HPC Application Performance Study on 4S Servers

by Ranga Balimidi, Ashish K. Singh, and Ishan Singh

What can you do with a big bad 4-socket machine with 60 cores with up to 6TB memory in HPC? To help answer that question, we conducted a performance study using several benchmark suites such as HPL, STREAM, WRF and Fluent. This blog describes some of our results that help illustrate the possibilities. The server that we used for this study is the Dell PowerEdge R920. This server supports the family of processors in the Intel architecture code named Ivy Bridge EX.

The server configuration table outlines the configuration details used for this study as well as the configurations from a previous study performed in June 2010 with the previous generation of technology. We use these two systems to compare performance across technology refresh.

Server Configuration

Power Edge R920 Hardware	
Processors	4 x **Intel Xeon E7-4870v2** @ 2.3GHz (15 cores) 30M cache 130W
Memory	512 GB =32 * 16GB 1333MHz RDIMMs
PowerEdge R910 Hardware	
Processor	4 x **Intel Xeon X7550** @ 2.00GHz (8 cores) 18M cache 130W
Memory	128GB = 32 * 4GB 1066MHz RDIMMs
Software and Firmware for PowerEdge R920	
Operating System	Red Hat Enterprise Linux 6.5 (kernel version 2.6.32-431.el6 x86_64)
Intel Compiler	Version 14.0.2
Intel MKL	Version 11.1
Intel MPI	Version 4.1
BIOS	Version 1.1.0
BIOS Settings	System Profile set to Performance (Logical Processor disabled, Node Interleave disabled)
Benchmarks & Applications for PowerEdge R920	
HPL	v2.1, From Intel MKL v11.1, Problem size 90% of total memory.

Stream	v5.10, Array Size 1800000000, Iterations 100
WRF	v3.5.1, Input Data Conus 12K, Netcdf-4.3.1.1
Fluent	v15, Input Data: eddy_417k, truck_poly_14m, sedan_4m, aircraft_2m

Results and Analysis

For this study, we compared the two servers across the four benchmarks described below.

The aim of this comparison is to show the generation-over-generation changes in this four socket platform. Each server was configured with the optimal software and BIOS configurations at the time of the measurements. The biggest difference in performance between the two server generations is the improvement in system architecture, greater number of cores, and memory speed. The software versions are not a significant factor.

STREAM

The STREAM benchmark is a simple synthetic program that measures sustained memory bandwidth in MB/s. It uses COPY, SCALE, SUM and TRIAD programs to evaluate the 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: $a(i) = b(i) + q*c(i)$

The chart below compares STREAM performance results from this study with results from previous the generation. In this study, STREAM yields 231GB/s memory bandwidth which is twice the memory bandwidth measured from the previous study. This increase is because of the improvement in the number of memory channels and DIMM speed.
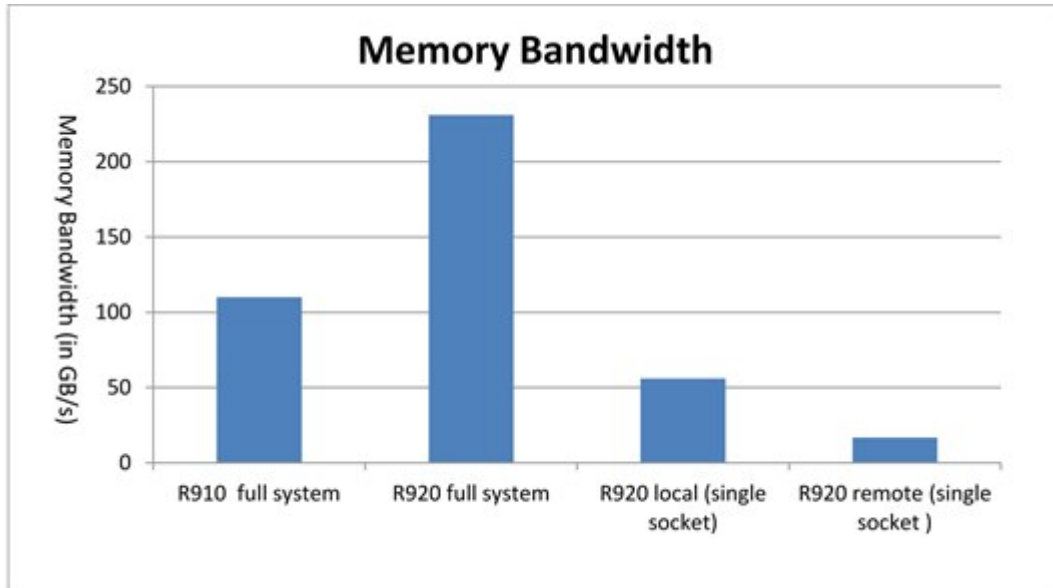
Memory Bandwidth

The graph also plots the local bandwidth and remote memory bandwidth. Local memory bandwidth is measured by binding processes to a socket and accessing only memory local to that socket (NUMA enabled, same NUMA node). Remote memory bandwidth is measured by binding processes to one socket and only accessing memory that is remote to that socket (remote NUMA node) where it has to go through QPI link to access this memory. The remote memory bandwidth is 72% lower than the local memory bandwidth due to the limitation of QPI link bandwidth.

LINPACK

The Linpack benchmark measures how fast a computer solves linear equations and measures a system's floating-point computing power. It requires a software library for performing numerical linear algebra on digital computers; for this study we used Intel's Math Kernel Library. The following chart illustrates results from a single server HPL performance benchmark.

HPL
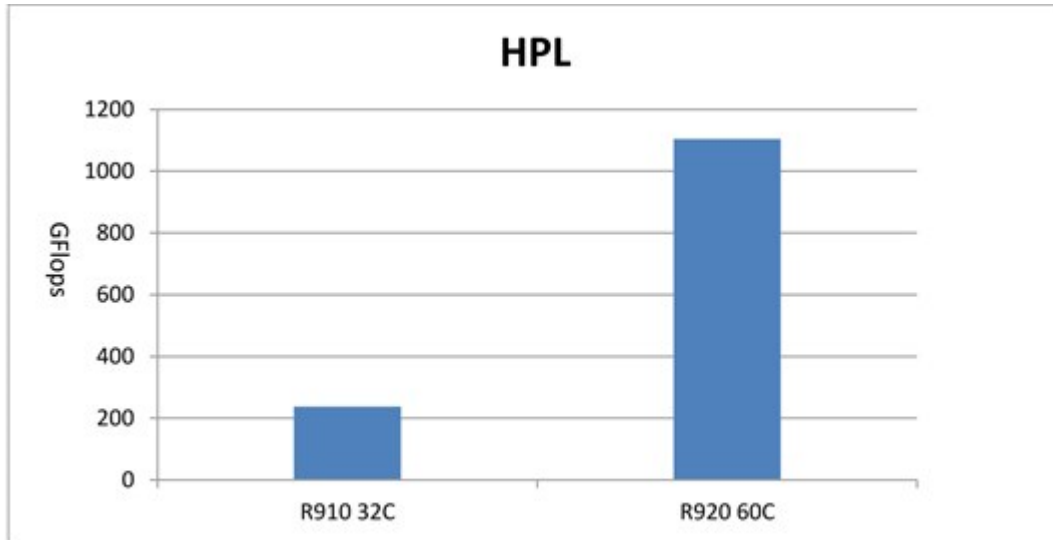
HPL yielded 4.67x sustained performance improvement in this study. This is primarily due to the substantial increase in the number of cores, increase in the FLOP/cycle of the processor and the overall improvement in the processor architecture.

WRF

The Weather Research and Forecasting (WRF) Model is a numerical weather prediction system designed to serve atmospheric research and weather forecasting needs. It features two dynamical cores, a data assimilation system, and a software architecture allowing for parallel computation and system extensibility.
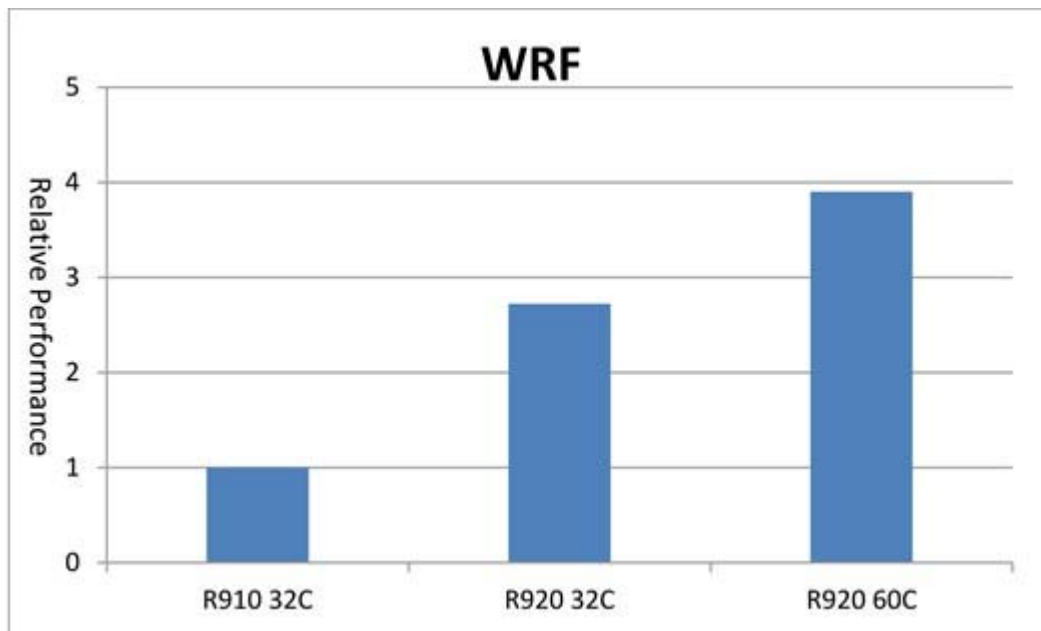
We have taken the average time step as the metric to measure WRF performance. We used Conus 12km data set for this application.

In the graph above we've plotted the WRF performance results from this study relative to results from the previous generation. Since there is an increase in number of cores on Intel E7-4870 v2 processor, we have scaled up WRF to 60 cores and observed significant performance increase while scaling. Matching the number of cores used on both platforms at 32 cores, we observed significant performance improvement (2.9x) over the previous generation platform. When using the full capability of the server at 60c there is an additional 35% improvement. When it comes to server-to-server comparison, the PowerEdge R920 performs ~4x better than PowerEdge R910. This is due to the overall architecture improvements including processor and memory technology.

ANSYS FLUENT

Ansys Fluent contains the broad physical modeling capabilities needed to 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.

In the charts below, we have plotted the performance results from this study relative to results from the previous generation platform.

eddy_417k

aircraft_2m

sedan_4m

truck_poly_14m
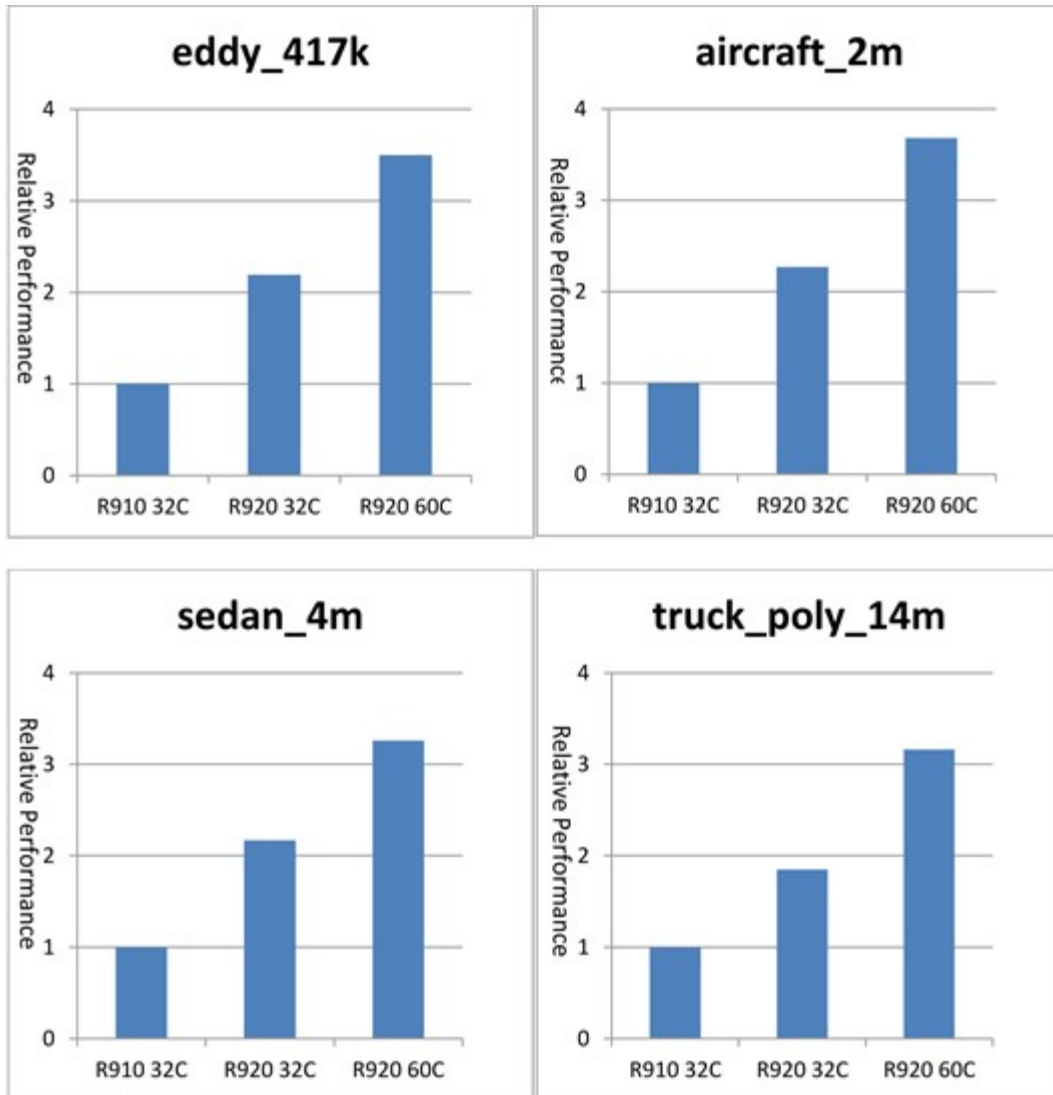
We've used four input data sets for Fluent. We've considered "Solver Rating" (higher is better) as the performance metric for these test cases.

For all the test cases, Fluent scaled very well with 100% CPU utilization. Comparing generation-to-generation, for 32 core-to-core comparisons we observed the R920 performance results are approximately 2x better over the previous generation in all the test cases. When it comes to server-to-server comparison using all available cores, it performs 3-3.5x better.

These results were gathered by explicitly setting processor affinity at the MPI level. To do this, the following two configuration options were used:

```
$export HPMPI_MPIRUN_FLAGS="-aff=automatic"
```

```
$cat ~/.fluent

(define (set-affinity argv) (display "set-affinity disabled"))
```

Conclusion

The PowerEdge R920 server outperforms its previous generation server in both benchmarks and applications comparisons studied in this exercise. The platform has its advantage over the previous generation platform in terms of latest processor support, increased memory speed and capacity support, and overall system architecture improvements. This platform is a good choice for HPC applications, which can scale-up with the high processor core count support (up to 60 cores) and large shared memory support (up to 6TB). It is also a great choice for memory intensive applications considering the large memory support.