Original Link: http://www.anandtech.com/show/9567/the-power-8-review-challenging-the-intel-xeon-

The IBM POWER8 Review: Challenging the Intel Xeon

49 Comments

by Johan De Gelas on November 6, 2015 8:00 AM EST

Posted in CPUs IT Computing Enterprise enterprise CPUs IBM POWER POWER8

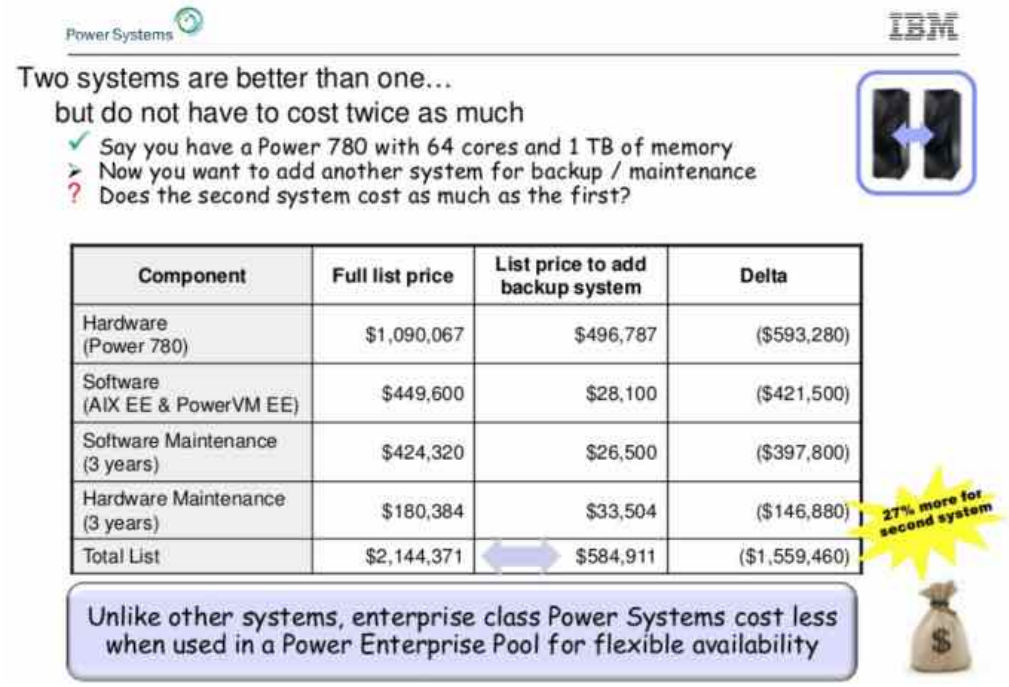
Five years. That is how much time has passed since we have seen an affordable server processor that could keep up with or even beat Intel's best Xeons. These days no less than 95% of the server CPUs shipped are Intel Xeons. A few years ago, it looked like ARM servers were going to shake up the market this year, but to cut a long story short, it looks like the IBM POWER8 chip is probably the only viable alternative for the time being.

That was also noticeable in our Xeon E7 review, which was much more popular than we ever hoped. One of the reasons was the inclusion of a few IBM POWER8 benchmarks. We admit that the article was however incomplete: the POWER8 development machine we tested was a virtual machine with only 1 core, 8 threads and 2 GB of RAM, which is not enough to do any thorough server testing.

After seeing the reader interest in POWER8 in that previous article, we decided to investigate the matter further. To that end we met with Franz Bourlet, an enthusiastic technical sales engineer at IBM and he made sure we got access to an IBM S822L server. Thanks to Franz and the good people of Arrow Enterprise Computing Solutions, Arrow was able to lend us an IBM S822L server for our testing.

A Real Alternative?

Some of you may argue that the POWER based servers have been around for years now. But the slide below illustrates what we typically associated IBM's POWER range with:



Proudly, the IBM sales team states that you can save 1.5 million dollars after you have paid them 2 million dollars for your high-end 780 system. There is definitely a market for such hugely expensive and robust server systems as high end RISC machines are good for about 50.000 clients. But frankly for most of us, those systems are nothing more than an expensive curiosity.

Availability can be handled by software and most of us are looking/forced to reduce our capital expenses rather than increase them. We want fast, "reliable enough" servers at low costs that are easy to service. And that is exactly the reason why the single and dual sockets Xeon servers have been so popular the past decade. Can an IBM POWER server be a real alternative to the typical Xeon E5 server? The short but vague answer: a lot has changed in the past years and months. So yes, maybe.

Challenging the Xeon

So what caused us to investigate the IBM POWER8 as a viable alternative to the mass market Xeon E5s and not simply the high-end quad (and higher) socket Xeon E7 parts? A lot. IBM sold its x86 server division to Lenovo. So there is only one true server processor left at IBM: the POWER family. But more importantly, the OpenPOWER fondation has a lot of momentum since its birth in 2013. IBM and the OpenPOWER Foundation Partners like Google, NVIDIA, and Mellanox are all committed to innovating around the POWER processor-based systems from the chip level up through the whole platform. The foundation has delivered some tangible results:

- *Open Firmware which includes both the firmware to boot the hardware (similar to the BIOS) ...*
- *... as OPAL (OpenPOWER Abstraction Layer) to boot and launch a hypervisor kernel.*
- *OpenBMC*
- *Cheaper and available to third parties (!) POWER8 chips*
- *CAPI over PCIe, to make it easier to link the POWER8 to GPUs (and other PCIe cards)*
- *And much more third party hardware support (Mellanox IB etc.)*
- *A much large software ecosystem (see further)*

The impact of opening up firmware under the Apache v2 license and BMC (IBM calls it "field processor") code should not be underestimated. The big hyperscale companies - Google, Amazon, Microsoft, Facebook, Rackspace - want as much control over their software stack as they can.

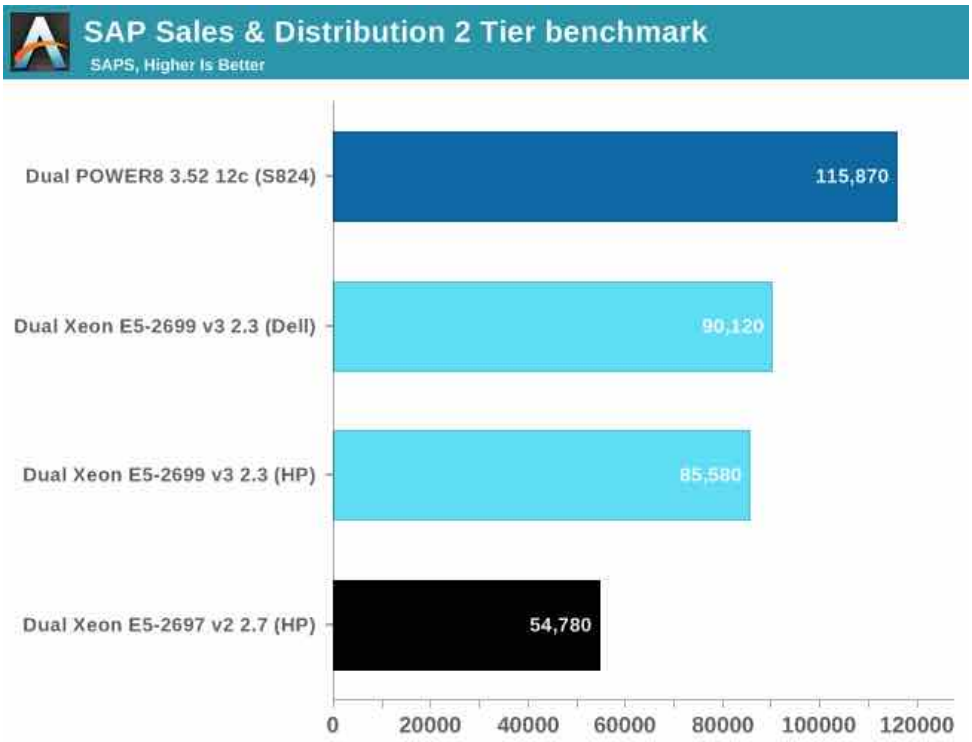
The resuls are that Google is supporting the efforts and Rackspace has even build their own OpenPOWER server called "Barreleye". While Google has been supportive and showing of proof of concepts, Rackspace is going all the way:

> ... and **aim to put Barreleye in our datacenters for OpenStack services early next year**.

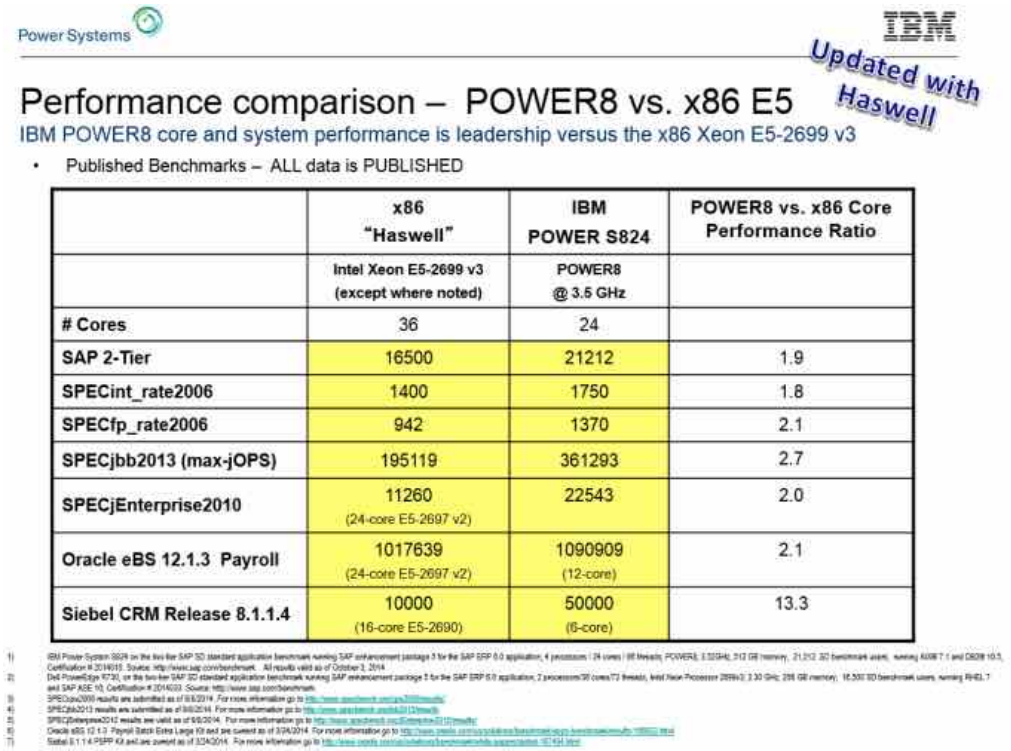
The end result is that the complete POWER platform, once only available in expensive high end servers, can now be found inside affordable linux based servers, from IBM (S8xxL) and third parties like Tyan. The opinions of usual pundits range from "too little, too late" to "trouble for Intel". Should you check out a POWER8 based server before you order your next Xeon - Linux server? And why? We started with analyzing the available benchmarks carefully.

Reading the Benchmarks

There are a lot of benchmarks available that compare the IBM POWER8 to Xeons. One example is the Enterprise Resource Planning (ERP) software SAP. We have used the Sales & Distribution 2 Tier benchmark many times because it is one of the very few benchmarks that is a very good representation of real world high-end enterprise workloads.



Now combine this with the benchmarks that IBM has compiled on their marketing slides and the fact that we know that the POWER8 chip has a TDP of 190W at nominal speed and 247W when running at "Turbo" clockspeeds.



	x86 "Haswell"	IBM POWER S824	POWER8 vs. x86 Core Performance Ratio
	Intel Xeon E5-2699 v3 (except where noted)	POWER8 @ 3.5 GHz	
# Cores	36	24	
SAP 2-Tier	16500	21212	1.9
SPECint_rate2006	1400	1750	1.8
SPECfp_rate2006	942	1370	2.1
SPECjbb2013 (max-jOPS)	195119	361293	2.7
SPECjEnterprise2010	11260 (24-core E5-2697 v2)	22543	2.0
Oracle eBS 12.1.3 Payroll	1017639 (24-core E5-2697 v2)	1090909 (12-core)	2.1
Siebel CRM Release 8.1.1.4	10000 (16-core E5-2690)	50000 (6-core)	13.3

It all seems very simple: the IBM POWER8 is a more power hungry chip but delivers much better performance. But as always you should take the time to read the benchmarks very closely. The IBM S824 is typically the one featured in the benchmarks. However, we are pretty sure that is not the system that will be able to sway the current Intel Xeon customers towards OpenPOWER. Nor are we convinced that the most widely reported benchmarks are accurately predicting the experience of those people.

There are three reasons for that. First of all, most of the benchmarks are run on **AIX (7)**, IBM's own proprietary UNIX. AIX is a high performance, extremely robust OS, but it does not have the rich software system and support that Linux has. Furthermore even with their common design elements, an excellent Linux administrator will have to invest some time to get the same level of expertise in AIX. But more importantly, the

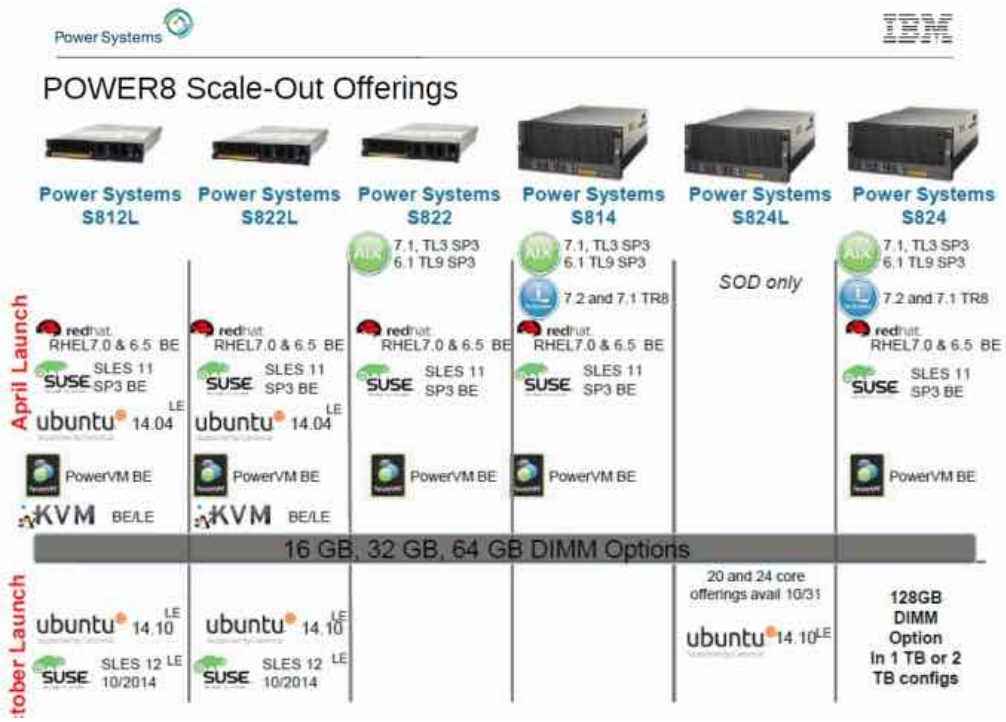
S824 is a pretty expensive machine, both in acquisition cost (starting at $21.000, up to $60.000 and more) and energy cost. That kind of pricing lands the system in hostile and more powerful quad Xeon E7 territory.

Lastly, the S824 uses two CPU cards or Dual Chip Modules (DCM), each containing two six-core POWER8 modules at 3.5 GHz. Now consider that the third party OpenPOWER servers have 190/247W TDP 10-core 3.4 GHz POWER8 CPUs. The power consumption does not increase linearly as you add more cores and higher clocks. So the CPU modules found inside the S824 are definitely more power hungry, probably well above 250W.

There is more. Take a look at IBM "Scale-out" server, the more affordable server range of IBM servers. First, a bit of IBM server nomenclature which is actually quite logical and easy to decipher (take note, Intel marketing).

- *S stands for "Scale-out"*
- *8 stands for POWER8*
- *1 or 2 is the number of sockets*
- *2 or 4 is the height, expressed in rack Us.*

So an S824 contains 2 sockets in a 4U chassis and a S812 is a one socket system. There is one designation left, : the "L" or Linux .



Notice that the non-L versions also support Linux, but a few months ago they supported only the Big Endian (BE) versions (the slide is from the beginning of this year). IBM told us that all POWER8 servers now support both Little Endian (LE) and BE Linux.

This is important since using an LE version (Ubuntu, SUSE) makes data migration from and data sharing (NAS, SAN) with an x86 system much easier, as x86 only supports LE.

Software Issues

So although the S824 is IBM's benchmark flagship for the scale-out range, the S812L and S822L are the servers that have the best chance at converting the kinds of users currently opt for x86 Xeons:

- *Support for Little Endian data*
- *Best Linux support (Suse, Redhat & Ubuntu)*
- *(Somewhat) lower power*
- *2U form factor which offers decent performance per U*
- *and probably the most important reason of all:* ***Affordable!*** *($10k-25k instead of $30-60k)*

So yes, the S822L looks like the first worthy alternative since 2010 for the dual Xeon servers. But the S822L

did not inherit all the strong points of the typical "Big Blue" servers. The clockspeeds are a bit lower to keep the power consumption in check, and more importantly the LE Linux support is still very young. Sure, POWERLinux has been around for ages, but the software ecosystem was mostly supporting a few Big Endian applications like heavy duty Java servers and SAP.

Let's make the issue at hand a bit more tangible. IBM offers a migration advisor that helps developers to port their applications. That is definitely a good thing, but it also clearly illustrates that building a software ecosystem is a lot more cumbersome than the POWERPoint slides let you believe. In case of IBM's LE Linux, porting the rich x86 Linux software ecosystem to OpenPOWER is not that straightforward:

- *Some code has inline x86 assembly such as thread resource locking code.*
- *Some code has x86 specific APIs*
- *No support for POWER in the make files which makes recompiling not straight forward*
- *POWER is 64 bit only.*

We have experienced ourselves that this was more than just theory.

Case in point: for X86-64 we simply installed well tuned, ready to run, pre compiled binaries. Benchmarking is pretty easy here with a minor scripting effort.

The story was very different on the IBM S822L. We installed Ubuntu 15.04 (3.19.0-15 - ppc64le). To satisfy our curiosity we did a quick benchmark run with Linux-Bench, an automated benchmarking tool that Ian also likes to use. The benchmark did almost nothing on our POWER system despite the fact that most of the software had some form of support for POWER based systems.

The same was true for most software out there: We had to port most of the software by delving deep in all kinds of config, Readme, and make files. In many cases, we had to search around for alternative libraries that did support OpenPOWER.

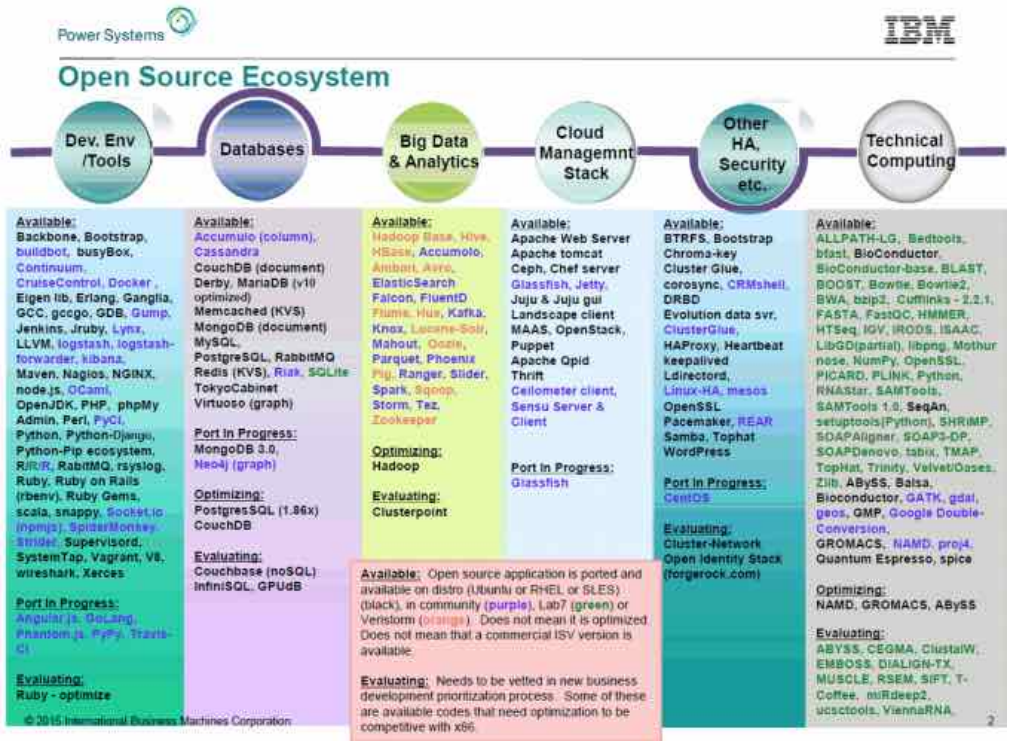
Although a lot of software had an entry for "IBM POWER" in the make files, we encountered a lot of trouble. The server nor IBM is to blame: it is simply a fact that most developers - especially those with HPC software - have put a lot more effort in optimizing and validating their Intel x86 version of their software than the more "exotic" platforms.

Linux Ecosystem Not at Full Throttle.. Yet

It is clear to us that the OpenPOWER Linux ecosytem is still young and as a result does not offer the same performance as the older PowerVM and AIX platforms. There is still quite a bit of performance headroom.

A good example is the crypto acceleration. The IBM POWER8 has a dedicated cryptographic unit supporting new POWER ISA instructions to accelerate AES (Encryption), SHA (Hashing), and CRC (Cyclic Redundancy Check) codes. A similar encryption unit was already available in the POWER7+ . We found out that an nx-crypto driver was available and part of the Linux 3.5 kernel. However, even though Ubuntu 15.04 LE for OpenPOWER is based upon the Linux kernel 3.19, the nx-crypto driver was nowhere to be found. You could argue that the same is true for Intel as they introduce new instructions, but as far as we could see, there was no encryption acceleration whatsoever possible, not even based upon the older POWER7+.

A few days after we have finished testing, we found out the vmx-crypto driver will be available in distributions using the Kernel 4.1 and later and will be enabled in OpenSSL 1.0.2 (currently 1.0.1f in the standard repositories). The slide below - found in a presentation given this month - show how fast the ecosystem is expanding but also that it is still in flux.



OpenPOWER gained traction in 2014, the POWER8 is the first POWER chip with LE support and the number of Linux servers on top of OpenPOWER systems is still very small compared to x86. It is pretty simple: it is a much smaller community than the x86 linux server community. According to "the platform", IBM claims that "scale-out POWER8 machines have seen double digit revenue growth in the first half of 2015" but those growth numbers are "against a very small base". That tells us a lot: it is indeed a very small community, but a quickly growing one.

Taking a Closer Look At IBM's S822L

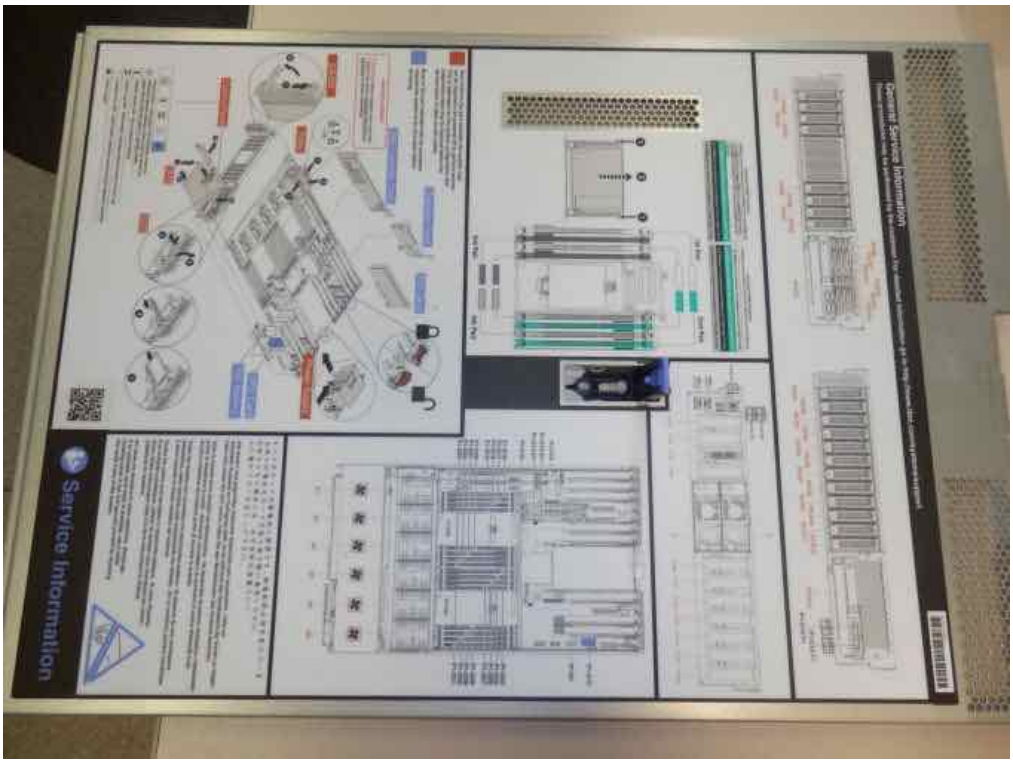
The S822L was mounted in the Xeon server dominated racks inside our experimental datacenter. The build quality of both the rails and server were apparent, a "locking mechanism" made sure the server was easy to mount without a screwdriver and was kept firmlyat its place.

The system booted by using the Flexible Service Processor (FSP), which is comparable to the Baseboard Management Controllers as they are both components that allow you to manage and monitor your server thanks to the IPMI specification. The main difference with Xeon system is that the FSP and its related firmware and software is the only way you can control your system. There is no "BIOS" screen or BIOS configuration setup, everything has to be configured and booted via the FSP software. You could say that the "BIOS" and "BMC management software" are now integrated into one central firmware.

To power on the S822L, you have to access the FSP using the open-source IPMItool. Once the server is

booted up, the "petitboot" bootloader of OPAL (the OpenPOWER Abstraction Layer firmware) takes over. It scans all bootable instances (disks, network, optical, etc.) for operating systems, and is similar to the GRUB bootloader. From there, you can install Linux like you would on an x86 system.

The cover was covered with a lot of interesting service information about the upgrading and replacing the hardware.



Once we removed the cover, lots of expansion slots became visible.

No less than nine hot plug (!) low profile PCIe Gen 3 slots are available. Four of them are x16, ready for some GPU action. Five are x8. Only one of the PCIe slots is used for the standard quad-gigabit Ethernet adapter. We also had one Emulex FC card installed.

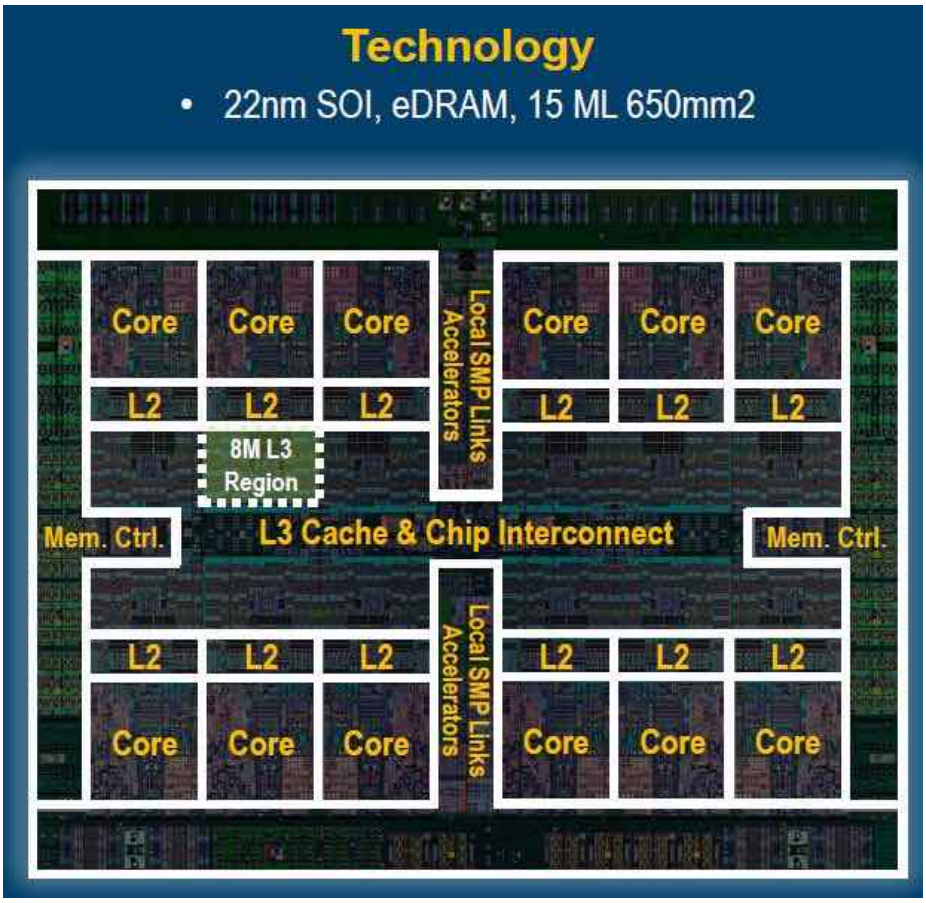
Also installed were two PowerPC based SAS RAID controller(s), capable of RAID-6 and all common RAID levels, which connectto a dual backplane that offers 12 Small Factor Form (2.5 inch) drives. These drives can be SAS SSD or hard disks, which is a reliable but rather expensive storage choice. A DVD drive was also present, which allowed us to install Linux the old-fashioned way.

At the back we find two hot-swappable PSUs, four gigabit Ethernet interfaces, two USB 2.0 ports, an HMC dual-gigabit interface (an HMC is a hardware applicance that can manage several IBM servers) and one system port.

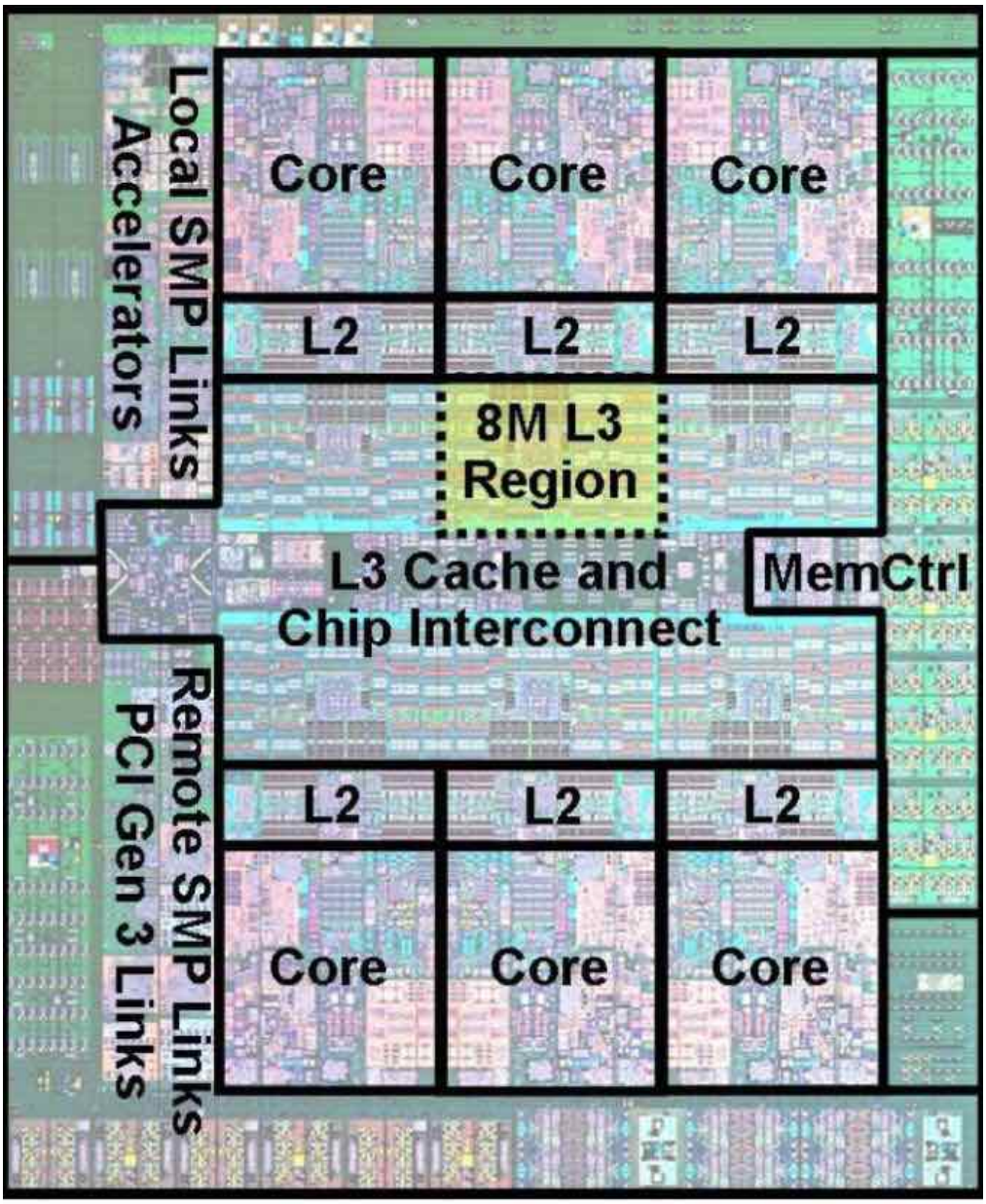
The server is powered by two redundant high quality Emerson 1400W PSUs.

Inside the S822L: Hardware Components

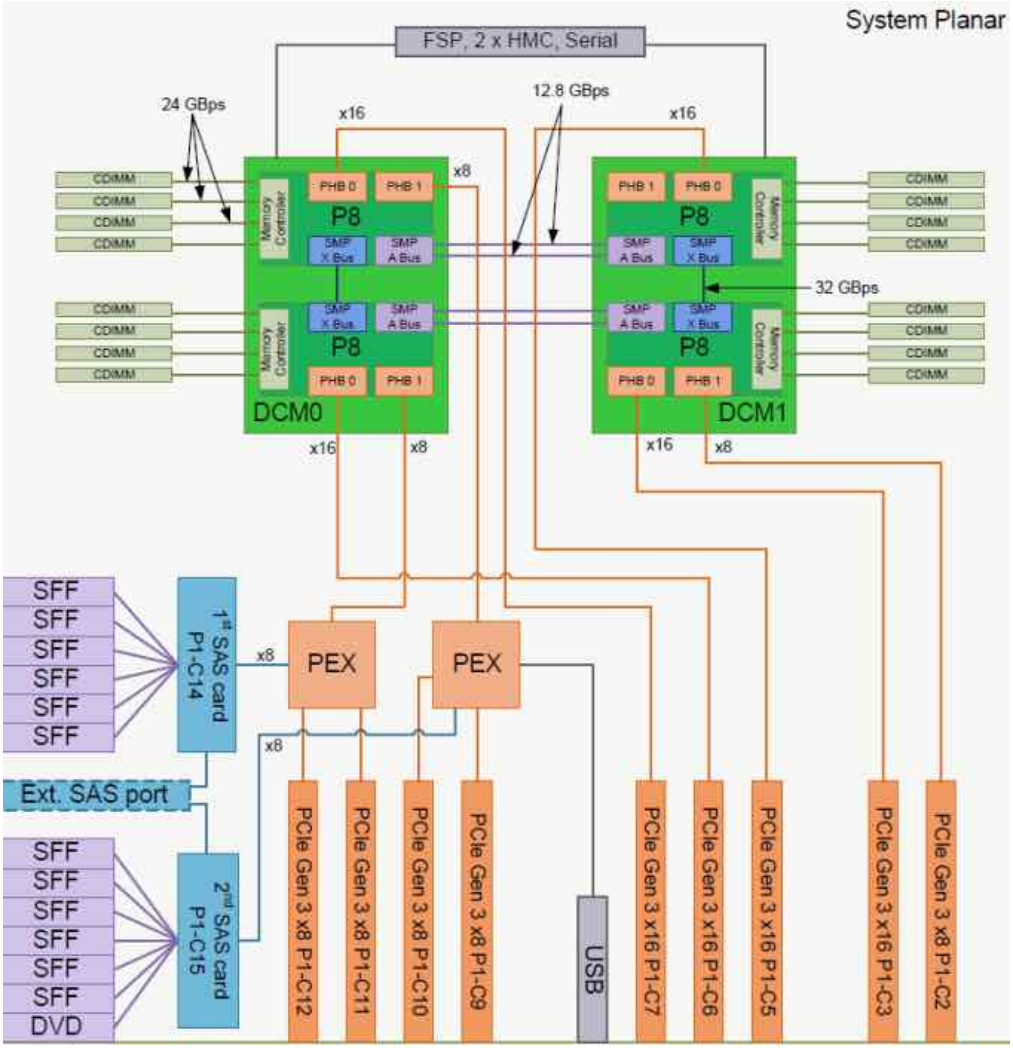
The 2U Rack-mount S822L server contains two IBM POWER8 DCM sockets. Each socket thus contains two cores connected by a 32GBps interconnect. The reason for using a Multi-Chip-Module (MCM) is pretty simple. Smaller five-to-six core dies are a lot cheaper to produce than the massive 650 mm² monolithic 12-core dies. As a result the latter are reserved for IBM's high-end (E880 and a like). So while most POWER8 presentations and news posts on the net talk about the multi-core die below...



... it is actually an MCM with two six core dies like the one below that is challenging the 10 to 18 core Xeons. The massive monolithic 10-12 core dies are in fact reserved for much more expensive IBM servers.



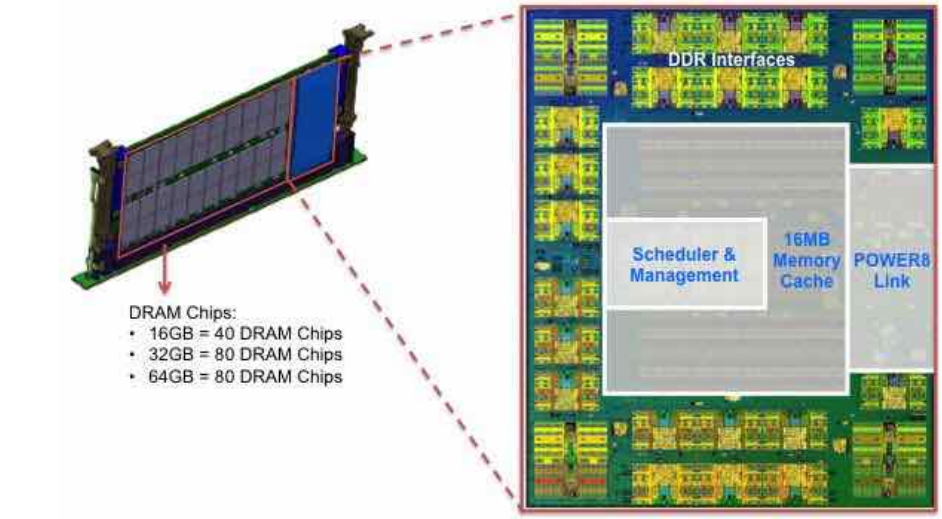
The layout of the S822L is well illustrated by the scheme inside the manual.



Each DCM offers 48 PCIe Gen 3 lanes. 32 of those lanes are directly connected to the processor while 16 connect to PCIe switches. The PCIe switches have "only" 8 lanes upstream to the DCM, but offer 24 lanes to "medium" speed devices downstream. As it unlikely that both your SAS controllers and your network controllers will gobble up the full PCIe x8 bandwidth, this is a very elegant way to offer additional PCIe lanes.

The L4-cache and Memory Subsystem

Each POWER8 memory controller has access to four "Custom DIMMs" or CDIMMs. Each CDIMMs is in fact a "Centaur" chip and 40 to 80 DRAM chips. The Centaur chip contains the DDR3 interfaces, the memory management logic and a 16 MB L4-cache.



The 16 MB L4-cache is eDRAM technology like the on-die L3-cache. Let us see how the CDIMMs look in reality.

Considering that 4Gb DRAM chips were available in mid 2013, the 1600 MHz 2Gb DRAM chips used here look a bit outdated. Otherwise the (much) more expensive 64GB CDIMMs use the current 4Gb DRAM chips. The S822L has 16 slots and can thus use up to 1TB (64GB x 16) in DIMMs.

Considering that many Xeon E5 servers are limited to 768 GB, 1 TB is more than competitive. Some Xeon E5 servers can reach 1.5 TB with 64 GB LR-DIMMs but not every server supports this rather expensive memory technology. It is very easy to service the CDIMMs: a gentle push on the two sides will allow you to slide them out. The black pieces of plastic between the CDIMMS are just place-holders that protect the underlying memory slots. For our testing we had CDIMMs installed in 8 of our system's 16 slots.

The Centaur chip acts as a 16MB L4-cache to save memory accesses and thus energy, but it needs quite a bit of power (10-20 W) itself and as a result is covered by heatsink. CDIMMs have ECC enabled (8+1 for ECC) and have also an extra spare DRAM chip. As result, a CDIMM has 10 DRAM chips while offering capacity of 8 chips.

That makes the DRAM subsystem of the S822L much more similar to the E7 memory subsystem with the "Scalable memory interconnect" and "Jordan Creek" memory buffer technology than to the typical Xeon E5 servers.

Benchmark Configuration and Methodology

For our testing we installed 64-bit Ubuntu 15.04 Linux (Kernel version 3.19.0) so that we were able to use GCC 4.9.2, which has better support for the POWER8. We tried to keep the colors inside our benchmark graphs consistent: dark blue is IBM, light blue is the latest Intel Xeon generation (Haswell, E5 v3), and gray was reserved for older Intel systems.

Meanwhile on a quick aside, we should point out that IBM's servers also support PowerVM and KVM virtualization, however we decided not to make use of it to keep the complexity of the tests under control. As we explained in the introduction, porting and tuning the usual benchmarks was quite a challenge, and virtualization makes benchmarking a lot more complex. Testing virtualized workloads was thus beyond the scope of this article.

All tests have been done with the help of Kirth and Wannes of the Sizing Servers Lab.

IBM S822L (2U Chassis)

CPU	Two IBM POWER8 3.425 GHz 10 cores
RAM	128GB (8x16GB) IBM CDIMMs
Internal Disks	2x 300GB 15K RPM SAS Disks (boot) 1x Intel DC P3700 400 GB (Data and benchmarks)
Motherboard	No idea
BIOS version	OPAL v3
PSU	Dual Emerson 1400W

Intel's Xeon E5 Server – "Wildcat Pass" (2U Chassis)

CPU	Two Intel Xeon processor E5-2699 v3 (2.3GHz, 18c, 45MB L3, 145W) Two Intel Xeon processor E5-2695 v3 (2.3 GHz, 14c, 35MB L3, 120W) Two Intel Xeon processor E5-2667 v3 (3.2 GHz, 8c, 20MB L3, 135W) Two Intel Xeon processor E5-2650L v3 (1.8GHz, 12c, 30MB L3, 65W)
RAM	128GB (8x16GB) Samsung M393A2G40DB0 (RDIMM)
Internal Disks	2x Intel MLC SSD710 200GB (boot) 1x Intel DC P3700 400 GB (Data and benchmarks)
Motherboard	Intel S2600WTT
BIOS version	version 1.01
PSU	Delta Electronics 750W DPS-750XB A (80+ Platinum)

All C-states are enabled in both the BIOS.

Other Notes

Both servers are fed by a standard European 230V (16 Amps max.) powerline. The room temperature is monitored and kept at 23°C by our Airwell CRACs.

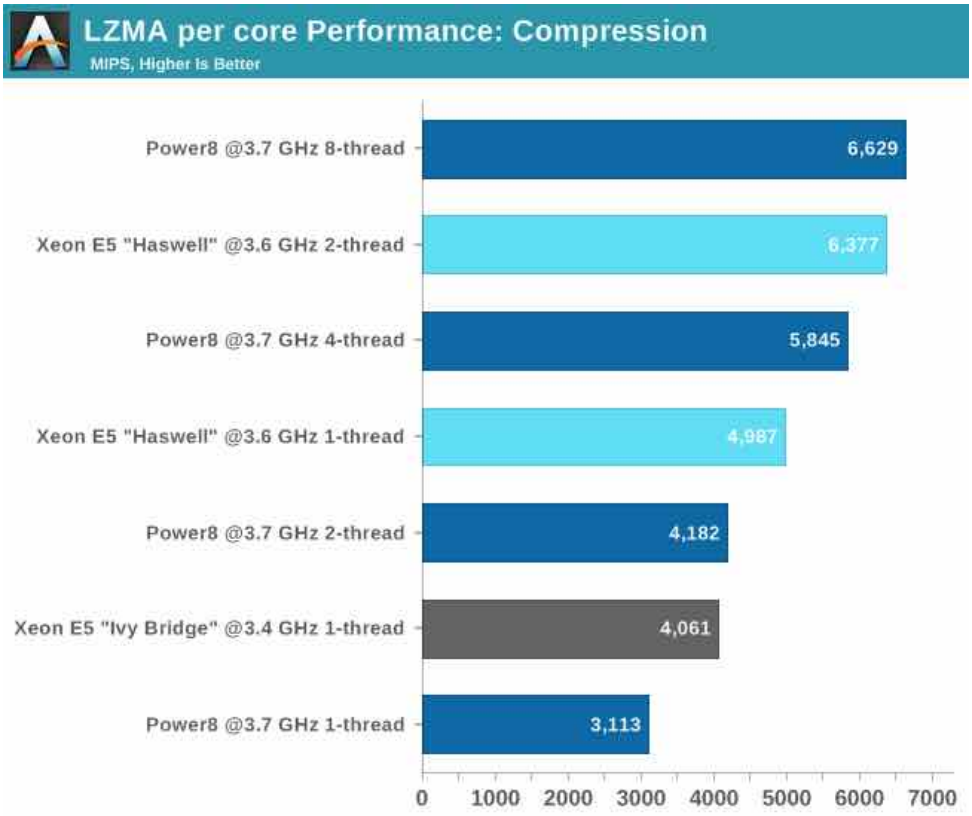
"Per Core" Integer Performance: 7-Zip

The profile of a compression algorithm is somewhat similar to many server workloads: it can be hard to extract instruction level parallelism (ILP) and it's sensitive to memory parallelism and latency. The instruction mix is a bit different, but it's still somewhat similar to many server workloads. Testing single threaded is also a great way to check how well the turbo boost feature works in a CPU.

We ran this benchmark on the POWER8 a few months ago, but there are several reasons to do this again. First of all, we can now use GCC 4.9.2, which has specific support for POWER8 (-mcpu=power8). It is good to note that POWER8 is not a radical new design compared to POWER7. So we only expect modest gains from the compiler.

Secondly, last time we ran on top of PowerKVM, inside a virtual machine. Although that should not make a big difference either - as the benchmark runs almost completely (99%) in user modus and thus runs at 100% - it's still worthwhle to rule out the influence of the virtual machine.

So we recompiled the 7-Zip source code on every machine with the *-O3* optimization with GCC 4.9.2.

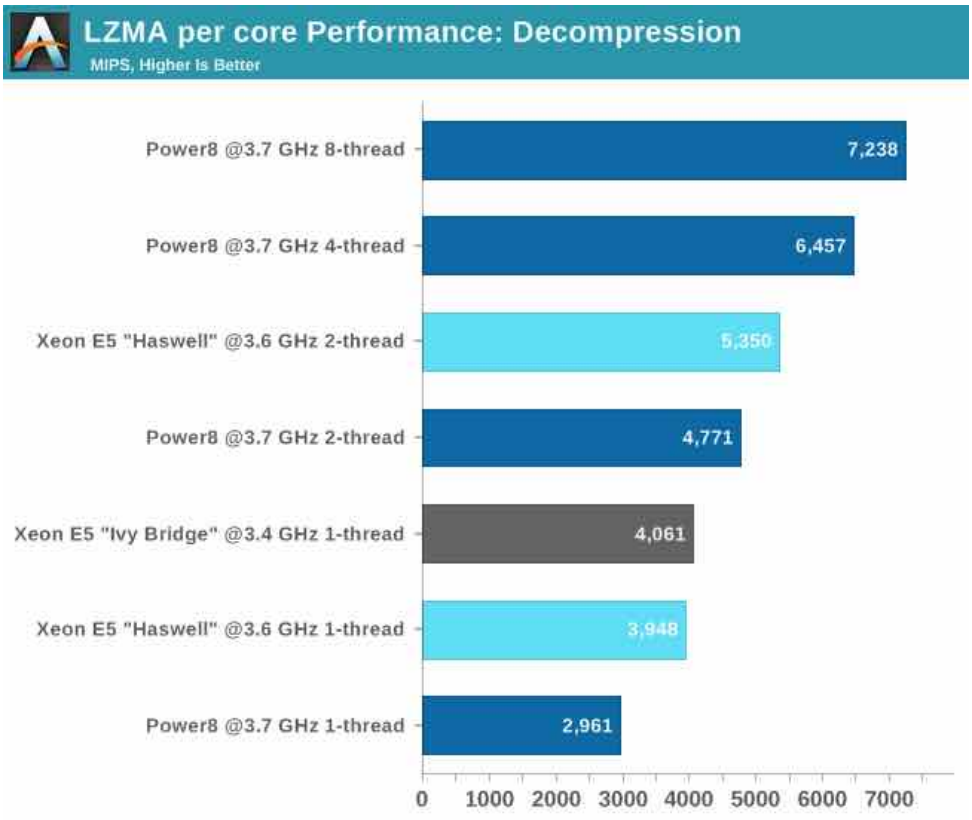


It is important to note that Intel is extremely aggressive with Turbo-boost on the Xeon E5-2699v3. Running code on one core causes the 2.3 GHz Xeon to boost to 3.6 GHz. As a result, the typical clockspeed advantage of the POWER8 was minimized to a measly 90 MHz, with the POWER8 CPUs boosting from 3.425 GHz to 3.690 GHz.

We found that the POWER8 needs more than one thread to deliver good performance: with one thread we only achieve 62% of the performance of a Haswell core at the same speed. Using the mcpu=power8 compiler flag did little more than boost the performance by 1-3%, which is within the margin of error of this benchmark. So your (occassional?) single threaded code will fare badly on POWER8.

Once you fire off 8 threads however, the POWER8 CPU outperforms the hyperthreaded Haswell core slightly (4%).

How about decompression which is even more (IPC) unfriendly to our brainiacs?

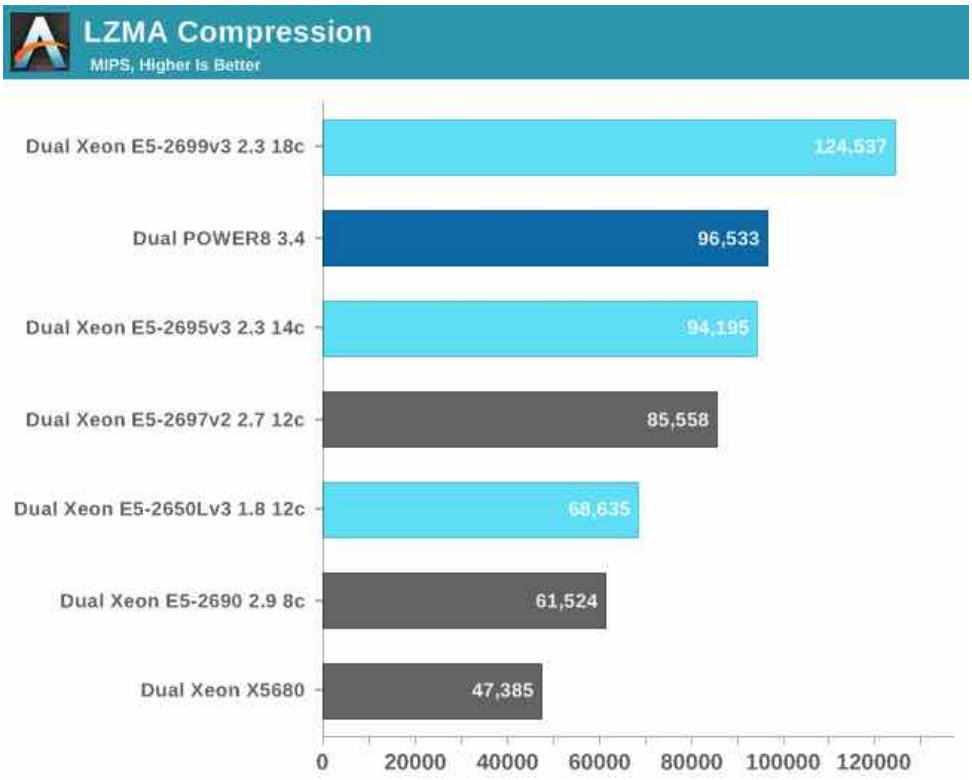


With a single thread, performance of a POWER8 core is about 25% slower than a Haswell core. The Haswell core is still clearly better in extracting Instruction Level Parallelism out of this ILP-unfriendly code. However, let there be no mistake about the integer crunching power of POWER8: it delivers 35% higher performance than the hyperthreaded Xeon E5, core per core, clock per clock (give or take a few MHz).

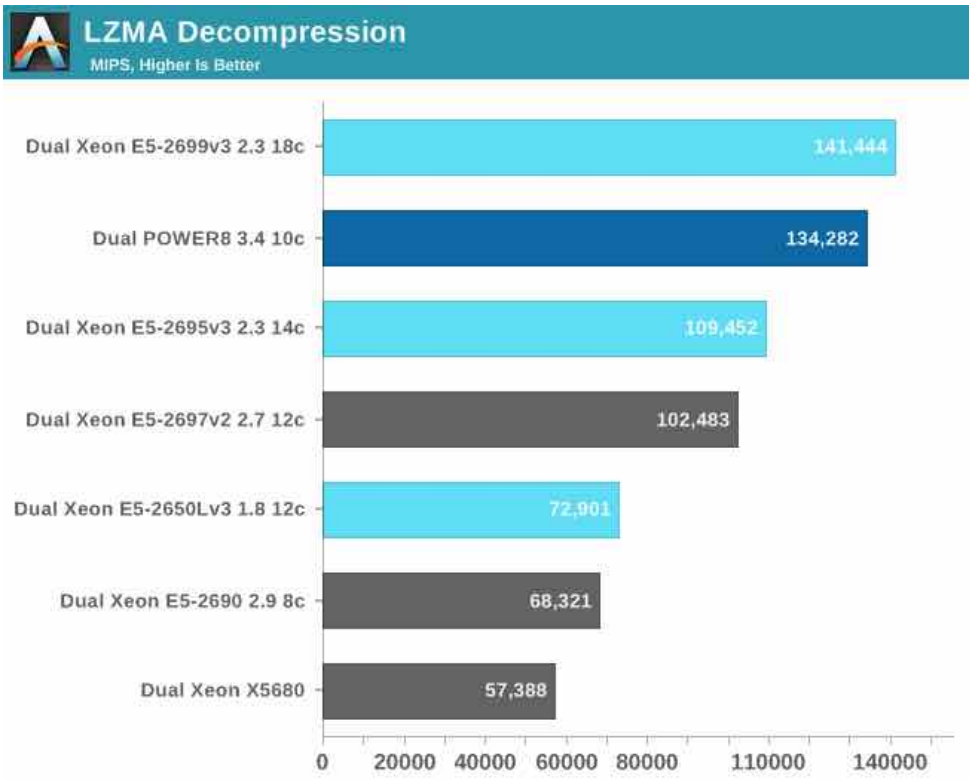
Compression depends more on the datacache and OoO engine. It is remarkable that the Haswell core with its smaller L1-datacache does a lot better than the POWER8. The many unpredictable branches of the decompression code underutilize these very wide modern cores, and as a result the SMT-8 capable POWER8 outperforms the dual-threaded (SMT-2) Haswell. Notice that running two threads instead one thread on the POWER8 offers 61% better performance. Running 8 threads delivers 2.4x higher performance, a clear indication that the POWER8 CPU has a very wide integer execution engine, but can only deliver if enough threads are active.

Multi-Threaded Integer Performance

Let's now take a look at multi-threaded workloads. Using all 18 cores causes the E5-2699 v3 to reduce the clock from 3.6 GHz to 2.8 GHz. Our POWER8 CPU on the other hand only has 10 cores.



The result is of course that the POWER8 cannot keep up with the 18-core Xeon, which runs most of the time at 2.8 GHz. But let's not forget that this is not the top-tier POWER8 CPU, which is the 12-core 3.5 GHz DCM inside the S824. All in all, this is solid performance as the POWER8 edges out a 14-core Xeon running at 2.8 GHz (E5-2695).



With LZMA decompression on the other hand, despite having only 10 cores the POWER8 almost keeps up with the 18-core Xeon running at 2.8 GHz. This is remarkable as it has been a long time since we have this kind of performance from a CPU that cost less than half as much as Intel's most expensive Xeons ($4000+).

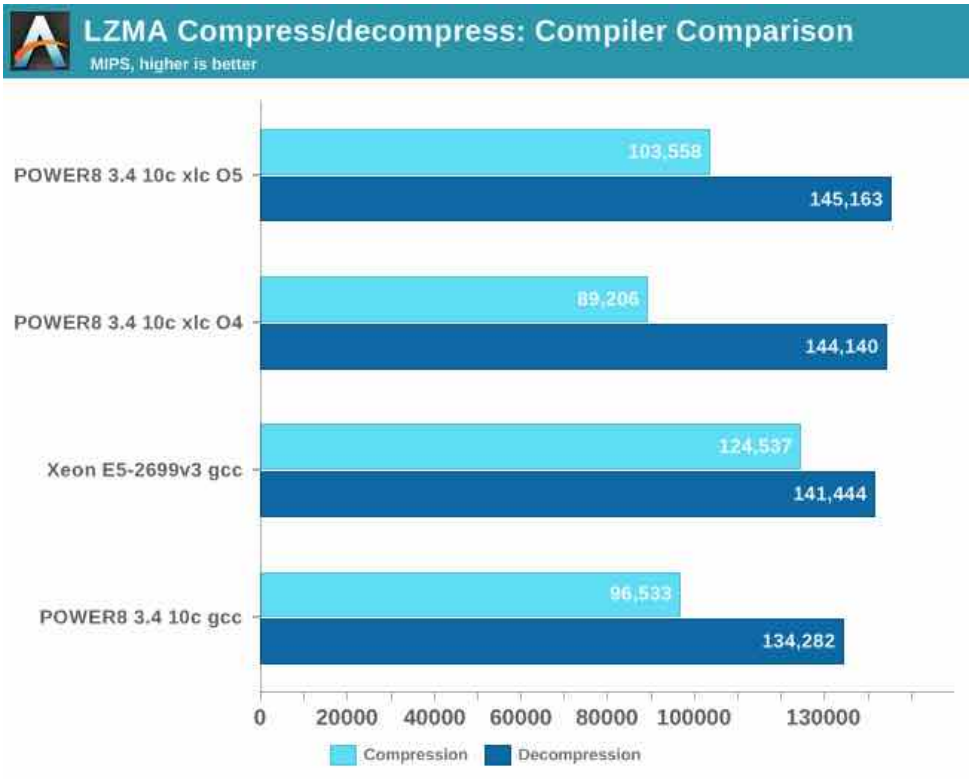
Trying Out a Different Compiler: IBM's XLC

Based on our previous results the IBM POWER8 has definitely attracted our interest, and we were motivated to test a lot more. Can IBM's own compiler "XLC" boost the scores even more?

To test that out, we joined the IBM OpenPOWER Linux community and downloaded the IBM XLC compiler. We compiled with two different and rather aggressive settings:

- *-qhot -O4 -qarch=pwr8*
- *-qhot -O5 -qarch=pwr8*

We've taken our XLC binaries and set them up against binaries compiled with GCC 4.9.2 using the "-o3 -mcpu=power8" flag. The mcpu=power8 flag has very little impact on performance, but we wanted to be sure that GCC was given every opportunity to optimize for the POWER8 CPU.



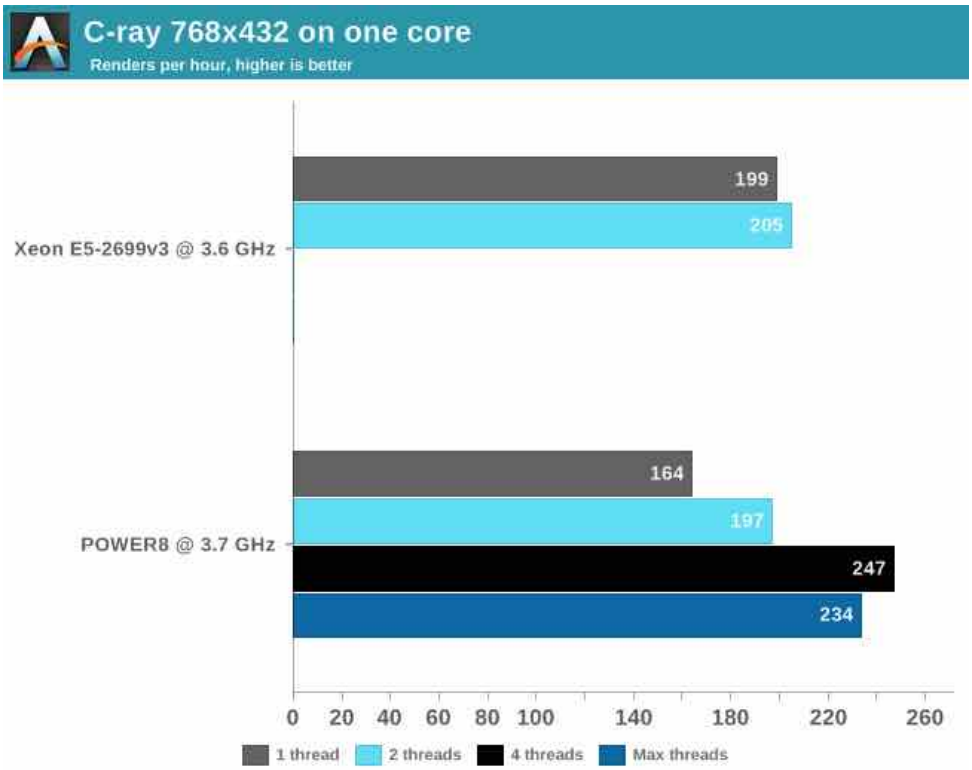
The results for XLC are very weird. Using the -O4 flag the XLC compiler does pretty badly on compression (-7%), while increasing performance by 7% in decompression. Nothing to write home about. Only when we use the -O5 flag do we get an increase in performance by 7-8%. However we also found that -O5 was too aggressive for most complex software that we ported to the POWER8, and as a result isn't very usable.

We suspect that the XLC compiler for LE Linux is still a bit immature and has still some room to improve. Which unfortunately isn't doing IBM any favors at this moment since XLC is a paid compiler.

Floating Point: C-ray

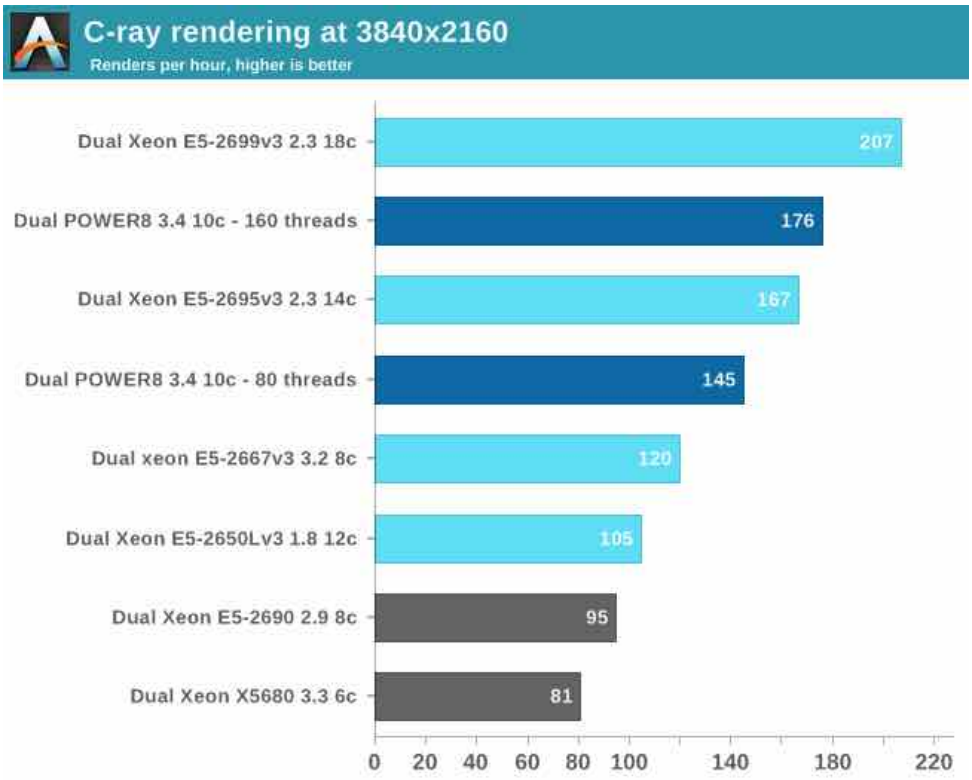
Shifting over from integer to floating point benchmarks we have C-ray. C-ray is an extremely simple ray-tracer which is not representative of any real world raytracing application. In fact, it is essentially a floating point benchmark that runs out the L1-cache. Luckily it is not as synthetic and meaningless as Whetstone, as you can actually use the software to do simple raytracing. That is not the kind of benchmark we like to use for the evaluations of server CPUs, but since our first efforts to port some of our favorite applications to OpenPOWER failed, we settled for something easier. We knew we would have the POWER8 system only for a few weeks, so we had to play it safe.

First we compiled the C-ray multi-threaded version with *-O3 -ffast-math*. To understand the CPU performance better, we limited C-ray with taskset to one or two threads (CPU 0 and 18) on the Haswell-based Xeon and one to eight threads on the POWER8. We also kept the output resolution at 768x432 to keep the render times in check. The "sphfract" file was used as input.



Real floating point intensive applications tend to put the memory subsystem under pressure, and running a second thread makes it only worse. So we are used to seeing that many HPC applications performe worse with multi-threading on. But since C-ray runs mostly out of the L1-cache, we get different behavior. Still, 8 threads of floating action seem to be too much: the POWER8 delivers the best FP performance at 4 threads. At this point, the POWER8 core is able to deliver 20% higher floating point performance than the Haswell Xeon.

Next we used all 160 (20 x 8 threads SMT) or 72 (36 x 2 threads SMT) threads and increased the resolution to 3840x2160.

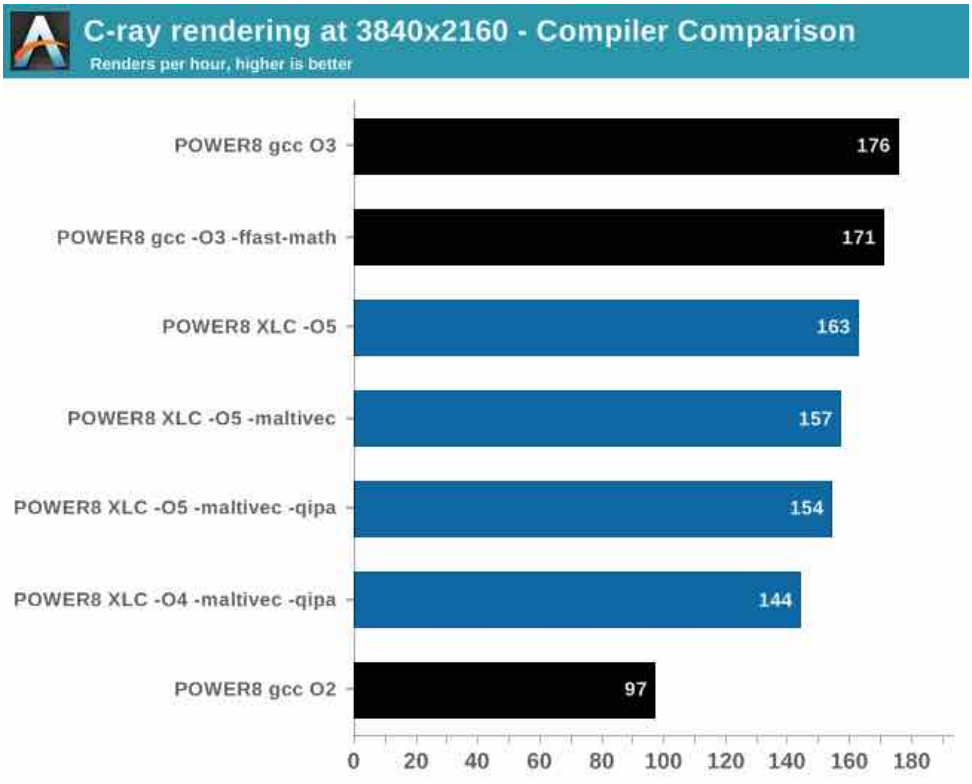


With a core count that is 80% higher, there is nothing stopping the Xeon E5-2699 v3 from taking the top spot. Still, the POWER8 delivers solid performance and outperforms the slower Xeon E5-2695 v3 by 5%. Although

the real world relevance of this benchmark is small, we now have an idea of how good the "basic FP" performance is. Otherwise in real world applications, the use of AVX-2/VSX and the available bandwidth will play a role.

Floating Point & Compilers

When looking at floating point performance we also tried out different compiler settings.



But nothing we tried gave us better results than GCC with the "-O3" setting. It looks like that setting is very effective as it almost doubles performance compared to "-O2". Once again, the XLC compiler fails to deliver any speed boost. In fact, XLC compiled code is slower than GCC 4.9.

Floating Point: NAMD

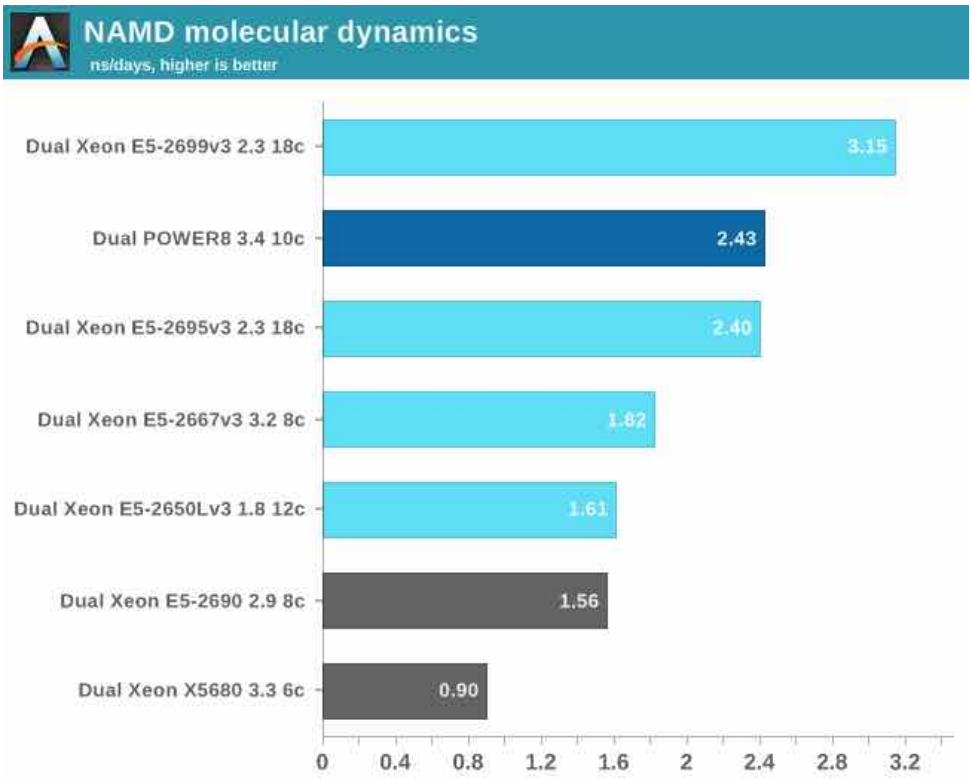
After quite a bit of trouble, we managed to port a real floating point application to our POWER8 system: NAMD.

Developed by the Theoretical and Computational Biophysics Group at the University of Illinois Urbana-Champaign, NAMD is a set of parallel molecular dynamics codes for extreme parallellization on thousands of cores. NAMD is also part of SPEC CPU2006 FP.

We used the "NAMD_2.10_Linux-x86_64-multicore" binary for our Xeons. Since there was no LE Linux version for POWER8, we built our own. We got it working with the g++ compiler and used these settings:

-O3 -mcpu=power8 -ftree-vectorize -mpopcntd

These setting should push GCC to generate as much VSX (Vector Scalar eXtenstion) code as possible. We used the most popular benchmark load, apoa1 (Apolipoprotein A1). The results are expressed in simulated nanoseconds per wall-clock day.



To put this in perspective: an early Xeon Phi (7120 1.2 GHz) scores about 4.4, A top NVIDIA GPU with CUDA based NAMD can score up to 20 and more. So it is clear that this kind of software will be run mostly on GPU accelerated servers.

But it is nonetheless a real world HPC benchmark. The IBM POWER8 is once again on par with the Xeon E5-2695v3. The NAMD binary does not seem to leverage AVX2, as the Xeon E5-2667 (16 cores) does not outperform the Xeon E5-2690 (AVX) with a large margin.

Database Performance: MySQL

Both MySQL and PostgreSQL do not scale well enough to make use of 72 threads (Dual Xeon E5), let alone 160 threads (Dual POWER8). We installed Percona MySQL Server 5.6, which is the most scalable InnoDB based MySQL server.

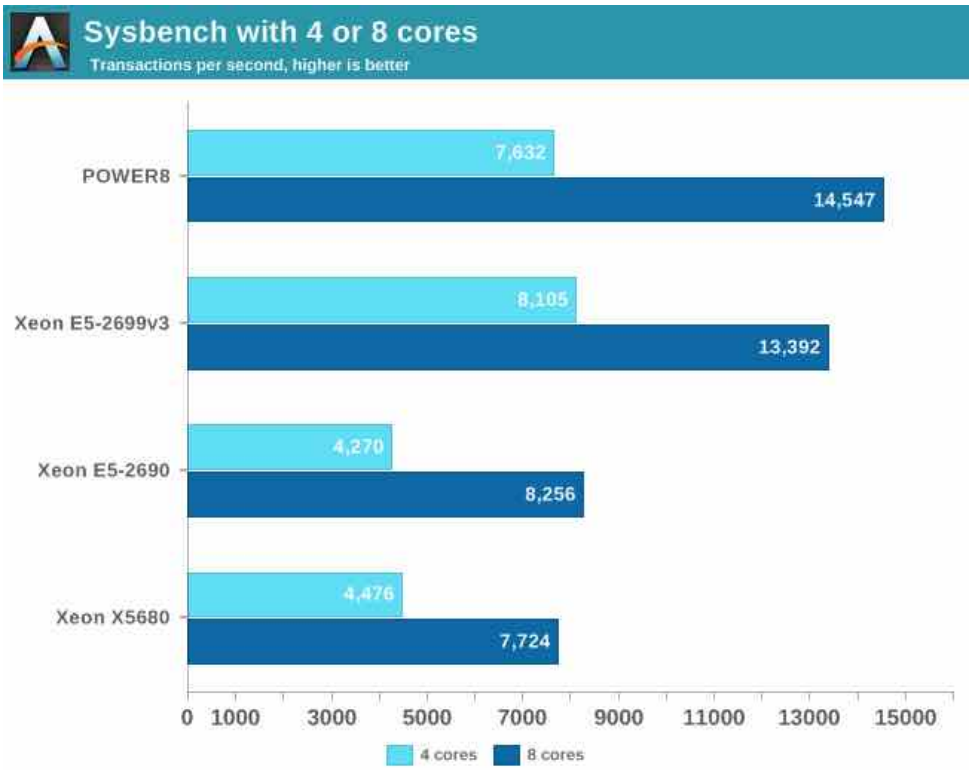
We used the MySQL Sysbench benchmark, but we limited MySQL with taskset to run on 4 or 8 physical cores. We verified that this was actually the case by running "nmon" on top of the IBM server.

You can clearly see that the first 32 threads are used (*CPU 0 - 7*).

Sysbench allows us to place an OLTP load on a MySQL test database, and you can chose the regular test or the read-only test. We chose read only as even with solid state storage, Sysbench is quickly disk I/O limited.

We tested with 10 million records and 100,000 requests. The main reason why we tested with Sysbench is to get a huge amount of queries that only select very small parts (a few or one row) of the tables, Sysbench allows you to test with any number of threads you like, but there is no "think time" feature. That means all queries fire off as quickly as possible, so you cannot simulate "light" and "medium" loads.

The response times are very small, which is typical for an OLTP test. To take them into account, we are showing you the highest throughput at around 3-4 ms. As the results tend to vary a bit, we give you the average of three runs.



With only 4 cores active, the Xeon E5-2699 v3 is still running at 3.3 GHz. Once we use 8 cores, the clockspeed lowers to 2.9 GHz, and the POWER8 outperforms the best Xeon by a small margin. However, we are only testing a part of the CPUs, similar to running only one VM. Ultimately what this means is that total performance will be:

- *the POWER8 will be around 36k (14400/8 * 20 cores)*
- *the Xeon E5-2699 v3 will be around 60k (13400/8 * 36 cores)*
- *the Xeon E5-2695 v3 will be around 45k (13000/8 * 28 cores)*

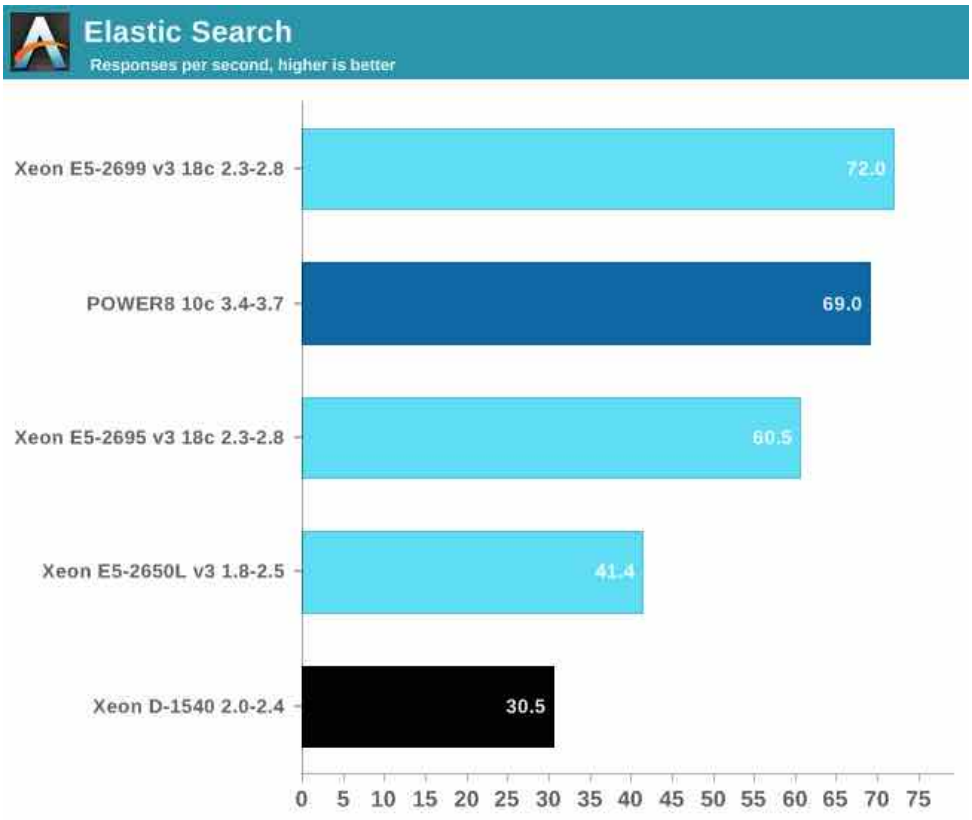
So the current MySQL performance on top of POWER8 is good, but MySQL runs still a lot better on the Xeons.

Scale-Out Big Data Benchmark: ElasticSearch

ElasticSearch is an open source, full text search engine that can be run on a cluster relatively easy. It's basically like an open source version of Google Search that can be deployed in an enterprise. It should be one of the poster-children of scale-out software and is one of the representatives of the so called "Big Data" technologies. Thanks to Kirth Lammens, one of the talented researchers at my lab, we have developed a benchmark that searches through all the Wikipedia content (+/- 40GB). Elasticsearch is – like many Big Data technologies – built on Java.

We are not sure why, but installing IBM's JDK caused a lot of headaches. For some reason the JVM stopped working in the middle of our tests. We got the same behavior running Apache Spark. This could be a result of our lack of experience with the IBM JDK, or the fact that the Linux LE ecosytem is still young. To cut a long story short, we ended up useing OpenJDK 8, which is part of the Ubuntu 15.04 distribution. OpenJDK is very similar to and based upon the same code as Oracle's HotSpot JDK.

We limited the systems to one socket to avoid the issues associated with garbage collection pauses and other scaling issues. There is reason why many Java benchmarks on these massive machines are using multiple JVMs.



Although the POWER8 can probably perform a bit better with the IBM JDK, performance is in the same league as the best Xeons. Meanwhile as a further point of comparison we also included the score of the Xeon D from our previous article.

Energy and Pricing

Unfortunately, accurately and fairly comparing energy consumption at the system level between the S822L and other systems wasn't something we were able to do, as there were quite a few differences in the hardware configuration. For example, the IBM S822L had two SAS controllers and we had no idea how power hungry that chip under the copper heatsink was. Still there is no doubt that the dual CPU system is by far the most important power consumer when the server system is under load. In case of the IBM system, the Centaur chips will take their fair share too, but those chips are not optional. So we can only get a very rough idea how the power consumption compares.

Xeon E5 299 v3/POWER8 Comparison (System)

Feature	2x Xeon E5-2699v3	2x IBM POWER8 3.4 10c IBM S822L
Idle	110-120W	360-380W
Running NAMD (FP)	540-560W	700-740W
Running 7-zip (Integer)	300-350W	780-800W

The Haswell core was engineered for mobile use, and there is no denying that Intel's engineers are masters at saving power at low load.

The mightly POWER8 is cooled by a huge heatsink

IBM's POWER8 has pretty advanced power management, as besides p-states, power gating cores and the associated L3-cache should be possible. However, it seems that these features were not enabled out-of-the box for some reason as idle power was quite high. To be fair, we spent much more time on getting our software ported and tuned than on finding the optimal power settings. In the limited time we had with the machine, producing some decent benchmarking numbers was our top priority.

Also, the Centaur chips consume about 16W per chip (Typical, 20W TDP) and as we had 8 of them inside our S822L, those chips could easily be responsible for consuming around 100W.

Interestingly, the IBM POWER8 consumes more energy processing integers than floating point numbers. Which is the exact opposite of the Xeon, which consumes vastly more when crunching AVX/FP code.

Pricing

Though the cost of buying a system might be only "a drop in the bucket" in the total TCO picture in traditional IT departements running expensive ERP applications, it is an important factor for almost everybody else who buys Xeon systems. It is important to note that the list prices of IBM on their website are too high. It is a bad habit of a typical tier-one OEM.

Thankfully we managed to get some "real street prices", which are between 30% (one server) and 50% (many) lower. To that end we compared the price of the S822L with a discounted DELL R730 system. The list below is not complete, as we only show the cost of the most important components. The idea is to focus on

the total system price and show which components contribute the most to the total system cost.

Xeon E7v3/POWER8 Price Comparison

Feature	Dell R730		IBM S822L	
	Type	**Price**	**Type**	**Price**
Chassis	R730	N/A	S822L	N/A
Processor	2x E5-2697	$5000	2x POWER8 3.42	$3000
RAM	8x 16GB DDR4 DIMM	$2150	8x 16 GB CDIMM (DDR3)	$8000
PSU	2x 1100W	$500	2x 1400W	$1000
Disks	SATA or SSD	Starting at $200	SAS HD/SSD	+/- $450
Total system price (approx.)		$10k		$15k

With more or less comparable specs, the S822L was about 50% more expensive. However, it was almost impossible to make an apples-to-apples comparison. The biggest "price issue" are the CDIMMs, which are almost 4 times as expensive as "normal" RDIMMs. CDIMMs offer more as they include an L4-cache and some extra features (such as a redundant memory chip for each 9 chips). For most typical current Xeon E5 customers, the cost issue will be important. For a few, the extra redundancy and higher bandwidth will be interesting. Less important, but still significant is the fact that IBM uses SAS disks, which increase the cost of the storage system, especially if you want lots of them.

This cost issue will be much less important on most third party POWER8 systems. Tyan's "Habanero" system for example integrates the Centaur chips on the motherboard, making the motherboard more expensive but you can use standard registered DDR3L RDIMMs, which are much cheaper. Meanwhile the POWER8 processor tends to be very reasonably priced, at around $1500. That is what Dell would charge for an Intel Xeon E5-2670 (12 cores at 2.3-2.6 GHz, 120W). So while Intel's Xeon are much more power efficient than the POWER8 chips, the latter tends to be quite a bit cheaper.

Comparing Benchmarks: AT vs IBM

Before we close things out, let's spend a moment summarizing our results and comparing the performance we saw to the kind of performance advantages that IBM advertises POWER8 is capable of.

From a high level perspective, the S822L is more expensive and consumes a lot more power than a comparable Xeon system.

With limited optimization and with the current Ubuntu 15.04, the performance-per-watt ratio favors Intel even more as the POWER8 barely outperforms the very efficient 120W TDP Xeons. So there is no denying that the Intel systems offer a superior performance/watt ratio.

However, it would be unfair to base our judgement on our first attempt as we have to admit this our first real attempt to benchmark and test the POWER8 platform. It is very likely that we will manage to extract quite a bit more performance out of the system on our second attempt. IBM POWER8 also has a big advantage in memory bandwidth. But we did not manage to port OpenFOAM to the POWER platform, probably the most likely candidate for leveraging that advantage.

We are less convinced that the POWER8 platform has a huge "raw CPU compute advantage," contrary to what for example IBM's SPECJBB (85% faster) and SAP (29% faster) results seem to suggest.

For example, IBM's own SPECjEnterprise®2010 benchmarking shows that:

- *Two Xeon E5-2697 v3, 14 cores at 2.6-3.1 GHz score 19282*
- *Two POWER8, 12 cores at 3.5-3.8 GHz score 22543*

SAP is "low IPC" (hard to run many instructions in parallel in one thread) software that benefits much from low latency caches. The massive L3-cache (12-cores, 96 MB) and huge thread count are probably giving the IBM POWER8 the edge. The RAM bandwidth also helps, but in a lesser degree. IBM clearly built POWER8 with this kind of software in mind. We had individual threadcount intensive benchmarks (LZMA decompression) and L3-cache sensitive benchmarks (ElasticSearch), but t o be fair to IBM, none of our benchmarks leveraged the three strongest points (threadcount, L3-cache size and memory bandwidth) all at once like SAP.

SPECJBB2013 has recently been discontinued as it was not reliable enough. We tend to trust the jEnterprise test a lot more. In any case, the best POWER8 has a 17% advantage there.

Considering that the POWER8 inside that S824 has 20% more cores and a 3% higher clockspeed, our 3.4 GHz 10-core CPU would probably be slightly behind the Xeon E5-2697 v3. We found out that the 10-core POWER8 is slightly faster than Xeon E5-2695 v3. The Xeon E5-2695 v3 is very similar to the E5-2697 v3, it is just running at a 10% lower clockspeed (All core turbo: 2.8GHz vs 3.1GHz). So all in all, our benchmarks seems to be close to the official benchmarks, albeit slightly lower.

Closing Thoughts: A Mix of Xeon "E5" and "E7"

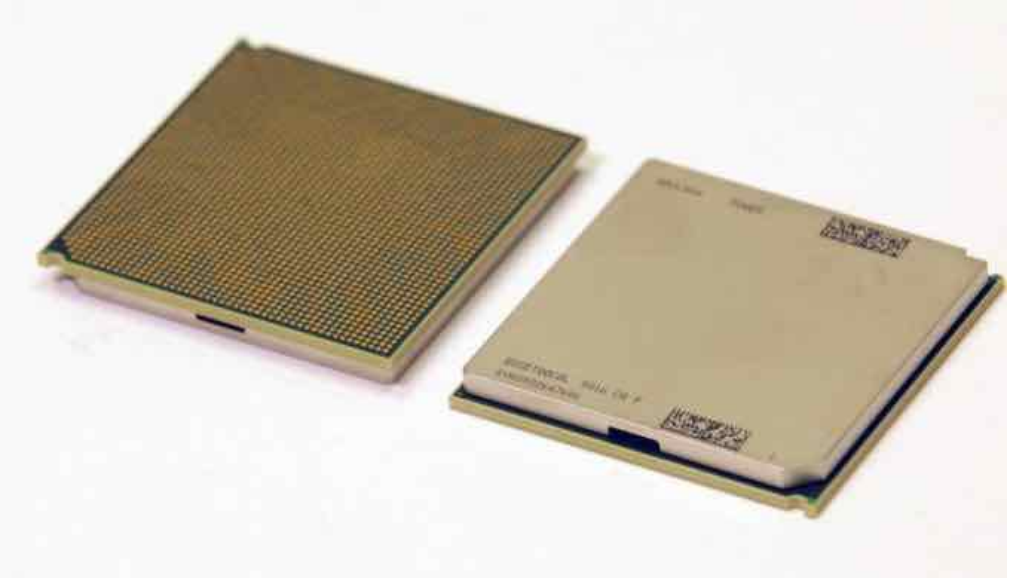
So let's sum things up. The IBM S822L is definitely not a good choice for those looking to lower their energy bills or operate in a location with limited cooling. The pricing of the CDIMMs causes it to be more expensive than a comparable Xeon E5 based server. However, you get something in return: the CDIMMs should offer higher reliability and are more similar to the memory subsystem of the E7 than the E5. Also, PCIe adapters are hot-pluggable on the S822L and can be replaced without bringing down the system. With most Xeon E5 systems, only disks, fans and PSU are hot-pluggable.

In a number of ways then, the S822L is more a competitor to dual Xeon E7 systems than it is to dual Xeon E5 systems. In fact, a dual Xeon E7 server consumes in the 600-700W range, and in that scenario the power usage of S822L (700-800W) does not seem outrageous anymore.

The extra reliability is definitely a bonus when running real time data analytics or virtualization. A failing memory chip may cost a lot when you running fifty virtual machines on top of a server. Even in some HPC or batch data analytics applications where you have to wait for hours for a certain result that is being computed in an enormous amount of memory, the cost savings of being able to survive a failing memory chip might be considerable.

One more thing: for those who need full control, the fact that every layer in the software stack is open makes the S822L very attractive. For now, the available "OpenCompute" Xeon servers that are also "open" seem to

mostly density optimized servers and the openess seems limited on several levels. Rackspace felt that the current OpenCompute servers are not "open enough", and went for OpenPOWER servers instead. In all those markets, the S822L is very interesting alternative to the dual Xeon E5 servers.



Ultimately however, the performance-per-dollar Xeon E5 competitors will most likely be OpenPOWER third party servers. Those servers do not use CDIMMS, but regular RDIMMs. Other components such as disks, networkcards and PSUs will probably be cheaper but potentially also slightly less reliable.

All in all, the arrival of OpenPOWER servers is much more exciting than most of us anticipated. Although the IBM POWER8 servers can not beat the performance/watt ratio of the Xeon, we now have a server processor that is not only cheaper than Intel's best Xeons, but that can also keep up with them. Combine that with the fact that IBM has lined up POWER8+ for next year and a whole range of server vendors is building their own POWER8 based servers, and we have a lot to look forward to!