

Application Performance Study on Intel Haswell EX Processors

by Ashish Kumar Singh

This blog describes, in detail, the performance study carried out on the E7-8800 v3 family of processors (architecture codenamed as Haswell-EX). The performance on Intel Xeon E7-8800 v3 has been compared to Intel Xeon E7-4800 v2 to ascertain the generation over generation performance improvement. The applications used for this study are HPL, STREAM, WRF and ANSYS Fluent. The Intel Xeon E7-8890v3 processors have 18 cores/36 threads with 45MB of L3 cache (2.5MB/slice). With AVX workloads the clock speed of Intel E7-8890 v3 reduced from 2.5GHz to 2.1GHz. These processors support QPI speed of 9.6 GT/s.

Server Configuration

	PowerEdge R920	PowerEdge R930
Processor	4 x Intel Xeon E7-4870v2 @ 2.3GHz (15 cores) 30MB L3 cache 130W	4 x Intel Xeon E7- 8890v3 @2.5GHz (18 cores) 45MB L3 cache 165W
Memory	512GB = 32 x 16GB DDR3 @ 1333MHz RDIMMS	1024 GB = 64 x 16GB DDR4 @1600MHz RDIMMS
BIOS Settings		
BIOS	Version 1.1.0	Version 1.0.9
Processor Settings > Logical Processors	Disabled	Disabled
Processor Settings > QPI Speed	Maximum Data Rate	Maximum Data Rate
Processor Settings > System Profile	Performance	Performance
Software and Firmware		
Operating System	RHEL6.5 x86_64	RHEL 6.6 x86_64
Intel Compiler	Version 14.0.2	Version 15.0.2
Intel MKL	Version 11.1	Version 11.2

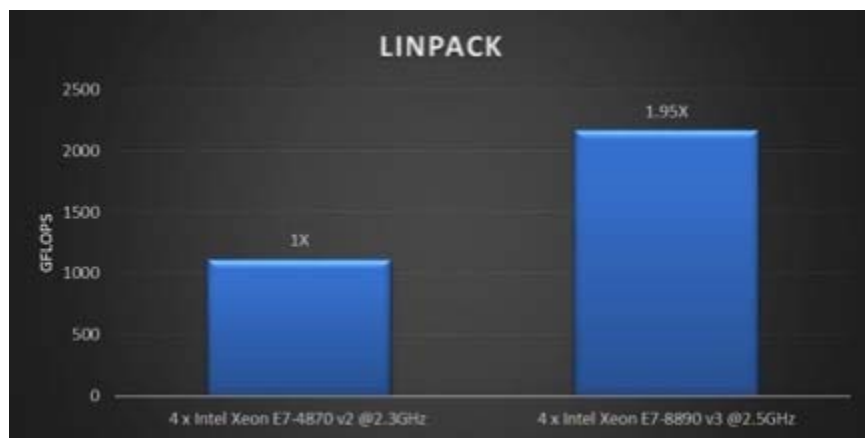
Intel MPI	Version 4.1	Version 5.0
Benchmark and Applications		
LINPACK	V2.1 from MKL 11.1	V2.1 from MKL 11.2
STREAM	v5.10, Array Size 1800000000, Iterations 100	v5.10, Array Size 1800000000, Iterations 100
WRF	v3.5.1, Input Data Conus12KM, Netcdf-4.3.1.1	V3.6.1, Input Data Conus12K, Netcdf-4.3.2
ANSYS Fluent	v15, Input Data: eddy_417k, truck_poly_14m, sedan_4m, aircraft_2m	v15, Input Data: eddy_417k, truck_poly_14m, sedan_4m, aircraft_2m

Analysis

The objective of this comparison was to show the generation-over-generation performance improvement in the enterprise 4S platforms. The performance differences between two server generations were because of the improvement in system architecture, greater number of cores and higher frequency memory. The software versions were not a significant factor.

LINPACK

High Performance LINPACK is a benchmark that solves a (random) dense linear system in double precision (64 bits) arithmetic on distributed memory systems. HPL benchmark was run on both PowerEdge R930 and PowerEdge R920 with block size of NB=192 and problem size of N=90% of total memory size.



As shown in the graph above, LINPACK showed 1.95X performance improvement with four Intel Xeon E7-8890 v3 processors on R930 server in comparison to four Intel Xeon E7-4870 v2 processors on R920 server. This was due to substantial increase in number of cores, memory speed, flop/second of the processor and processor architecture.

STREAM

STREAM is a simple synthetic program to measure sustained memory bandwidth used COPY, SCALE, SUM and TRIAD programs to measure memory bandwidth.

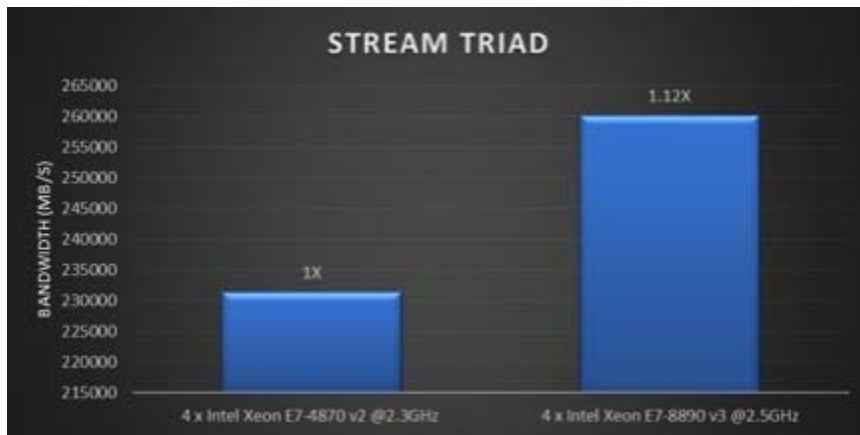
Operations of these programs are shown below:

COPY: $a(i) = b(i)$

SCALE: $a(i) = q*b(i)$

SUM: $a(i) = b(i) + c(i)$

TRIAD: $a(i) = b(i) + q*c(i)$

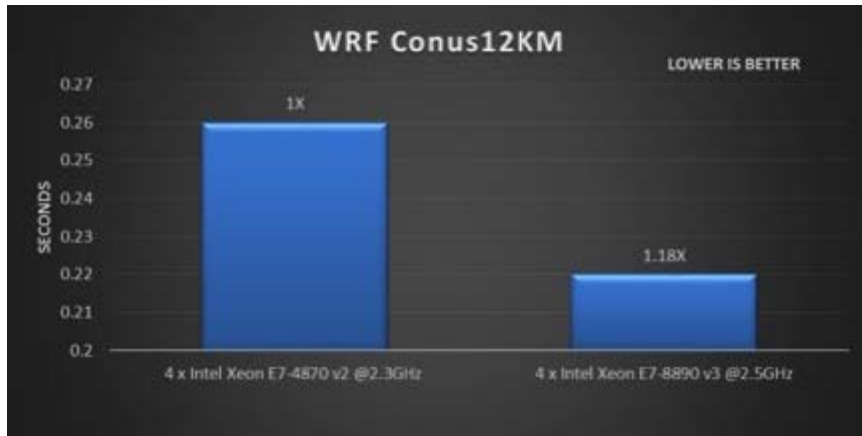


This chart showed the comparison of sustained memory bandwidth between PowerEdge R920 and PowerEdge R930 servers. STREAM showed 231GB/s on PowerEdge R920 and 260GB/s on PowerEdge R930, which is 12% improvement in memory bandwidth. This increase is because of the improvement in DIMM speed available on PowerEdge R930.

WRF

The WRF (Weather Research and Forecasting) model is a next-generation mesoscale numerical weather prediction system designed to serve both atmospheric research and operational forecasting needs. The model serves a wide range of metrological applications across scales from tens of meters to thousands of kilometers. WRF allows to generate atmospheric simulations based on real data (observations, analysis) or idealized conditions.

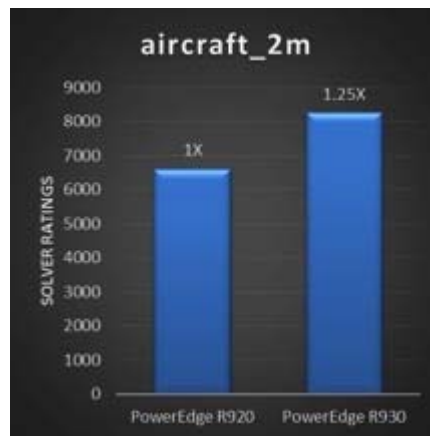
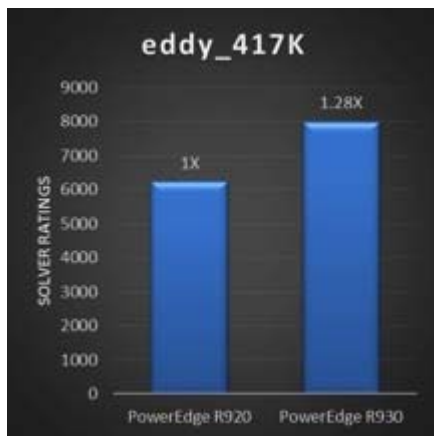
WRF performance analysis was run for conus12KM dataset. Conus12KM data is a single domain, medium size 48-hours 12KM resolution case over continental US (CONUS) domain with a time step of 72seconds.

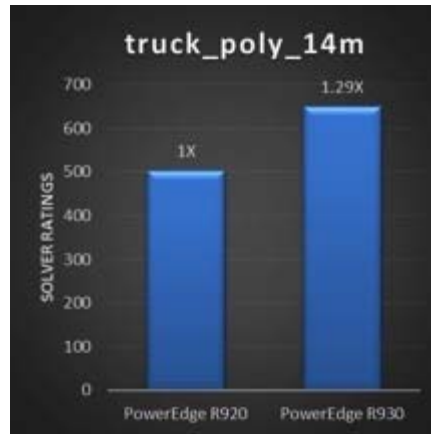


With Conus12KM dataset, WRF showed 0.22seconds average time on PowerEdge R930 server, while 0.26seconds on PowerEdge R930 server, which is an 18% improvement.

ANSYS Fluent

ANSYS Fluent contains the broad physical modeling capabilities for model flow, turbulence, heat transfer, and reactions for industrial applications ranging from air flow over an aircraft wing to combustion in a furnace, from bubble columns to oil platforms, from blood flow to semiconductor manufacturing, and from clean room design to wastewater treatment plants.





We used four different datasets for Fluent. We considered 'Solver rating' (higher is better) as the performance metric. For all the test cases with PowerEdge R930 Fluent showed 24% to 29% performance improvement in-comparison to PowerEdge R920.

Conclusion

PowerEdge R930 server outperforms its previous generation PowerEdge R920 server in both benchmarks and application comparison. Due to latest processors with higher number of cores, higher frequency memory and CPU architecture improvement PowerEdge R930 gave better performance than PowerEdge R920. PowerEdge R930 platform with four Intel Xeon EX processors is very good choice for those HPC applications, which can scale up to the large number of cores and memory.

Reference

http://en.community.dell.com/techcenter/high-performance-computing/b/general_hpc/archive/2014/05/21/hpc-application-performance-study-on-4s-srvers