Despite the Pandemic, IBM POWER10 is on the Near Horizon

By Jean S. Bozman

Even in the teeth of the pandemic, IBM is expected to deliver its POWER10 processors in calendar 2021, continuing its drive to place a new generation of POWER microprocessor engines in the world's cloud data centers. Cloud migrations, accelerated by the pandemic's impact on business worldwide, are making it possible to deploy many types of CPUs and GPUs in the cloud infrastructure, which is dominated by servers based on Intel x86 processors.

POWER10 processors are aimed at running in the hybrid cloud, supporting AI/ML analytics applied to enterprise applications and scalable enterprise workloads (e.g., SAP, SAS). That strongly aligns the POWER10 processors with IBM's strategic marketing priorities for hybrid cloud and AI/ML.

The POWER10's life in the extended hybrid cloud, coexisting with other types of processors from Intel and AMD, will be possible because cloud infrastructure supports mixed hardware types in a scale-out hyperscale world where clusters are the cloud's go-to resources for running specific workloads.

POWER10 is positioned for use in highly scalable (scale-up) systems – and in optimized scale-out cloud-native workloads. So, it is intended for supporting workloads that are migrating from the enterprise data center into the hybrid cloud. The emphasis is on faster processing of enterprise applications; support for accelerated AI analytics; support for multi-PB (petabyte) memory clusters, embedded math acceleration and extremely high security.

How Will POWER10 Fit Into Hybrid Clouds?

POWER10's presence in the cloud is significant, because IBM's stated corporate strategy focuses on hybrid clouds and AI. POWER processors are already being deployed in the IBM Cloud and in Google Cloud Platform. POWER-based servers are already in public clouds and private on-prem clouds, in addition to longtime deployments in enterprise customers' data centers.

However, one could argue that in 2021, the customer mix is changing: POWER engines are being deployed into the cloud that doesn't require a traditional IT sales cycle that is aimed at installing POWER-based hardware into an enterprise data center. Rather, POWER will increasingly be viewed as a resource for customers

46 Kent Hills Lane, Wilton, CT. 06897; (203) 429 8951;
http://www.rfgonline.com/; Contact: inquiry@rfgonline.com

running applications that can leverage POWER best, whether that's in data centers or in cloud-service-provider (CSP) public cloud infrastructure.

In the pre-cloud era, POWER's position in the marketplace would have been viewed solely according to unit shipment counts of POWER servers. In the hybrid cloud, it will be based on the capability, expanded memory and the speedup associated with POWER10's on-chip acceleration that will lead to wider use of POWER capabilities.

Top features of POWER10's 7 nm design include 3X greater energy efficiency than POWER9; up to 20X better AI performance than POWER9; increased scalability in POWER-based clusters; and 4X the number of encryption engines per core.

In a pragmatic move for DevOps teams, POWER10 processors have been optimized to support Red Hat OpenShift containers, making it easier to mix POWER and x86 processors in deployments of servers, storage, and converged infrastructure. In a cloud-enabled world, most customers are focusing at the application level, leaving the hardware decisions to IT departments and cloud service providers (CSPs).

POWER10 Processors

As described at the **Hot Chips** conference in August 2020, POWER10 adds extended support for an expanded memory architecture. Given the expanded memory and the on-chip math accelerators, IBM has disclosed that it expects to see a three-fold improvement in capacity and processor energy efficiency, compared to POWER9 processors, as introduced in 2017.

Using a technology called Memory Inception, POWER10 processors can support multi-petabyte (PB) memory clusters. This capability will support memory-intensive workloads, such as enterprise applications from SAP and SAS Institute, when used alongside large-model AI inference engines.

The POWER10 processors, based on a 7 nm design and fabricated by Samsung Electronics, will deliver new capabilities to hybrid clouds housing clusters, or server nodes, for AI/ML, for scalable enterprise applications (e.g., SAP, SAS) and for speeding up HPC applications that were originally deployed with POWER9.

We expect cloud customers to be a bigger factor in driving future demand for POWER technology. For this generation of POWER, many customers will leverage the POWER10 processors via IBM Cloud or Google Cloud services -- gaining new capabilities without directly deploying POWER10-based hardware on-site. In that scenario, some customers may be less aware of POWER10 supporting cloud services – although they could request that CSPs provision it for specific workloads.

The Long Road to POWER10

Some may ask: Why IBM is pressing forward with the POWER10 RISC chip, which will inevitably enter a new round of competition with AMD's EPYC processors and Intel's x86 Ice Lake processors? Both EPYC and Ice Lake are both expected to be updated in the 2021-2022 timeframe, and there is substantial investment going into both the AMD and Intel platforms.

A large group of IBM engineers – estimated to be 1,000 engineers working over five years – designed POWER10's circuits and shared-memory capabilities. But AMD and Intel aren't standing still: AMD just announced a new group of EPYC processors, code-named Milan. Intel is making changes, as well: Intel's new CEO, Pat Gelsinger, is expected to accelerate the process of bringing a 7 nm Intel x86 processor to the server marketplace.

The answer has more to do with target workloads for IBM's systems and storage than it does with the volume of processor shipments. IBM is looking to optimize the performance and the efficiency of scale-out clusters of high-performance systems, while keeping its semiconductor R&D costs in check by partnering with Samsung.

The evolution of POWER has seen a steady cadence, with POWER8's release in 2015; POWER9 in 2017; and POWER10 in 2020/2021. To make that last jump to 7 nm, IBM is partnering with Samsung Electronics to fab the POWER10 chips that IBM CPU engineers are designing. IBM has disclosed that it expects to see a three-fold improvement in capacity and processor energy efficiency, compared to POWER9.

Looking back on five years of POWER announcements, we can see that there was a steady progression from 22 nm to 14 nm microprocessors – and now to microprocessors built with a 7 nm process technology. Today, that remains a technical differentiator with Intel's Ice Lake x86 processors, which are based on a 10nm design, even though Intel is expected to go to 7 nm designs in the next product generation. All three types of processors (AMD, Intel and IBM) run Linux, so customers will view all major hardware platforms as actively competing to run the cloud's VMs, containers, databases and applications.

Building on AI/ML and Cloud Affinity

If you look at the recent generations of POWER, the goals for each of the POWER processors – 8, 9 and 10 – were slightly different. POWER 8 focused on Big Data, database workloads and analytics, working with NVIDIA on high-speed NVlink interconnects. POWER9 focused on AI/ML and Cognitive workloads, and high-performance (HPC) supercomputing, using POWER9-based IBM AC922 servers as

building blocks for highly scalable AI analytics systems and HPC in the data center. POWER9–based supercomputers were deployed at U.S. national labs for energy and nuclear research, including the Sierra supercomputer at Lawrence Livermore National Labs in California.

Now, IBM is looking to the new wave of hybrid cloud migrations to increase demand for POWER processors, Red Hat OpenShift containers and POWER-based scalable storage. IBM has been focused on what it calls "Chapter 2" cloud migrations for several years. This refers to a first wave of cloud workloads that led to migrations of apps for rapid deployments, saving on Capex costs – and the second wave for enterprise applications, many of them transactional and mission-critical..

OpenPOWER Ecosystem May Expand

OpenPOWER microprocessors, modeled on POWER designs, are also likely to show up in the world's cloud data centers (e.g., IBM Cloud and Google Cloud Platform) to support specific, demanding workloads.

OpenPOWER processors can be integrated into computer systems makers' OEM designs, including Inspur in China, and Fujitsu Ltd. and Hitachi Ltd. in Japan. All three of these companies have shown their OpenPOWER-based products at industry events in recent years. We should note that OpenPOWER specifications have become open-source technology, allowing them to be used more widely by OEMs worldwide, expanding the TAM of the POWER-based ecosystem.

Catching the Next Wave of Enterprise Cloud Migrations

As for IBM itself, IBM executives have pointed to a second wave of cloud migrations that leverages the company's strengths in supporting enterprise apps – which IBM has long called "industrial-strength workloads." Those migrations will combine POWER10 processors with Red Hat OpenShift and modernized software to move traditional enterprise workloads to hybrid cloud data-centers. We expect those types of cloud migrations to be increasingly important to IBM and its customers.

Clearly, IBM plans an "inside-the-hybrid-cloud" strategy for POWER10-based systems and POWER10-based clusters, supporting cloud-centered deployments, such as AI/ML, automation, scale-up and scale-out solutions. IBM will still compare processing time and benchmarks for specific workloads (POWER vs. x86), but the tacit understanding for IBM's POWER strategy is that IBM has a new and expanded audience for hybrid cloud: cloud providers, system integrators (SIs) and enterprise data centers – all of them building and deploying hybrid cloud infrastructure.

46 Kent Hills Lane, Wilton, CT. 06897; (203) 429 8951;
http://www.rfgonline.com/; Contact: inquiry@rfgonline.com