SPONSORED BY

IBM

GEEK GUIDE

Linux on Power:

Why Open Architecture Matters

LINUX™ JOURNAL

Table of Contents

TED SCHMIDT is the Senior Project Manager and Product Owner of Digital Products for a consumer products development company. Ted has worked in Project and Product Management since before the agile movement began in 2001. He has managed project and product delivery for consumer goods, medical devices, electronics and telecommunication manufacturers for more than 20 years. When he is not immersed in product development, Ted writes novels and runs a small graphic design practice at http://floatingOrange.com. Ted has spoken at PMI conferences, and he blogs at http://floatingOrangeDesign.Tumblr.com.

GEEK GUIDES:
Mission-critical information for the most technical people on the planet.

About the Sponsor

IBM

IBM is a globally integrated technology and consulting company headquartered in Armonk, New York. With operations in more than 170 countries, IBM attracts and retains some of the world's most talented people to help solve problems and provide an edge for businesses, governments and non-profits.

Innovation is at the core of IBM's strategy. The company has reinvented itself through multiple technology eras and economic cycles, creating differentiating value for its clients. Today, as the IT industry is fundamentally changing at an unprecedented pace, IBM is much more than a "hardware, software, services" company. IBM is now emerging as a cognitive solutions and cloud platform company.

Cognitive solutions powered by the cloud are the key to clients' digital transformation. This transformation requires breakthroughs at every level of the enterprise IT foundation, from processors and computer design to storage, networking and the integration layer. IBM Power Systems, built with open technologies and designed for mission-critical applications, offers the infrastructure that is designed for cognitive workloads.

Linux on Power: Why Open Architecture Matters

TED SCHMIDT

Overview

With all the industry talk about the benefits of Linux on Power and all the performance advantages offered by its open architecture, you may be considering a move in that direction. If you are thinking about analytics, big data and cloud computing, you would be right to evaluate Power. The idea of using commodity x86 hardware and replacing it every three years is an outdated cost model. It doesn't consider the total cost of ownership, and it doesn't consider the advantage of real processing power, high-availability

and multithreading like a demon.

This ebook takes a look at some of the practical applications of the Linux on Power platform and ways you might bring all the performance power of this open architecture to bear for your organization. There are no smoke and mirrors here—just hard, cold, empirical evidence provided by independent sources. I also consider some innovative ways Linux on Power will be used in the future.

If you want to get the best possible performance out of your database engines, want to scale according to your business' seasonality without incurring huge infrastructure cost increases when demand spikes, or want to run transactional and analytics concurrently, so you can provide decision-makers with real-time data, you will want to take a very close look at what IBM Power systems have to offer.

Linux on Power: an Architecture Built for Performance

The Linux on Power architecture originally was conceived through an alliance of Apple, IBM and Motorola, which came to be known as the AIM Alliance. The AIM Alliance wanted to challenge the dominant Wintel platform in the early 1990s by delivering a RISC-based operating system. As you all know, RISC (Reduced Instruction Set Computer) improves performance tremendously, which allows system functionality to be moved from hardware to software. RISC originally was designed to power workstations and mid-range servers, but IBM wanted to go beyond mid-range and workstations and into big data and cloud computing, so it developed Power, which stands for Performance Optimization with Enhanced

Suddenly, with 96 threads providing eight threads for each core and addressable memory up to 16TB, huge performance gains result in screaming-fast query speeds and reduced IO times, which are important if you want to run big data or analytics, and also if you want to crush typical x86 performance.

RISC, to challenge traditional x86 systems.

A typical x86 processor offers 30 concurrent threads at a maximum processor speed of 3.4GHz, on-chip cache of 37.5MB and 1.5TB of addressable memory per socket. Power, with its enhanced RISC, provides an impressive 96 concurrent threads, processor speeds as high as 5GHz, on-chip cache of 64MB and up to 16TB of addressable memory per socket. Suddenly, with 96 threads providing eight threads for each core and addressable memory up to 16TB, huge performance gains result in screaming-fast query speeds and reduced IO times, which are important if you want to run big data or analytics, and also if you want to crush typical x86 performance.

In computer memory, "endianness" refers to how bytes are ordered in digital memory. Big-endian stores data big end first, so it treats the number one hundred twenty-three as "123" (one hundred being the "big end"). Little-endian, on the other hand, stores the same number with the little end first, which means one hundred twenty-three is stored as "321".

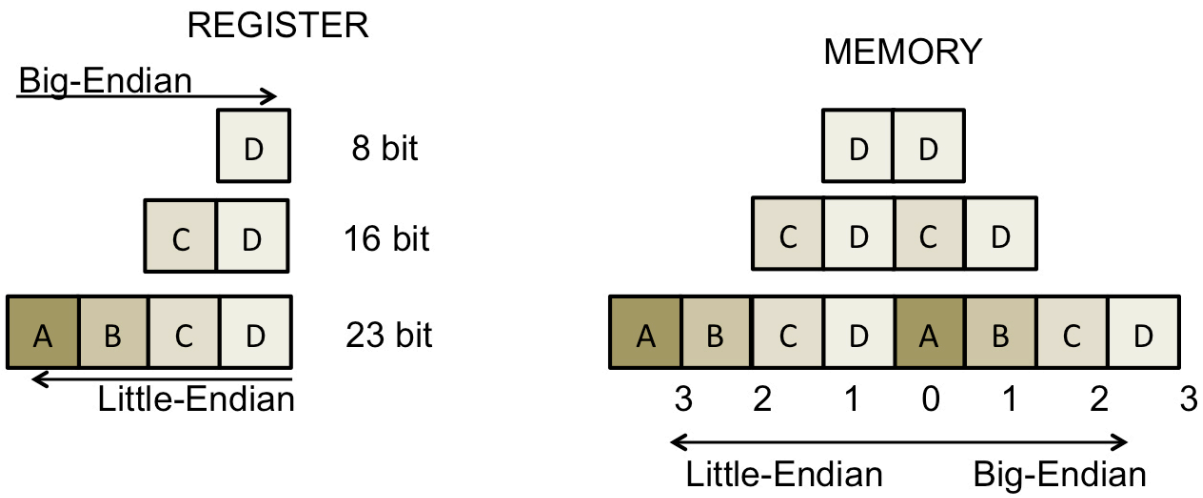

FIGURE 1. Big-Endian versus Little-Endian

Power is bi-endian, so it can store data as either little-endian or big-endian, which is important if you want to dominate a market where x86 currently prevails. Intel x86 processors support little-endian Linux, and RISC defaults to big-endian. Because of this bi-endianness, it now becomes much easier to move an application currently running on an LE x86-processor to a Power system. Without this capability, it would require a massive effort to move applications from x86 to Power.

Canonical (Ubuntu Linux distributor) migrated 40,000 Linux applications over to the Power platform in six months because it avoided having to recode and recompile millions of lines of code. What normally would have taken years to accomplish, Canonical managed in 160 days, directly because of Power's built-in flexibility.

As well, IBM has more than 80% of its own Linux applications ported over to the Power platform. SUSE and

It's one thing to be able to crunch massive amounts of data, but if you can't handle the seasonality of your business or keep your systems up and running and secure, what's the point?

Red Hat have Linux releases running on Power. Anyone currently looking for more power and speed to support analytics or cloud computing will appreciate the commitment these organizations are making to the Power platform.

Finally, Power has a neat little feature called CAPI (Coherent Accelerator Processor Interface), which allows hardware to bypass the operating system and communicate directly with the CPU. CAPI allows the CPU to access external flash drives directly as an extension of its memory, rather than bringing the data stored on that drive into memory, like other systems. What results is even faster data access speeds and more data being accessed. Think of it as the difference between on-line banking and having to drive to the bank to conduct every financial transaction.

Power's RISC-based architecture clearly provides a much faster and more powerful platform than x86. Canonical and IBM both have shown how easy it is to migrate to Linux on Power for current users of x86 processors running Linux. But, let's not forget about scalability, high availability/disaster recovery and security. It's one thing to be able to crunch massive amounts of data, but if you can't handle the

seasonality of your business or keep your systems up and running and secure, what's the point?

Power provides multiple virtualization options, depending on your needs. If you're looking for multiple LPARs running different operating systems, then PowerVM will be the answer, unless you're running the Linux on Power variation, which obviously supports only Linux VMs. If you want to take the next step in scalability and provide more of a cloud-computing solution by virtualizing your memory and I/O, or if you need to perform live VM migrations, PowerKVM is the answer. But you'll have to go to a POWER8 platform, as PowerKVM isn't supported on POWER7 systems. Kindly enough, IBM provides the ability to switch between PowerVM and PowerKVM with a simple reboot of your VM.

PowerKVM is a very nice option because it allows for memory virtualization, I/O virtualization and live VM migration. This means your memory can be over-committed, as well as swapped out when underutilized, just like in a regular Linux system. Additionally, I/O virtualization offers the benefits of a wide range of I/O resource sharing options, including PCI passthrough to allow VMs to access dedicated devices directly, and I/O configurations that include Fiber Channel, NFS, iSCSI and Ethernet.

So, if you're running a mix of Linux and Windows machines, you may need to stick with PowerVM on a POWER7 platform. But, if you are running Linux exclusively and want to take advantage of some really nice virtualization options to improve performance all around, running the PowerKVM option on a POWER8 platform could be the answer.

PowerKVM Benefits

- Open-source server virtualization.
- Optimizes Linux workload consolidation.
- Improves total cost of ownership.
- Leverages performance, scalability and security.
- Avoids expensive and proprietary x86 virtualization.
- OpenStack, libvirt and open Linux tools.
- Enables single, cross-platform virtualization.
- Continues to leverage open-source solutions.
- Traditional Linux admin skills on Power systems.
- Exploits new advanced POWER8 processor features.
- Live virtual machine migration.

If you do decide to leverage the benefits of KVM, keep in mind that it runs only on POWER8 servers because PowerKVM needs the OPAL (Open Power Abstraction Layer) interface to access hardware.

Also note that Red Hat does not yet support RHEL if it is running on PowerKVM, so you need to use its own virtualization: Red Hat Enterprise Virtualization (RHEV). Although POWER8 also provides a bare-metal option (which is really no virtualization), it's enabled only for Ubuntu at the time of this writing. But, bare metal is worth mentioning because removing the virtualization layer from between the hardware and the virtual machine results in a performance boost of 3–20%.

PowerVM provides the ability to create multiple LPARs of different operating systems, with the exception of the PowerLinux variation, which supports only Linux. PowerKVM is the POWER8 solution that takes the next step by providing memory and I/O virtualization, along with live VM migration. These are important if you need the scalability of cloud computing. Just be careful and remember the limitations of some of these options before choosing which virtualization option to use. The best approach is the cautious approach.

High Availability and Disaster Recovery: To provide high-availability and disaster recovery capabilities, Power makes use of native Distributed Replicated Block Device (DRBD). DRBD mirrors an entire block device to another networked host during runtime by using a configuration of redundant array of independent disks (RAID-1) that works based on mirroring. This allows the development of high-availability clusters for block data.

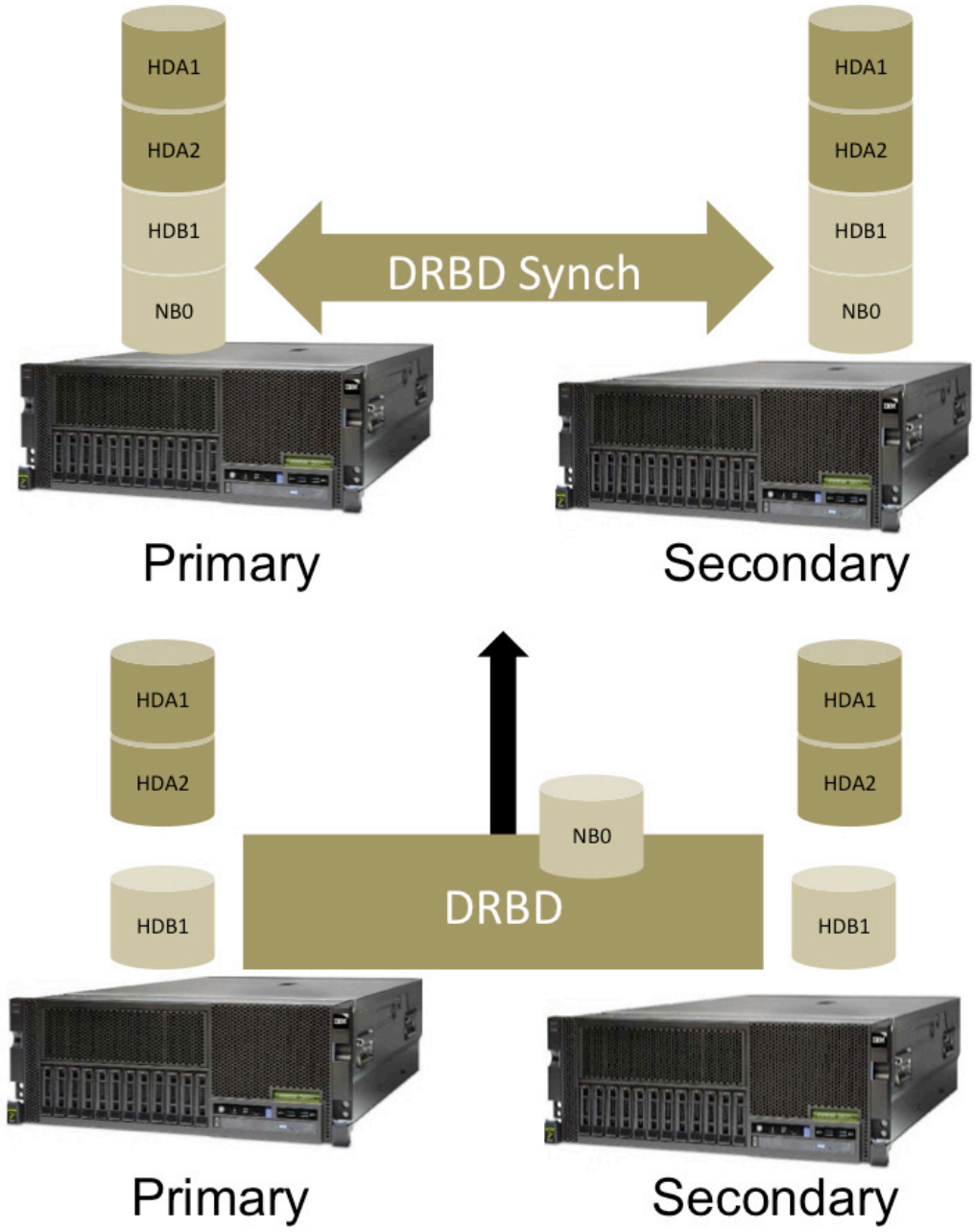

FIGURE 2. Distributed Replicated Block Device

If you are looking for more data protection and willing pay a small performance price to get it, DRBD supports a variation on synchronous and asynchronous data

replication that ends up providing more data protection than asynchronous data replication. DRBD makes use of the Linux Crypto API and calculates message integrity codes to go with data moving between nodes that ensures the integrity of data. Using this method allows the receiving node to validate its incoming data and request retransmission when an error is found. DRBD also verifies local and peer devices in an on-line fashion (while input/output occurs), and provides several automatic methods for recovering from split-brain errors, which happens when each node thinks it is primary.

When it comes to security and compliance, Power can be configured to notify an administrator automatically if a virtual machine falls out of compliance on any of the major governance protocols. HIPAA, DOD, PCI, CObIT and other governance protocols are pre-configured into Power's security and compliance component, PowerSC. Because the hypervisor for Power resides directly in the firmware of the system, it makes the virtual machines just as secure as the physical servers on which they reside. By using built-in compliance reporting, PowerSC can offer suggestions for bringing VMs back into compliance when they fall out of compliance. PowerSC also ensures the complete separation of virtual networks. As a result, security compliance is assured. PowerSC also provides local firewall services on the virtualization layer. This feature helps improve performance and reduce network resource consumption. When it comes to security, Power has globally recognized certifications from the Department of Homeland Security's United States Computer Emergency Readiness Team (US-CERT) and MITRE

Corporation. Both US-CERT and MITRE reported zero vulnerabilities against PowerVM.

So far, I've summarized the many performance benefits Power provides over Intel's x86 systems. The RISC-based architecture provides incredible performance gains and throughput advantages over the x86 architecture. Power provides a huge benefit for analytics applications through the CAPI feature, which allows the CPU to access flash storage as if it were memory. And, Power supports both little-endian and big-endian data storage, which means the pain of migrating little-endian applications from x86 platforms is eliminated. I also reviewed the advantages provided by Power's high availability/disaster recovery capabilities, along with its inherently secure environment that provides reporting on compliance to multiple governance protocols. Now, let's move on to look at some real-world configuration and applications of the Power platform.

Configuration Options for Power

Typically, when you talk about configuration, you really are talking about managing the balance between two competing priorities: performance and power-consumption. At the simplest level, there are three places you can configure a system: you can configure hardware settings, you can tune the operating system, and you can tune applications. Independent studies continue to prove that Linux on Power is a superior platform to x86.

One such example is Quru, which recently used the PGBench SELECT benchmark and the TPC-C benchmark

Overall, Quru was able to tune a 99% increase of performance on the PGBench benchmark and a performance gain of 52% with the HammerDB benchmark.

to prove that a tuned Linux on Power system will significantly outperform an x86 system. Starting with the PowerLinux 7R2, in factory default configuration with a clean install of RHEL 6.4 and version 9.2-4 of Postgres Plus Advanced Server, combined with the open version PostgreSQL as benchmarks, Quru went on to show results ranging from an 8% performance gain realized from OS tuning, to a 14% improvement gained from hardware tuning, to a staggering 24% improvement realized from application optimization. Overall, Quru was able to tune a 99% increase of performance on the PGBench benchmark and a performance gain of 52% with the HammerDB benchmark. Clearly, Quru has shown that a significant performance improvement can be achieved when properly tuning a deployment of Linux on Power.

For more details on the results achieved by the Quru study, go to https://www.quru.com/whitepapers/postgreSQLonPower.

Early in 2015, Forrester conducted a study of North American IT strategists and architects to understand motivations behind OS selection and use. What they found was that, due to inherent limitations around agility,

scalability and limits in analytics capabilities, IT professionals being interviewed were beginning to hit the wall with the capabilities of so-called "commodity" systems built on x86 architecture. Because of emerging performance and agility requirements stemming from newer, more demanding business-critical applications, x86 server farms were found to be lacking. Furthermore, the cost model that made these commodity server farms initially appealing has started giving way to a broader, TCO model of understanding value.

In other words, driving a four-cylinder sedan seems great from an initial cost model when you are thinking about traveling a long distance, but when you have to tow a huge load or get somewhere fast, the value proposition changes. As a result, IT organizations are beginning to rely less on the x86 server-farm model to run their newer, more power-hungry applications, paving the way for powerful, scalable, open systems, like POWER7 and POWER8.

See http://www-03.ibm.com/systems/power/advantages/smartpaper/performance-flexibility.html for more details on the Forrester study.

Real-World Applications of Power

According to the Forrester study mentioned above, 37% of the 150 respondents reported that they ran more than half or all of their applications on Linux and half of the respondents ran on RISC processors. With the need for crunching big numbers, there's no doubt of the need for more processing power, but who is actually using Power systems and to what advantage?

Redis Labs: Redis Labs points to the example of a client

with more than 31TB of raw genome data that was able to achieve 97% cost savings using a POWER8/CAPI Flash solution over standard AWS solutions. Leveraging Power's open community, Redis Labs is able to offer this solution for Power using the IBM NoSQL Data Engine and using Redis on Flash to extend RAM up to a 90% Flash ratio. The solution offers asynchronous, non-blocking operations, and a 4U unit supports 40TB Redis over Flash. Redis Labs also points out that costs are reduced when deploying on Power—deployment costs have been reduced as much as 70%, when deploying on Power with an IBM Flash System.

You can read more about how Redis Labs leverages IBM Power here: https://redislabs.com/solutions-redis-labs-on-power.

Bon-Ton Stores: In 2013, Bon-Ton Stores (a $3B, omni-channel retailer located in the Eastern and Midwestern US) replaced their x86 servers with PowerLinux 7R2 running PowerVM, SUSE and WebSphere Commerce. Proving out the solution first, running their own stress tests, they improved per-core throughput by two-and-a-half times over their legacy x86 servers. Since Bon-Ton licenses its software on a per-core basis, the improved throughput provided incredible cost savings when demand peaked during the holiday season.

For more details on how Bon-Ton Stores leveraged Power, see http://www-03.ibm.com/systems/power/advantages/whypower/bon-ton.html.

SAP HANA: HANA, SAP's real-time analytics and aggregation system that brings transactional and analytics processing together into a single, in-memory database

system, has been available on Power since 2015. Since Power clearly exceeds the capabilities of x86 systems, especially where caching of data, high bandwidth and multithread processing are required, it makes sense for organizations using HANA to want a Power platform on which to run it. Organizations with very large databases, which are becoming more prevalent, will realize huge benefits from the analytics capabilities of HANA on Power, combined with the faster processing that comes from performing calculations at the database level, rather than the application level. The benefits of running HANA on Power are serious and abundant. In fact, the SAP BW Enhanced Mixed Load Standard Application Benchmark, which was performed in May 2015, highly suggests that HANA clients would be much better off running on an IBM Power platform.

You can view the complete benchmark, certified by SAP on June 1, 2015, here: http://global.sap.com/solutions/benchmark/pdf/Cert15024.pdf?cm_mc_uid=62494968479914641009170&cm_mc_sid_50200000=1464116566.

Basically, the benchmark shows that customers who choose to run HANA on Power can expect to get twice the performance benefit over x86 servers. How does Power succeed at handling the incredible requirements of HANA? First, Power feeds the massive throughput demands of HANA in several ways. Because POWER7 provides twice the threads per core, and POWER8 provides four-times the threads per core as standard architectures, Power responds well to HANA's tendency to initiate multiple threads. Furthermore, Power handles the memory throughput demands of HANA by providing up to eight high-speed

channels, 32 DDR ports, and up to 128MB L4 cache and 1TB memory per processor socket. Finally, with 15 layers of copper wire, Power's chip technology provides much faster throughput per core than standard architectures.

Power also ensures proper failover of database IP addresses during database failover with Tivoli, predictive unit failure alerts and a "chipkill" feature, which avoids the need for having to double the amount of memory as a mitigation strategy against chip failure. Finally, Power provides high LPAR redundancy with multiple, virtual IO servers and flexible, on-demand scaling to respond to peaks in demand, which are key for a platform that has increasing accessibility demands.

For a more detailed analysis of HANA on Power, see https://www.ibm.com/marketing/iwm/dre/signup?source=stg-UK&S_PKG=ov38808&lang=en_gb.

MariaDB: In 2015, using the sysbench 0.5 OLTP benchmark, MariaDB conducted an independent test of the performance of the MariaDB, a community-developed fork of MySQL (version 10.0.19), running on POWER8 using RHEV/RHEL with PowerKVM 2.1.1 in comparison to an HP ProLiant DL380, also using RHEV/RHEL. Results found that, on a per-core basis, Power ran 2.6 times faster on read-only performance and 2.2 times faster on read/write. When scaled from one to two sockets, the Power system ran 80% faster, compared to a 50% gain for the HP. The conclusion reached by MariaDB was that a POWER8 system, with its speed and scalability advantages, is very well suited to handle the data-intensive workloads of a relational database like MariaDB.

To read more details on the MariaDB's benchmark, see https://mariadb.com/products/mariadb-ibm-power8.

In order to drive innovation, IBM realized that all intellectual property regarding the Power platform had to be made completely available to this new community of OpenPOWER Foundation members.

Redis Labs, Bon-Ton Stores, SAP HANA and MariaDB are but a few examples of how organizations are achieving real gains from moving from x86 to Linux on Power. If your goals are performance and scalability, while maintaining, and in many cases, reducing your total cost of ownership, Linux on Power may be the answer. With Power, IBM also has shown the benefits brought about by providing a truly open ecosystem. Let's briefly look at that open ecosystem and why it's so beneficial.

Looking Toward the Future

Learning to Be More Open: In 2013, IBM took a leap of faith by partnering with Google, Mellanox, Tyan and NVIDIA to form the OpenPOWER Consortium, which is now known as the OpenPOWER Foundation. In doing this, IBM released all the Power architecture into the wild, making the system completely open. This was a huge departure from what IBM had done in the past, but IBM viewed it as necessary. In order to drive innovation, IBM realized that all intellectual property regarding the Power

platform had to be made completely available to this new community of OpenPOWER Foundation members.

With more than 200 members in 22 countries, the strategy appears to be working. Google has invested in building hyperscale processors based on the POWER8 architecture. Using POWER8 as a model, Tyan has developed a customer reference system to allow end users to deploy software that is customized to their unique requirements. And NVIDIA is working on an alternative to the Power CAPI, which would provide immediate access to Graphics Processing Units (GPUs).

Because of its open nature, the OpenPOWER Foundation makes Power available to a much broader population of interested contributors and therefore will continue to offer a place and community where innovation drives advancement for the Power platform. Where open standards are crucial, such as analytics and mobility, this ends up being very beneficial. IBM will continue to work closely with channel partners who already have investments in AIX and IBMi solutions. Partners like HelpSystems, who know what customers are looking for because of their significant investments in AIX and IBMi, are more invested in expanding the capabilities of the Power platform. OpenPOWER members who appreciate the value of big data and cloud computing, and who understand the value that Power on Linux offers their customers, will continue to advance Power precisely because of the open community. Creating and working with an open community certainly will make Power more relevant.

The OpenPOWER Foundation helps promote collaboration

on multiple levels, from chip architecture, to hardware acceleration, to systems management software. At the time of this writing, the Foundation has 11 chartered workgroups, such as Systems Software, Memory, Accelerator and Hardware Architecture, just to name a few. These workgroups are focused on delivery of projects such as LE Linux KVM, OpenCL SDK to OpenPOWER Platform interface and IO Device Architecture (IODA2), all of which will increase the platform's accessibility to channel partners. IBM and the OpenPOWER Foundation are committed to expanding the open-source ecosystem for Linux on Power. With more than 200 members and events like the OpenPOWER Developer Challenge (read: hackathon), IBM is becoming an example of open platform collaboration. For more on the OpenPOWER Developer Challenge, follow this link: http://openpower.devpost.com.

IBM maintains its own on-line community that is equally as open as the Foundation. The only real difference between IBM's community and the Foundation may be that where the Foundation is created by and for channel partners, IBM's on-line community tends to deliver content aimed more at helping individuals in the understanding and implementation of IBM Power solutions. IBM's on-line community offers a wonderful myriad of technical resources that include blogs, wikis, forums and direct access to Power technical experts. It's all aimed at providing as much information and assistance to individual members of the Power community.

And, it's precisely this open community that is determining how best to leverage Power for future

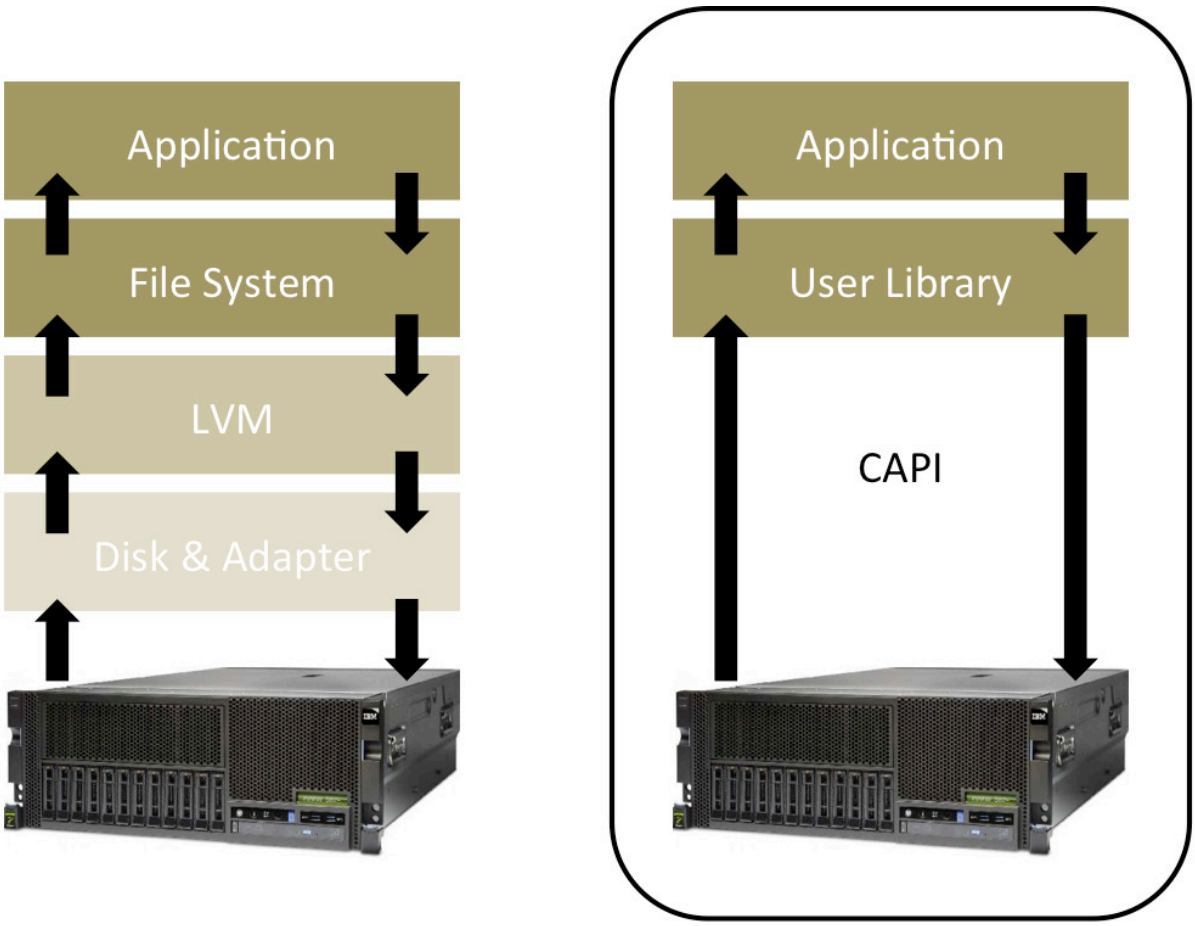

FIGURE 3. **CAPI eliminates 95% of instructions.**

applications. For instance, the CAPI feature of Power, which allows hardware to communicate directly with the CPU by bypassing the operating system, is providing all kinds of opportunities in high-performance computing (HPC). Because of CAPI, and the open community around it, almost 2,400 software ISVs and tens of thousands of open-source packages have been made available, specifically designed to leverage the accelerated processing available on Power.

Because CAPI eliminates the middleman of the operating system between the OS and hardware, it effectively can

eliminate more than 95% of the instructions required to access data, which results in huge gains. In the HPC community, gains like that are huge. And, the openness of the OpenPOWER Foundation only encourages future advancements of this technology.

The bottom line is that the future is focused on high-performance computing that supports immediate access to massive amounts of data and cloud computing that provides scalability and security. With its advanced technology, like CAPI, and its open community focused on driving innovation, the Power platform is well positioned to be the computing platform of the future.

Conclusion

Scalability, speed and a completely open architecture—these are the qualities that Linux on Power offers and the qualities that make it the platform of the future. Anyone with needs around analytics or crunching massive amounts of data, scalability and security should take a hard look at the Power architecture's total cost of ownership. The math behind disposable, commodity x86 hardware doesn't work anymore because of the performance demands of new applications. Applications like HANA that need power to handle transactional and analytical processing at the same time or big data applications that work with data at the petabyte-level demand more than x86 farms can deliver. But Power, with its powerful, fast and scalable architecture, coupled with an open community built around enhancing the Power platform for future applications, can provide an efficient and cost-effective solution that's built to last.■