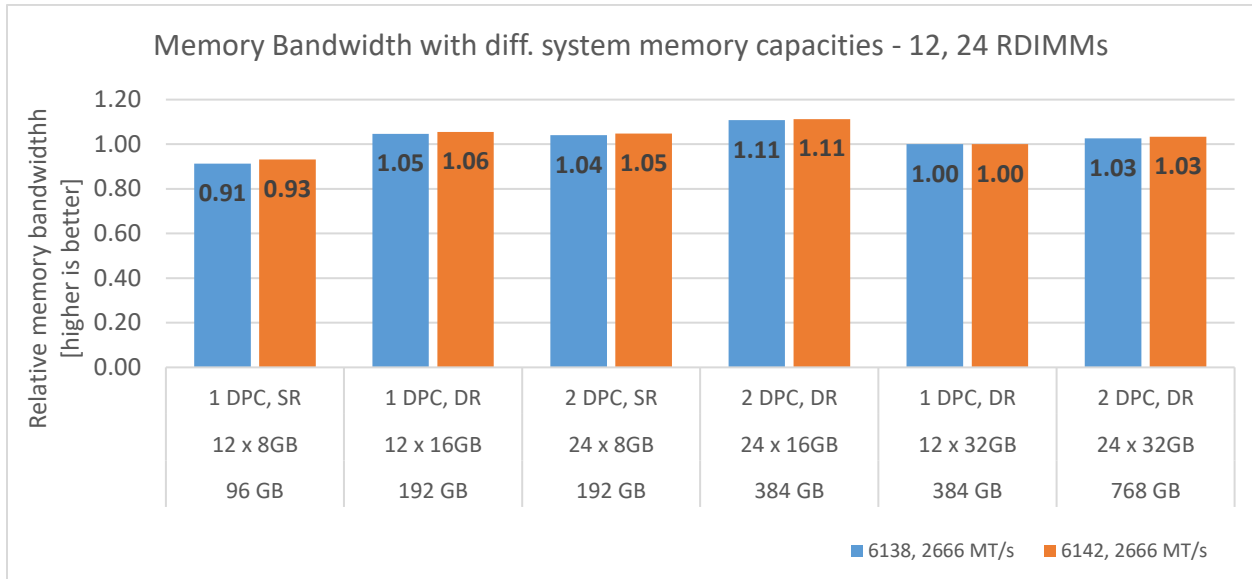

Figure 3.   Relative memory bandwidth for different system capacities (12D, 24D balanced configs). PowerEdge R640. Processor and DIMM configuration as noted. All 2666 MT/s memory.

Unbalanced memory configurations

In previous generation systems, the processor supported four memory channels per socket. This led to balanced configurations with eight or sixteen memory modules per dual socket server. Configurations of 8x16GB (128 GB), 16x16 GB or 8x32GB (256 GB), 16x32 GB (512 GB) were popular and recommended.

With 14G and Skylake, these absolute memory capacities will lead to unbalanced configurations as these memory capacities do not distribute evenly across 12 memory channels. A configuration of 512 GB on 14G Skylake is possible but suboptimal, as shown in Figure 4. Across CPU models (Platinum 8176 down to Bronze 3106), *there is a 65% to 35% drop in memory bandwidth when using an unbalanced memory configuration* when compared to a balanced memory configuration! The figure compares 512 GB to 384 GB, but the same conclusion holds for 512 GB vs 768 GB as Figure 2 has shown us that a balanced 384 GB configuration performs similarly to a balanced 768 GB configuration.
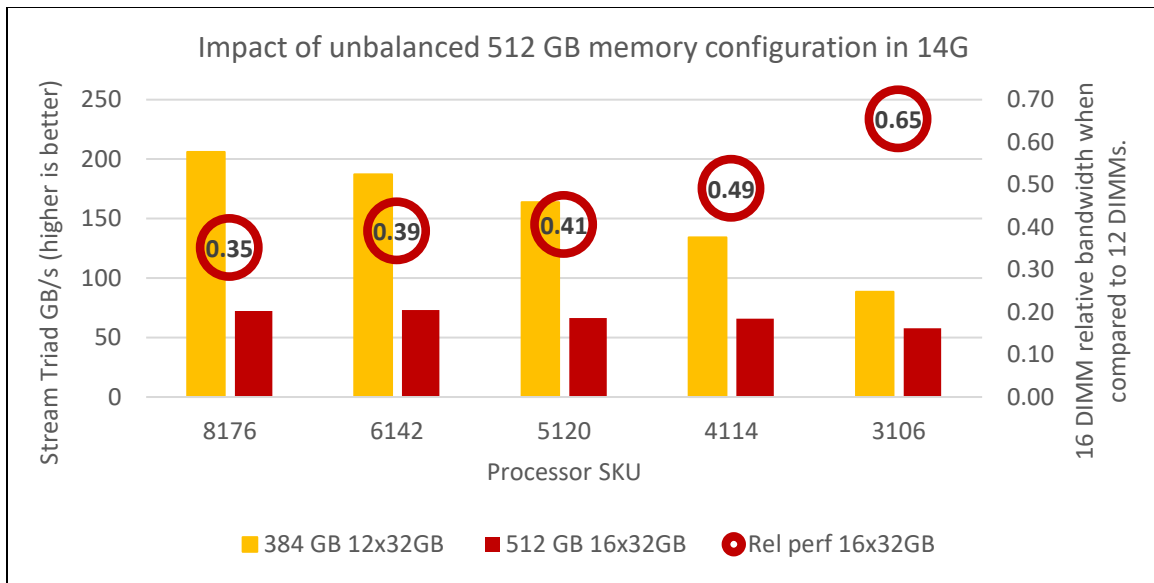
Figure 4. **Impact of unbalanced memory configurations.**
PowerEdge C6420. Processor and DIMM configuration as noted. All 2666 MT/s memory.

Near-balanced memory configurations

The question that arises is - Is there a reasonable configuration that would work for capacities close to 256GB without having to go all the way to a 384GB configuration, and close to 512GB without having to raise the capacity all the way to 768GB?

Dell EMC systems do allow mixing different memory modules, and this is described in more detail in the server owner manual. For example, the Dell EMC PowerEdge R640 has 24 memory slots with 12 slots per processor. Each processor's set of 12 slots is organized across 6 channels with 2 slots per channel. In each channel, the first slot is identified by the white release tab while the second slot tab is black. Here is an extract of the memory population guidelines that permit mixing DIMM capacities.

> THE POWEREDGE R640 SUPPORTS FLEXIBLE MEMORY CONFIGURATION, ENABLING THE SYSTEM TO BE CONFIGURED AND RUN IN ANY VALID CHIPSET ARCHITECTURAL CONFIGURATION. DELL EMC RECOMMENDS THE FOLLOWING GUIDELINES TO INSTALL MEMORY MODULES:
>
> RDIMMS AND LRDIMMS MUST NOT BE MIXED.
>
> POPULATE ALL THE SOCKETS WITH WHITE RELEASE TABS FIRST, FOLLOWED BY THE BLACK RELEASE TABS.
>
> WHEN MIXING MEMORY MODULES WITH DIFFERENT CAPACITIES, POPULATE THE SOCKETS WITH MEMORY MODULES WITH THE HIGHEST CAPACITY FIRST. FOR EXAMPLE, IF YOU WANT TO MIX 8 GB AND 16 GB MEMORY MODULES, POPULATE 16 GB MEMORY MODULES IN THE SOCKETS WITH WHITE RELEASE TABS AND 8 GB MEMORY MODULES IN THE SOCKETS WITH BLACK RELEASE TABS.
>
> MEMORY MODULES OF DIFFERENT CAPACITIES CAN BE MIXED PROVIDED OTHER MEMORY POPULATION RULES ARE FOLLOWED (FOR EXAMPLE, 8 GB AND 16 GB MEMORY MODULES CAN BE MIXED).
>
> MIXING OF MORE THAN TWO MEMORY MODULE CAPACITIES IN A SYSTEM IS NOT SUPPORTED.
>
> IN A DUAL-PROCESSOR CONFIGURATION, THE MEMORY CONFIGURATION FOR EACH PROCESSOR SHOULD BE IDENTICAL. FOR EXAMPLE, IF YOU POPULATE SOCKET A1 FOR PROCESSOR 1, THEN POPULATE SOCKET B1 FOR PROCESSOR 2, AND SO ON.
>
> POPULATE SIX MEMORY MODULES PER PROCESSOR (ONE DIMM PER CHANNEL) AT A TIME TO MAXIMIZE PERFORMANCE.

*One important caveat is that 64 GB LRDIMMs and 128 GB LRDIMMs cannot be mixed; they are different technologies and are not compatible.

So the question is, how bad are mixed memory configurations for HPC? To address this, we tested valid "near-balanced configurations" as described in Table 1, with the results displayed in Figure 5.

Table 1.   Near balanced memory configurations

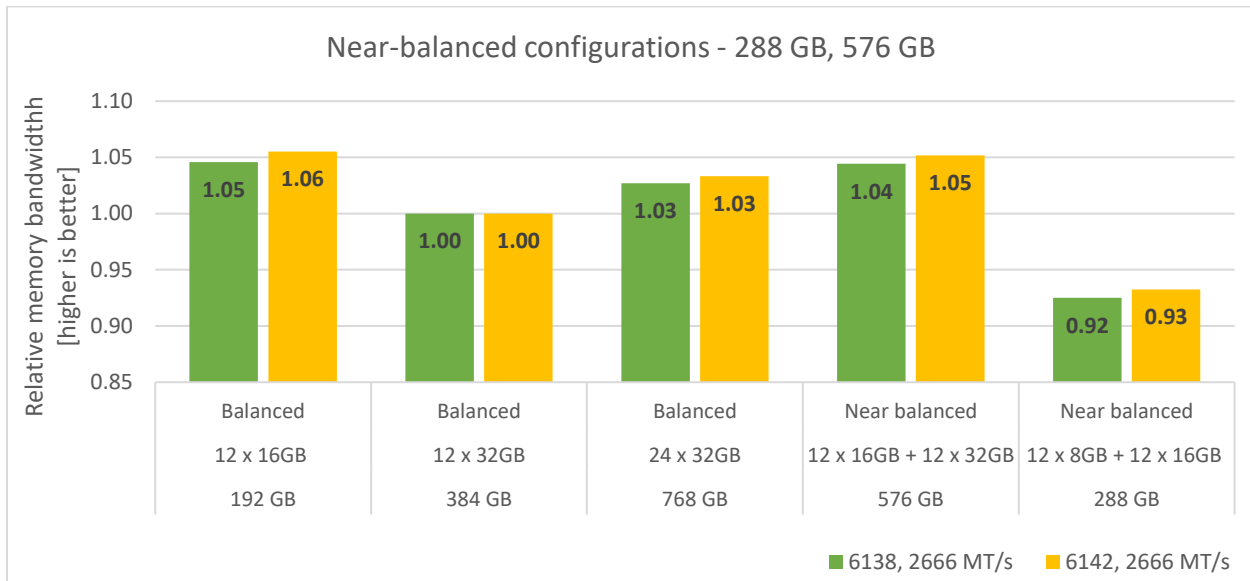| Name	Memory Capacity	CPU 1 memory configuration	CPU 2 memory configuration
Balanced 192 GB	192 GB	6x 16GB DR DIMMS	6x 16GB DR DIMMS
Balanced 384 GB	384 GB	6x 32GB DR DIMMS	6x 32GB DR DIMMS
Near balanced 288 GB	288 GB	6x 16GB DR DIMMS (DIMM1) 6x 8GB SR DIMMS (DIMM2)	6x 16GB DR DIMMS (DIMM1) 6x 8GB SR DIMMS (DIMM2)
Near balanced 576 GB	576 GB	6x 32GB DR DIMMS (DIMM1) 6x 16GB DR DIMMS (DIMM2)	6x 32GB DR DIMMS (DIMM1) 6x 16GB DR DIMMS (DIMM2)
Balanced 768 GB	768 GB	12x 32GB DR DIMMS	12x 32GB DR DIMMS

Figure 5.   Impact of near-balanced configurations.
PowerEdge R640. Processor and DIMM configuration as noted. All 2666 MT/s memory.

Figure 5 illustrates that near-balanced configurations are a reasonable alternative when the memory capacity requirements demand a compromise. All memory channels are populated, and this helps with the memory bandwidth. The 288 GB configuration uses single ranked 8GB DIMMs and we see the penalty single ranked DIMMS impose on the memory bandwidth.

Conclusion

The memory subsystem performance differences with balanced vs. unbalanced configurations and with different types of memory modules is not new to Skylake or Dell EMC servers. Previous studies for previous generations of servers and CPUs are listed below and show similar conclusions.

- Memory should ideally be populated in a balanced configuration with all memory channels populated and populated identically for best performance. The number of memory channels are determined by the CPU and system architecture.

- DIMM rank, type and memory speed have an impact on performance.

- Unbalanced configurations are not recommended when optimizing for performance

- Some near-balanced configurations are a reasonable alternative when the memory capacity requirements demand a compromise

Previous memory configuration studies

1) [Performance and Energy Efficiency of the 14th Generation Dell PowerEdge Servers](#) – Memory bandwidth results across Intel Skylake Processor Family (Skylake) CPU models (14G servers)

2) http://en.community.dell.com/techcenter/extras/m/white_papers/20444326/download

3) [13G PowerEdge Server Performance Sensitivity to Memory Configuration](#) – Intel Xeon 2600 v3 and 2600 v4 systems (13G servers)

4) [Unbalanced Memory Performance](#) – Intel Xeon E5-2600 and 2600 v2 systems (12G servers)

5) [Memory Selection Guidelines for HPC and 11G PowerEdge Servers](#) – Intel Xeon 5500 and 5600 systems (11G servers)

D&LLEMC