CLABBY ANALYTICS

Research Report

POWER8 and POWER9: Accelerated Computing Gone Wild

Executive Summary
With improvements to its POWER8 microprocessor architecture, and with its forthcoming POWER9, *we believe that IBM is poised to pull way ahead of its competitors when it comes to accelerated system designs*. We see distinct advantages in performance and cost/performance; in leadership accelerator-attach technologies; and, in a rich ecosystem that has developed around POWER architecture.

In this *Research Report*, *Clabby Analytics* describes what "accelerated systems" are; how they work – and then we take a closer look at some of the present and future plans for IBM POWER-based system designs.

Introduction
Over the past several years, all major server makers have focused on redesigning their server architectures in order to eliminate data flow bottlenecks – building what have become known as "accelerated systems". Likewise, the major developers of database software have also found new ways to accelerate database processing.

> ***The combination of accelerated system designs coupled with database acceleration has created a data-processing panacea for fast, efficient and affordable processing of very large databases (Big Data).***

As far back as early 2013, *Clabby Analytics* started to report on the progress being made in accelerated system designs (our reports on VelociData, The Now Factory, IBM's PureData system, IBM's DB2 Analytics Accelerator and IBM Power Systems accelerated designs are available upon request). What intrigued us about the early designs was that several systems were making use of field programmable gate arrays (FPGAs) to accelerate data transfer (processing data a line speed); and some designs were also making use of graphical processing units (GPUs) to accelerate parallel processing. We also found that the makers of general purpose CPUs (especially POWER and Intel processors) were more focused than ever before on processing large amounts of data (Big Data).

In 2014 we started tracking database accelerators (our report on IBM's DB2 BLU Acceleration vs. SAP's Hana vs. Oracle's Exadata is available upon request). What we observed was that new algorithms combined with new data processing methodologies were being used to accelerate database processing. Also, in 2014, we noted that sales of in-memory database software from IBM, SAP, Oracle, Altibase, Exasol, Kognito, McObject, ParStream, Quartet FS, VMware and VoltDB were on the rise – helping further accelerate the processing of large amounts of data (see this report for further details).

Today, in 2016, we're starting to see one systems vendor pull ahead of its competitors when it comes to architecting accelerated systems. In a recent briefing, IBM shared with us its plans for further accelerating its current POWER8-based servers – and IBM also shared its future plans for its next generation POWER9 architecture. In this *Research Report* we share with our readers some of our observations based upon this briefing.

> ***The mantra for POWER8 when it was introduced was "designed for data" – meaning that several of the features were designed to accelerate the processing of large databases. With POWER9, we believe the new mantra should be "designed for accelerated systems leadership" because the new POWER9 processor and system designs should significantly outperform competing Xeon E7, Sparc M and Sparc64 designs. For enterprises planning their future Big Data strategies, it is time to become very familiar with IBM's fast approaching POWER9 architecture.***

What Are Accelerated Systems?

From a performance perspective, the primary goal for any information system should be to achieve *balanced performance between processors, memory, and I/O (input/output)*. Accordingly, information technology (IT) executives should strive to avoid scenarios where a workload overpowers the processor, memory or the I/O subsystem. To avoid scenarios like this, IT executives need to look at CPU performance characteristics, off-load characteristics, memory utilization, processor core efficiency, execution styles, instruction set characteristics, communications and network facilities, and a few other design elements. A workload optimized environment should keep processors, memory and I/O very busy ─ but should not overwhelm them.

The total amount of work a system can execute is known as the system's "capacity". To maximize ROI, IT buyers need to get as much work as possible out of the system (so buyers should strive to reach maximum capacity). The way to measure the amount of work being done is known as "throughput" and is measured in data packets per second.

> ***Accelerated system designs focus on speeding workload processing by improving system throughput. Improved throughput is achieved in a variety of ways including increasing internal bus speed, reducing communications overhead, off-loading tasks to other types of processors, improving memory management, reducing I/O drag, tuning execution methods, creating new interfaces that streamline peripheral access to processing power, and more.***

Why Accelerated Systems?

Traditional server designs can be used to process Big Data, technical and cognitive workloads. But the big question is: "how efficiently?" Accelerated systems can process workloads exponentially faster than traditional designs – and this speed advantage results in the ability to process more and more data at a significantly lesser cost.

> ***Annecdote: When the computer system salesman told an IT executive that his new system could process data exponentially faster than the executive's current system – the IT executive said "I don't care about***

> ***'exponentially faster' claims". The salesman, not missing a beat, then said "okay – let me rephrase. How would you like to use significantly fewer systems and pay for significantly fewer software licenses in order to process your workloads?" The IT executive was suddenly more interested.***

IBM POWER8 Accelerated Systems

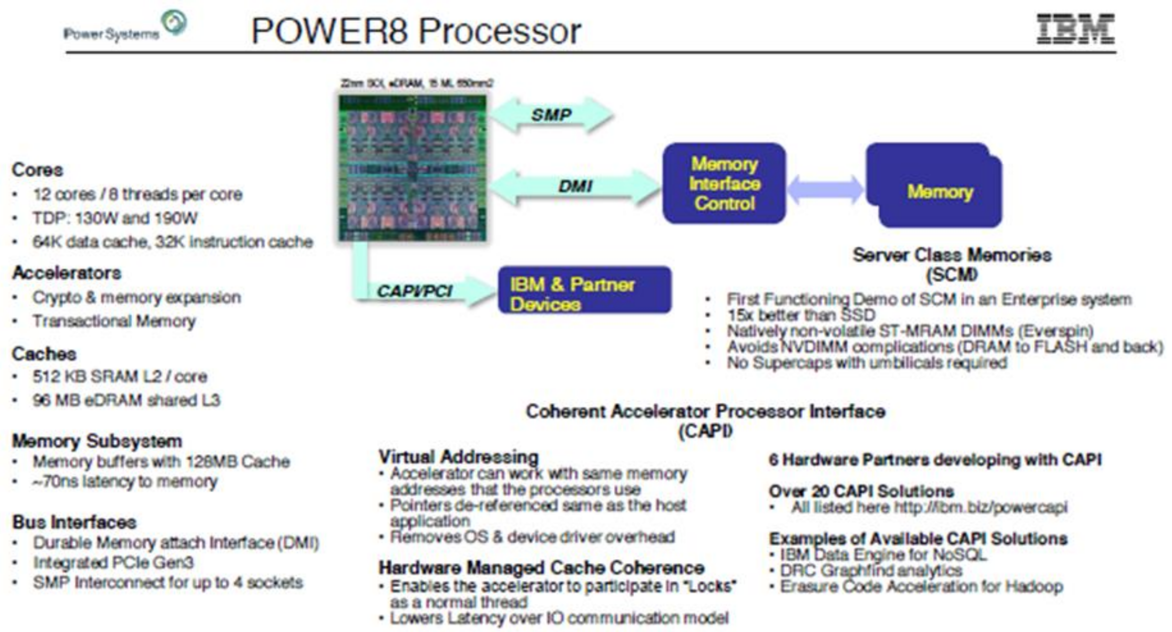
In June, 2014, *Clabby Analytics* took a close look at IBM's POWER8 architecture when we compared it to Intel's E5 v2 architecture (see this report). In that report we found that both microprocessor/server environments had been designed to process Web, file and print, email, database, vertical-specific applications, high performance computing and cloud workloads. We also found that POWER8 processors were more efficient than its Xeon competitor (due to processing and bandwidth advantages, POWER8-based servers were able to deliver results more quickly). When we looked at Power Systems designs, we found that POWER8-based servers were better suited for data-intensive environments from a performance and price/performance perspective (POWER8-based servers cost less than E5 v2-based competitors due to aggressive IBM pricing and numerous efficiency advantages).

Today, as we look more closely at Power8-based system designs, we find that that IBM has further accelerated its Power Systems by:

- Using fast – and sometimes different – processors to handle specific workloads;
- Expanding memory access using crypto and memory expansion – and by introducing transactional memory;
- Increasing on-chip cache sizes and increasing the size of memory buffers to make it possible to place data closer to central processing units (CPUs) and graphical processing units (GPUs) where that data can be processed more quickly;
- Increasing internal bus speed to accelerate data flow within a server;
- Streamlining input/output (I/O) access to processors using interfaces such as CAPI and RDMA (and will soon use NVLink) to reduce I/O overhead and simplify I/O device to processor data movement;
- Streamlining networking subsystems to speed communications;
- Using virtual addressing to allow accelerators to use the same memory that processors use in order to remove operating system and device overhead; and,
- Introducing hardware managed cache coherence.

Figure 1 (next page) illustrates the technologies that IBM uses to accelerate its POWER8-based servers.

<u>Figure 1 – POWER8-based Acceleration Technologies</u>



Source: IBM Corporation – June, 2016

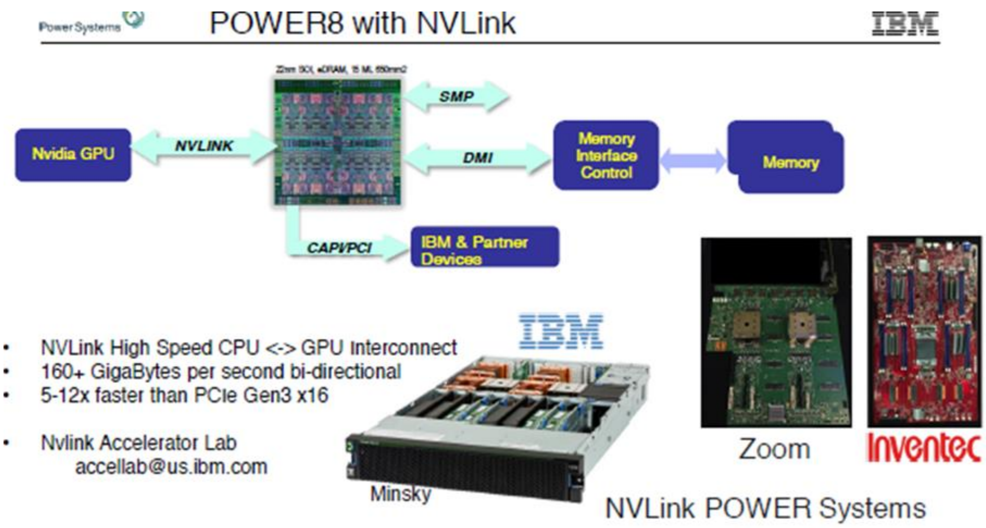
A Major Differentiator: Accelerator-Attach Technologies

Especially worthy of note in Figure 1 is the work that IBM has been doing to accelerate I/O access to its POWER8 processors. When POWER8 was introduced, IBM had placed PCIe Gen 3 logic directly on the chip. IBM then built an interface to this logic known as the coherence attached processor interface (or CAPI). CAPI is a customizable *hardware accelerator* that enables devices, Flash and coprocessors to talk directly and at very high speeds with POWER8 processors. Examples of CAPI-based solutions include IBM's Data Engine for NoSQL (which allows 40GB of flash storage to be used like extended memory); DRC Graphfind analytics; and Erasure Code Acceleration for Hadoop.

> ***The reason this CAPI interface is so important is because it <u>eliminates the need for a PCIe bridge as well as the need to launch the thousands of operating system and driver instructions</u> (perhaps as many as 22.5K instructions) that are run every time PCIe I/O resources need to be used. Instead, the logic for driving I/O resides on the chip where the number of instructions are reduced and the speed of interactions between the CPU and associated hardware devices is dramatically improved.***

Later this year IBM will introduce a new interface known as NVLink to its POWER8 processor. NVLink, created by NVIDIA, will make it possible to tightly link NVIDEA graphical processing units (GPUs) with POWER8, enabling POWER8 and NVIDEA GP100 GPUs to jointly process data. This new interface streamlines the GPU to CPU (central processing unit) interface and is 5X to 12X faster than PCI Gen 3 connections – leading to even more rapid processing of data-intensive workloads. Figure 2 (next page) shows how NVIDIA GPUs can be attached to POWER8 processors – and also describes the interconnect speed (160+ GBps bidirectional, yielding 5-12x faster performance that PCIe Gen3 x16 interconnects).

Figure 2 – POWER8 NVLink Acceleration



Source: IBM Corporation – June, 2016

Another method that IBM Power Systems use to accelerate data flow is known as remote direct memory access (RDMA). RDMA is a facility that enables data to be moved from the memory/storage of one system to the memory or storage of another system – at line speed. The types of workloads that benefit from the use of RDMA are network-intensive applications that suffer from bandwidth/latency-related data retrieval issues. These include:

- Large scale simulations, rendering, large scale software compilation, streaming analytics and trading decisions – the kinds of applications found most often in massively parallel, high performance computing (HPC);
- Hyper-appliance, hyperconverged and hyperscale environment where large volumes of data needs to be moved between servers and associated storage; and,
- Workloads where network latency slows database performance and interferes with virtual machine (VM) density.

IBM's accelerator-attach technologies such as CAPI and NVLink feed data to the POWER8 CPU exponentially faster than previous generation interconnects such as PCIe. To us, IBM appears to be more aggressive in accelerator-attach interfaces than its competitors, and – accordingly – we see this as a distinct competitive differentiator.

Also Noteworthy: Memory Activities

When POWER8-based systems were introduced, IBM announced that memory channel speed has been improved – again enabling more data to be delivered more quickly to processors, thus accelerating the processing of large volumes of data. IBM also announced its Durable Memory Interface (DMI). This interface is agnostic, meaning it enables different types of memory to be attached to the bus (so Power Systems are no longer only tied to DRAM memory).

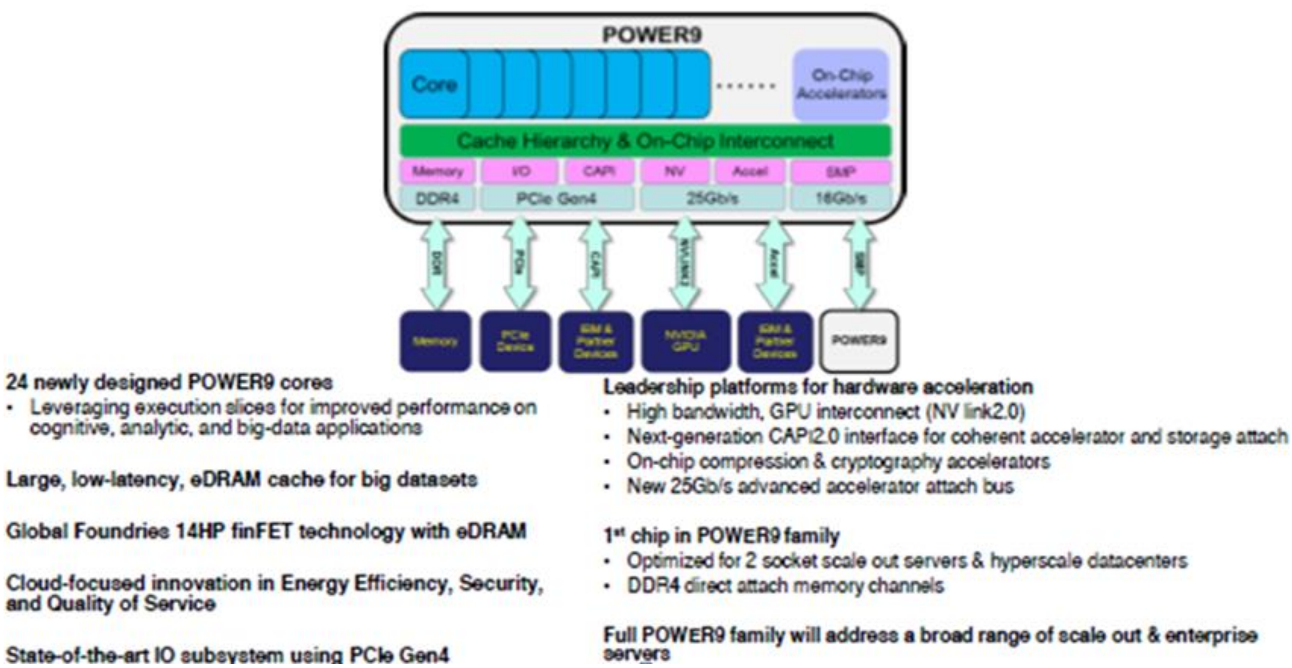
What POWER9 Means to the Future of Accelerated Systems
Just over a year from now (in the second half of 2017), IBM will then introduce its next generation POWER processor: the POWER9. POWER9 will consist of a family of chips that will be focused on 1) analytics and cognitive; 2) new opportunities in cloud and hyperscale; and, 3) technical computing.

Some of the improvements that can be expected when POWER9 is delivered include:

- 24 cores;
- Improvements to IBM's Vector Scaling eXtension (VSX) – which improves floating point extensions (especially useful for cryptographic operations);
- Improvements in branch prediction and a shorter pipeline;
- The use of "execution slices" to improve performance;
- The use of large, low-latency eDRAM cache to accommodate big datasets;
- State-of-the-Art I/O speed (by leveraging PCI Gen4) – giving POWER9 3X faster bandwidth to access I/O and storage as compared with POWER8;
- A new 25Gbps advanced accelerator bus;
- On-chip compression and cryptographic accelerators;
- Access to next generation NVLink 2.0 to increase speed by 33%, plus coherence across GPUs and CPUs memory to enhance usability of GPUs; and,
- Optimization for 2 socket servers using direct attached DDR4 direct attach memory channels.

The new POWER9 features are illustrated in Figure 3.

<u>Figure 3 – POWER9 Feature Set</u>



Source: IBM Corporation – June, 2016

POWER9-based servers will be delivered in multiple different scale-up and scale-out form factors. And IBM, as could be expected, will continue to concentrate on making its next generation POWER processor energy efficient, rich in security features and rich in Quality-of-Service functionality (such as high availability and resiliency).

The Ecosystem Advantage

In 2013, IBM announced the formation of the OpenPOWER Foundation, an organization dedicated to expanding the ecosystem that surrounded IBM's POWER8 (and successive generations) processors. More specifically, IBM announced that it would make processor specifications, firmware and related software available to the evolving OpenPOWER community – essentially making the POWER processor an open standard microprocessor.

From our perspective (as we state in this report), the reasons why IBM opened it POWER architecture were three-fold: 1) the company had seen a steep decline in sales of its Power Systems (Power System subsequently returned to profitability as we show in this report); 2) the company recognized that a broader ecosystem would foster even greater innovation – resulting in new POWER-based systems designs that would also help drive sales; and, 3) IBM recognized and responded to changing market dynamics relative to Moore's Law (silicon alone isn't going to provide the necessary cost/performance improvements in server designs any longer – hence, new aggressive systems designs that exploit other components and that use accelerators were needed).

The member-run OpenPOWER Foundation has been a big success for IBM in terms of promoting POWER8 innovation. In this blog we describe what we learned at the annual OpenPOWER Foundation Summit held in San Jose, California in April, 2016. At this summit the OpenPOWER Foundation was able to demonstrate great progress in building an open, collaborative ecosystem around the Power8 architecture. 59 new products have become part of the POWER8 ecosystem including FPGA cards, chip liquid coolers, memory cards, high-speed switches, new POWER-based system designs, new boards and expansion kits, GPU accelerators, and even a new POWER chip for the China market. Further, OpenPOWER Foundation membership has grown to more than 200 members.

The biggest eye-opener at this Summit was a project being jointly developed by Google and Rackspace that will create a new open server specification based on the forthcoming POWER9 architecture. Both companies plan to submit a candidate server design to the Open Compute Project (a Facebook-inspired initiative that allows vendors and users to share advanced data center and server designs with the information technology industry).

> ***There are huge processing efficiencies to be gained by using POWER processors instead of Intel processors for many types of workloads. POWER processors can process twice as many threads as their Intel competitors, they offer more cache, and new interfaces such as CAPI can deliver lower cost storage/memory connections with far less latency. I have long been surprised that POWER processors don't dominate more industry clouds – and attribute the lack of adoption of POWER-dominated clouds to tardy development and promotion of Linux on POWER. If companies like Rackspace and Google are starting to open the door for POWER-based servers – and if they can prove processing and energy efficiencies – it may not take long for other large cloud vendors to get the message.***

Summary Observations

In days gone by, systems makers focused on increasing processor speed in order to improve processing performance. For almost 50 years, the number of transistors on a processor doubled about every two years (known as "Moore's Law) – so improving system performance constantly centered on improving processing power in order to increase systems performance. But in the late 2000s processors reached their peak in terms of the number of transistors that could be placed on a chip, forcing systems designers to focus more heavily on tuning memory and input/output (I/O) device access, on improving internal bus speed, on improving communications and on streamlining memory access in order make systems perform more quickly. These highly tuned systems have become known as "accelerated systems".

IBM has been particularly aggressive in driving breakthroughs that attack bottlenecks that limit bandwidth. It has created hardware accelerators; it has embedded accelerators at the chip level; it has implemented its own hardware interface (CAPI); and it is using NVIDIA's NVLink to help connect GPUs directly to its POWER8 and POWER9 processors. From a memory perspective it uses RDMA to speed memory access; it has increased memory channel bandwidth – and now it will soon offer its customers the opportunity to use multiple different types of memory in its Power Systems. POWER8 offers large memory caches – as will POWER9.

It should be clear to readers that Power Systems are all about moving data efficiently. IBM has focused strongly on overcoming latency issues and other bottlenecks – delivering some of the most powerful servers in the industry for data processing. Meanwhile, an entire ecosystem has grown-up around POWER8 – enabling major vendors from around the world to contribute new, innovative solutions to the POWER ecosystem.

> ***As we look at IBM's competitors, we see all leading systems makers working on accelerated system designs. However, we see several distinct differentiators in IBM's Power Systems lines including several accelerator-attach technologies, high performance and strong cost/performance. We also see contributions and innovations introduced through the OpenPOWER Foundation as a major differentiator. From our perspective, IBM's aggressive positions in server acceleration, in database acceleration – combined with its broad portfolio of analytics software – make IBM a formidable competitor in the world of accelerated, Big Data servers.***

Clabby Analytics
http://www.clabbyanalytics.com
Telephone: 001 (207) 239-1211