WHITE PAPER

Innovations to IBM Power Systems for the Virtualization, Multitenancy, and Cloud Demands of the 3rd Platform

Sponsored by: IBM

Matthew Eastwood
April 2014

EXECUTIVE SUMMARY

The world of IT is changing. We are in the early stages of what IDC describes as the 3rd Platform of computing, in which Big Data, cloud computing, mobile, and social technologies are fundamentally transforming the computing and business landscape. This 3rd Platform is enabling enterprises to find new ways of using their datacenters to drive business value and as a source of competitive differentiation. Organizations that do not incorporate the use of Big Data, cloud, and other 3rd Platform technologies risk being left behind.

But leveraging 3rd Platform technologies such as Big Data and cloud must be done thoughtfully. Maximizing the value of Big Data requires comprehensive data management that allows different workloads to be deployed across a single architecture with an integrated set of analytic and management tools. Cloud requires support for virtualization and virtual machine (VM) optimization at the server level and new levels of security to allay privacy and security concerns. And as Big Data and other 3rd Platform technologies increasingly become mission critical, they require an infrastructure with the reliability, availability, and serviceability (RAS) features typically used to support resource-intensive, mission-critical workloads such as OLTP, business applications, and data warehouses. And IT organizations want protection for their mission-critical investments, a goal that is generally best achieved by working with vendors that have the resources, scale, and ecosystem to ensure that they are in it for the long haul.

IBM Power Systems servers are designed to address each of these needs. IBM Power Systems have a long track record of supporting computationally intensive applications, including enterprise systems of record, information storage, and retrieval systems that act as the authoritative source of data for key business applications and services. IDC expects that with the POWER8 release, IBM will continue to support the types of performance and RAS features that have kept Power Systems at the forefront of platforms supporting mission-critical systems. In addition, with POWER8, IBM is introducing a series of new innovations such as greater multithreading, support for higher VM density and VM mobility, and Coherent Accelerator Processor Interface (CAPI) acceleration to better support 3rd Platform requirements.

IBM clearly intends POWER8 to provide further headroom for enterprises already running compute-intensive, mission-critical workloads on Power Systems servers. In addition, IBM seems to be focused on organizations that are running 3rd Platform and other applications on x86 platforms that need to upgrade their infrastructure with better virtualization, greater workload density, and performance acceleration for Big Data. And with CAPI, IBM is demonstrating its understanding of the importance of platform choice and ecosystem support for a wide variety of traditional as well as next-generation workloads.

SITUATION OVERVIEW: THE 3RD PLATFORM AND BIG DATA ARE CHANGING THE IT LANDSCAPE

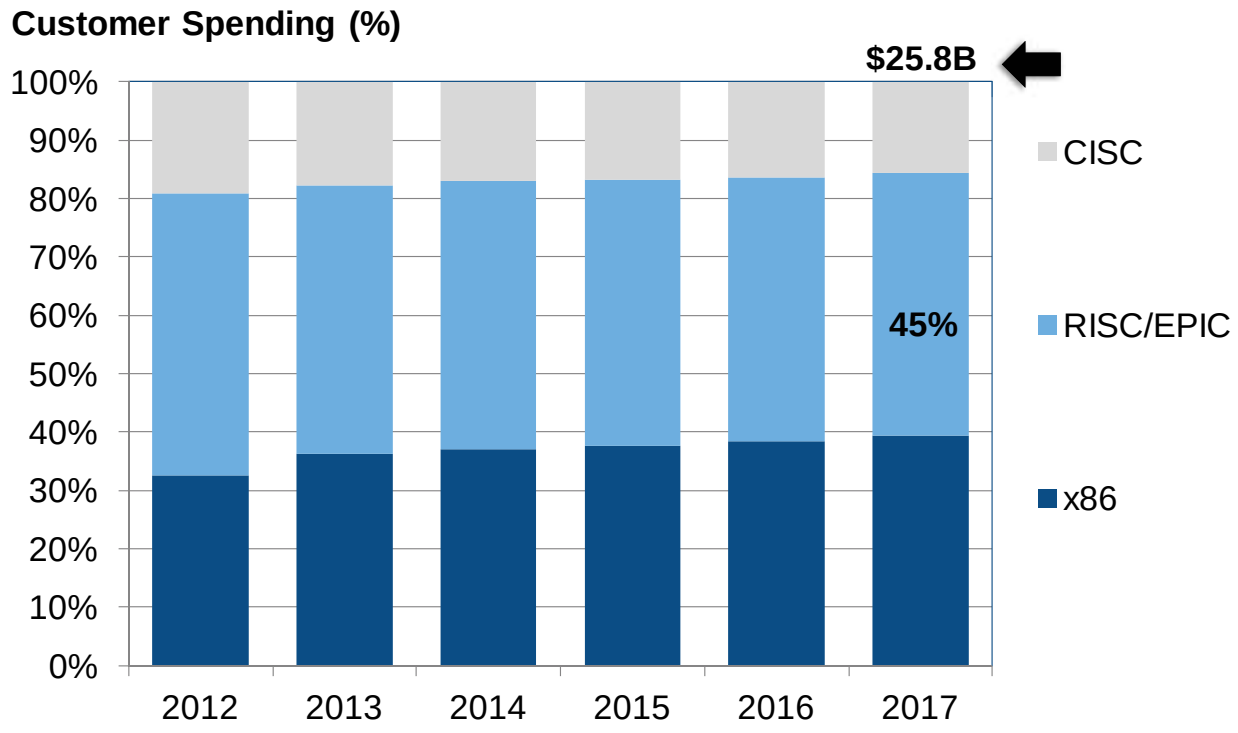
In recent years, there has been a major change in computing marked by an explosion in devices and fundamental shifts in the way users interact with technology. IDC refers to this as the 3rd Platform of computing, and unlike the 2nd Platform, which was defined by the predominance of PCs and client/server computing, the 3rd Platform is built on a foundation of Big Data, cloud, mobile, and social technologies.

Many of the original instances of these 3rd Platform technologies were deployed for consumer technologies, but as consumers grew to appreciate the convenience and control they provided, they have been increasingly demanding that their employers and other companies they do businesses with support them as well. Today, 3rd Platform technologies are common in most enterprises and are the basis for the next generation of enterprise business innovation that organizations across all industries will need to stay at the forefront in order to remain competitive.

But even as we transition to a new model of IT delivery, companies must leverage their large existing investments in current-generation technologies. It is neither financially nor logistically viable for companies to completely change out their IT infrastructure, and it makes sense for companies to move to the 3rd Platform in a way that builds on current investments. To support this point, IDC workloads research indicates that the core high-value workloads — 45% of all server spend — remain predominantly on established reduced instruction set (RISC), typically Unix environments, and complete instruction set (CISC), typically System z server environments (see Figure 1). It is important to understand that Big Data is orthogonal to a broad set of workloads running in a typical enterprise datacenter and not in and of itself inclusive of traditional data warehousing and data analytics. IDC notes that a significant part of infrastructure spending for high-value workloads like those in Figure 1 also covers 3rd Platform needs such as security, privacy, compliance, RAS, and virtualization.

FIGURE 1

Business Processing, OLTP, Data Warehousing, Data Analytics, and Application Development and Deployment: High-Value Workloads and Their Server Platforms, 2012-2017



Note: High-value workloads include business processing, OLTP, data warehousing, data analytics, and application development and deployment (AD&D).

Source: IDC, 2014

Supporting Big Data Workloads

Big Data is one of the four pillars of the 3rd Platform. It is transforming today's enterprise and the way it interacts with the marketplace. Every year, the volume of data that enterprises must handle continues to grow, with IDC data showing it is doubling every 18 months. This is creating huge challenges for IT organizations, not just in terms of housing all this data but in terms of putting the right tools, technologies, and processes in place to use it effectively.

IT infrastructure is separated into three tiers: edge, application, and database. Big Data includes structured data, which is housed on systems of record in the database tier usually on very large, high-end servers, as well as unstructured data, which is typically kept on scale-out server volumes. While organizations have deployed analytics applications designed to access the data in these repositories, it is impossible to take advantage of the data in a meaningful way unless you can inform real-time business decisions differently than you could without it. Truly leveraging these Big Data systems and providing this real-time response require staging data appropriately and housing analytics applications like Hadoop close to the systems of record and not downloading the data to be analyzed on separate siloed servers in a different portion of the organization or infrastructure. This in turn requires a comprehensive data management solution to allow these different workloads to be deployed across a single platform architecture with an integrated set of analytic and data management tools on a system with sufficient scalability and throughput to handle very computationally intensive workloads.

Supporting Cloud and Virtualization

Another major pillar of the 3rd Platform is cloud technologies. Cloud is taking the place of the traditional IT service delivery model, provides organizations greater flexibility in their IT spending and deployments, and offers the promise of greater scalability and agility. Cloud offers a model for new service platforms, marketplaces, and developer communities where the most sought-after solutions will live, and the choice of cloud platforms will either expand a company's ability to find and leverage the best new solutions or constrain it.

There are several models to deploy cloud – public, private, or hybrid – and whether you are an enterprise deploying a private or hybrid model or a cloud provider scaling your public cloud offerings, the choice of underlying technologies is critical to your success. The platforms you choose must be cloud enabled with support for large numbers of VMs, application isolation for multitenancy, workload density/utilization, and the ability to scale rapidly. Organizations are also looking to further define their datacenter environments by using a software model that heavily leverages open source cloud frameworks. Power Systems include a commitment to a number of important areas of open source innovation including KVM and OpenStack as well as expanding support for additional Linux distributions around the globe. These technologies help enterprises achieve the disaggregation of software from hardware necessary to pursue private, public, and hybrid cloud computing initiatives.

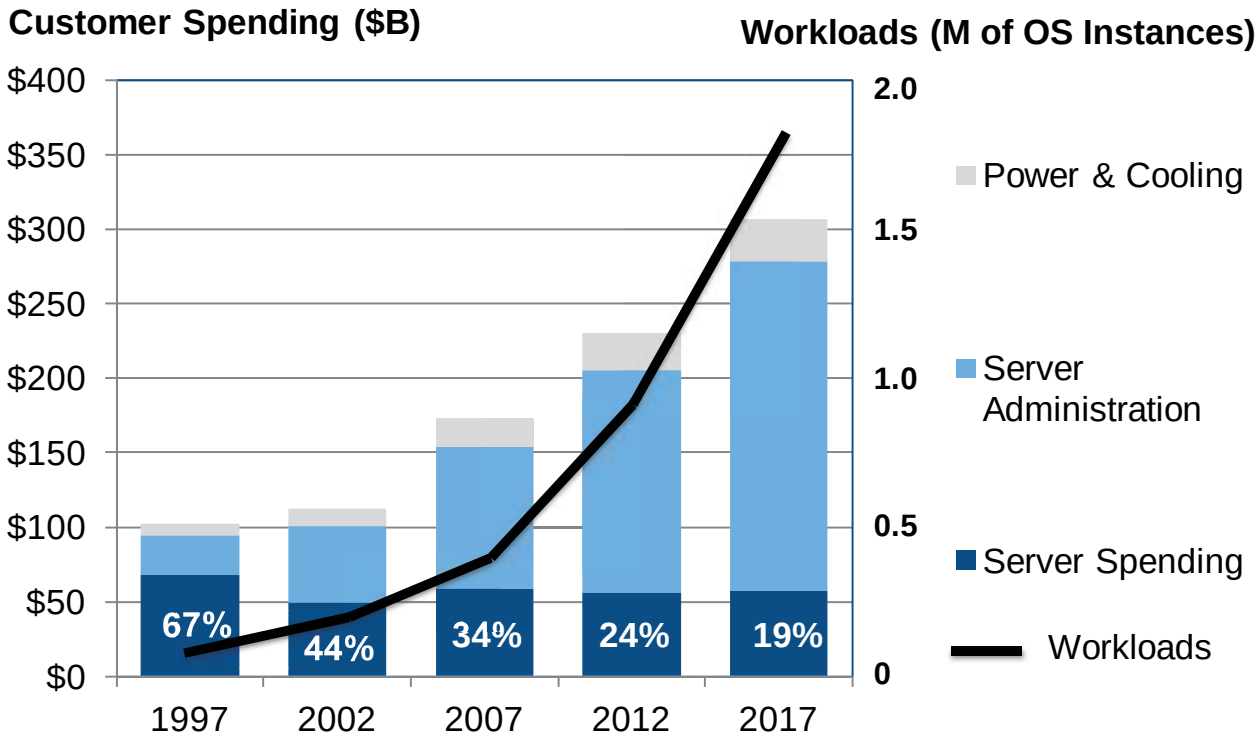
Alleviating Operating Costs

IDC data shows that in recent years, IT operating costs have skyrocketed. While the cost for computing is declining – dropping by 50% every two years– other cost factors are on the rise, such as data volumes, which are doubling every 18 months, and the number of applications IT must support, which is doubling every 4 years.

The net effect is that opex costs for IT organizations have been doubling every 8 years. In 1995, the ratio of IT opex to capex spending was 0:5. This ratio grew to 1:5 in 2005 and is projected to grow to 3:9 by 2015 (see Figure 2). The growth in opex spending, most of which is required simply to "keep the lights on," is crowding out IT's ability to spend on new innovations and to support new initiatives in the organization. This makes it even more critical that organizations maximize the use of their existing investments so that the maximum available budget can be spent on new innovations and strategic initiatives.

FIGURE 2

Workload Growth and the Consequent Customer Spending on Servers, Server Administration, and Power and Cooling: 1997-2017



Notes:

We expect capex server spending, as a percentage of total spending on server infrastructure, to drop from over 65% in 1997 to just under 20% of total spend by 2017.

Workloads, as measured by the number of instances or virtual machines (VMs), will have increased 24 times between 1997 and 2017.

Source: IDC, 2014

To combat this trend, IT organizations are under intense pressure to reduce operational costs, including IT staff costs, management and maintenance costs, and the costs of unplanned downtime. Some of the most common, and effective, cost-reducing strategies are consolidation/virtualization, automation and integration, and cloud sourcing. In fact, in IDC's *Server Virtualization 2013 Study* (December 2013), datacenter managers indicated that initiatives centered on these cost-containment strategies are the highest single set of priorities in IT organizations, ranking ahead of initiatives to grow the business, support business innovation, and manage risk.

Consolidation, virtualization, and private cloud initiatives enable organizations to reduce their server footprints, which can yield significant IT operating cost reductions. There are two primary infrastructure strategies for customers looking to efficiently achieve consolidation in the datacenter. One is larger SMP system configurations that are designed to support mixed workloads and have the ability to

manage workload peaks efficiently, which offer customers increased infrastructure flexibility. The second area of focus for customers seeking efficiently consolidated environments is capacity-on-demand (COD) capabilities, which allow dynamic expansion and contraction of resource pools.

Taking maximum advantage of these initiatives requires enterprise-class servers that offer flexible licensing models; provide large, flexible resource pools that can supply capacity on demand; and can support mixed workloads.

3rd Platform Workloads Combine Reliability, Availability, and Serviceability Requirements with High Degrees of Scalability

Traditionally, some of the most demanding, mission-critical workloads that IT organizations have had to support are transaction-oriented workloads. For these workloads, IT organizations are investing in solutions that remove single points of failure. This includes high-reliability infrastructure and high-availability architecture and software solutions to ensure continuous delivery. This also includes choosing servers with specific RAS features built into them to support greater levels of resiliency, fault tolerance, and automated service and support.

With the 3rd Platform, an increasing number of 3rd Platform workloads are crossing into the "mission critical" category. As a result, more of these technologies now require the same level of RAS, security, and performance that was previously reserved only for transaction-oriented workloads. More organizations are looking to deploy Big Data and cloud workloads on enterprise-class systems with robust RAS features but that still offer the features to scale out and support large VM environments such as multithreaded processors and high workload density. IBM Power Systems continue to improve these capabilities with features such as Enterprise Pools and the Power Integrated Facility for Linux (Power IFL). Additionally, 3rd Platform workloads stand to benefit from acceleration technologies such as GPUs, field-programmable gate arrays (FPGAs), and flash cache, which allow CPUs to securely and directly access memory as if it were a standard CPU core. Mobile-based transactions are also increasing rapidly and driving the need for infrastructure that can meet greater workload demands, including increased performance for data, scalability, security/privacy, and high degrees of availability (as systems must now be able to provide support to users 24 x 7).

IBM POWER8

With POWER8, the successor to POWER7 and POWER7+, IBM will now be increasingly focused on supporting Linux in addition to the installed base of IBM AIX Unix and IBM i on IBM POWER8 Systems. This will enable it to support more open source workloads, including OpenStack-based cloud workloads. IDC notes that many of the POWER8 capabilities — including scale, performance, RAS, security, and virtualization — are available in POWER7+. That said, POWER8 includes significant enhancements to these core features.

POWER8 Feature and Performance Improvements

Chief POWER8 features include the following:

- 12 cores per POWER8 chip, compared with the 8 cores used in POWER7 and POWER7+ (POWER8 is manufactured using a 22nm [nanometer] process.)
- Eight threads per core in place of the four threads per core in POWER7+
- Optimizations for greater VM density and for power efficiency
- Accelerators for memory expansion, encrypting/decrypting code; transactional memory; VMM assist and VM Mobility to improve and simplify the development, porting, and operation of mission-critical, demanding applications
- Larger caching structures, with 96MB in the L3 cache and 128MB in the L4 cache, for handling database and analytics workloads by staging larger chunks of data closer to the chip for processing
- Memory buffer chip, which improves end-to-end processing across POWER8; sustained memory bandwidth reaching 230GBps per socket compared with 86GBps on POWER7+ processors
- Coherent Accelerator Processor Interface (CAPI) includes an accelerator feature that allows PCIe devices, such as GPUs, FPGAs, and flash cache, to directly and securely address memory as if it were a standard CPU core.)
- Native support for the PCIe server/IO interface standards; support for PCIe Gen 3

POWER8 has faster single-thread processing than the earlier POWER chips, with processing up to 1.9 times what it was in POWER7. IDC expects that at maximum, the POWER8 core will deliver twice the overall processing power of POWER7. IDC also understands that POWER8 will improve VM densities and will be more efficient than earlier generations of POWER technology via new hardware virtualization assists.

POWER8 also provides improved support for IBM DB2 with BLU Acceleration, a new generation of in-memory data management technology, which typically delivers 8x to 25x improvements in analytic reporting and 10x storage space savings compared with DB2 10.x. Based on these new capabilities, organizations should be able to significantly decrease their server footprint for equivalent computing performance, reduce software licensing costs, and improve single thread performance and virtualization density.

Designed to Bring Performance and RAS to 3rd Platform Workloads

With POWER8, IBM appears to be pursuing a strategy of bridging the gap between the needs of traditional OLTP workloads, with high degrees of reliability, accessibility, and security, and the needs of the 3rd Platform, with support for virtualization, VM mobility, and CAPI acceleration. IBM has also emphasized the centrality of Big Data analytics workloads in its design with increased parallelism, higher memory, and I/O bandwidth. In addition to traditional scalable Unix workloads, POWER8 is also designed to provide an attractive and cost-effective alternative to x86, where the organization has a scale-out strategy, where data-centric application performance is the most important characteristic, or where small x86 clusters can be consolidated into a single 2-socket POWER8 box for efficiency.

As such, IBM is targeting POWER8 to provide enterprise-class characteristics to workloads that were once the domain of scale-out servers that offer less robust RAS and security features. At the same time, it is designed to remain a viable investment where you need maximum efficiency, maximum RAS, a platform to support a broad mix of more traditional Unix applications with a need for shared resources delivered via a scalable single system image, or large-scale consolidation.

Push for Linux and Open Standards

While Power Systems processors have supported Linux for some time, with POWER8, IBM is supporting a broad initiative aimed at expanding the number of Power Systems shipping with Linux as the primary operating system. Further, new POWER8 features instantiated in firmware will allow more code to run on Power Systems than before, which could expand IBM's footprint in Linux-based computing and help the company gain a greater foothold in cloud computing and Big Data Hadoop analytics. IBM plans to support the open source KVM hypervisor, in addition to its own PowerVM hypervisor, which IBM is using as another way both to compete more strongly against x86 where scale-out is the primary deployment model and to compete more strongly in the enterprise segments where POWER is the highest logical unit but scale-out has been on x86.

To tap this opportunity, IBM has announced that it is spending $1 billion over three years. That investment includes the opening of a number of Power Systems Linux Centers worldwide — including sites in Beijing; New York City; Austin, Texas; and Montpellier, France.

Building the POWER8 Ecosystem

IBM plans to leverage its Power Systems investments into a wider ecosystem encompassing Linux, IBM AIX, and IBM i — and including a broader portfolio of Linux and open source workloads. One example is the OpenPOWER Foundation for which IBM is working with Google, Tyan, NVIDIA, Mellanox, Research Institute of Jiangsu Industrial Technology, Samsung, and others to enable a variety of workloads on POWER8 technology.

The OpenPOWER Foundation will allow members to build server, networking, storage, and GPU-acceleration technology by leveraging IBM's POWER architecture. Much of this work will focus on hyperscale computing datacenters, such as those used by cloud service providers (CSPs). Open source firmware will be made available to OpenPOWER members.

In addition, in February 2014, IBM launched a Linux on Power development cloud. Aimed at programmers and developers, this is a free cloud service provided and supported by IBM. It will support building, porting, and testing of Linux workloads on Power Systems and will also allow programmers to test workloads for IBM AIX and IBM i. In its announcements, IBM has noted that 400 ISVs, with more than 1,000 Linux applications, are certified for use on Power Systems servers. IBM will be targeting a range of applications, including the Oracle and SAP application ecosystems, to run on Power Systems.

Importantly, through the OpenPOWER Foundation, organizations will be allowed to license POWER intellectual property (IP) from IBM for use in systems targeting a range of new workloads for mobility and cloud computing. The first products from alliance partners leveraging OpenPOWER will use IBM's POWER8 technology, which is expected to ship in 2014.

OPPORTUNITIES AND CHALLENGES

IBM has a long history of investment in strategic ecosystem enablement for its technologies, and the Power Systems ecosystem is also benefiting from these investments in two important areas. First, IBM announced a $1 billion investment in enabling additional Linux on Power by working closely with independent software developers around the world to increase the number of applications optimized to run on Power. Second, IBM announced the creation of the OpenPOWER Foundation in August 2013. The members of the consortium, including Google, NVIDIA, and Mellanox, have committed to work together and build server, storage, networking, and GPU acceleration technologies that leverage IBM's POWER microprocessor technology. This consortium will be able to license POWER8 intellectual property from IBM for use in systems targeting a number of rapidly emerging 3rd Platform workloads spanning mobile, cloud, and Big Data analytics.

IBM has invested billions of dollars on the development of the core POWER micro-architecture, associated software, and silicon fabrication capabilities. IDC sees a number of opportunities and challenges for IBM as it rolls out its new POWER8 offerings. Opportunities include:

- **Refresh opportunities within the IBM Power Systems installed base.** With its improved performance and scalability, POWER8 presents an opportunity for organizations already running performance-intensive, mission-critical applications on the Power Systems platform to extend the life of those applications. New performance capabilities with POWER8 also enable smaller systems to run more diverse workloads, enabling smaller businesses to leverage Power for 3rd Platform workloads more efficiently.
- **Platform migration for demanding workloads.** IDC research shows stabilization in the pace of migration of mission-critical, transaction-oriented workloads away from traditional Unix servers. IBM's Power Systems remain an attractive landing place for Unix workloads running on competitive platforms from HP and Oracle. IBM's commitment to the platform, coupled with the ecosystem development investments noted previously, helps ensure that Power Systems remain optimized for a wide range of tier 1 mission-critical workloads. For example, Power Systems servers' multithreaded cores and RAS features could support database, transaction, and Big Data analytic workloads as they move onto a highly virtualized and managed platform.
- **Support for cloud computing.** Whether for running an internal cloud or for cloud service providers, POWER8 support for virtualization at the chip level enables service organizations to spin up multiple virtual machines quickly while reducing the performance overhead associated with typical virtual machine instances.
- **Enabling reduced IT operating costs through server consolidation.** Housing both the application and the data layer on a single POWER8 system enables reduction in the hardware footprint, enabling enterprises to achieve higher IT staff productivity, increased resource utilization (CPU, memory, and network), reduced software licensing costs, reduced energy consumption, reduced IT infrastructure costs, and extended datacenter life. IDC's Business Value research shows that consolidating workloads onto fewer hardware systems can reduce IT operational costs for IT staff by 50%+, power/cooling up to 20%+, and use of datacenter real estate by 30%+, freeing up spending for more strategic investments. While consolidation is attractive on POWER7+ systems, new capabilities on POWER8 extend these benefits to a broader set of customers and workloads.

Challenges include:

- **Ability of IBM to expand a complex ecosystem.** Systems of record and systems of engagement require deep levels of investment and integration. To have viability, a healthy ecosystem requires commitment from other partners — ISVs, systems integrators, and VARs — to support and extend their own investments deeper into the environment. IBM has been investing heavily in its POWER architecture and has made a number of investments intended to expand the Power Systems ecosystem, which are noted previously. However, IBM will face a challenge to build the critical mass necessary to fully realize the expansion of the Power Systems ecosystem among other partners.
- **Investments in open standards may not generate intended returns.** IDC believes IBM is making a number of critically important investments intended to expand the POWER architecture deeper into open computing environments. These investments include the creation of the OpenPOWER Foundation as well as significant expenditures designed to expand the Linux on Power ecosystem. IDC notes that these investments could face headwinds in the market primarily from the x86 ecosystem, which intends to continue to move upmarket and deliver richer enterprise services into the Linux and Unix server market.

IDC OPINION

IBM and Oracle/Fujitsu are the remaining server competitors investing in RISC processors — with their POWER and SPARC processors, respectively. Both POWER and SPARC run Unix operating systems, which is what makes their continued head-to-head competition in RISC processors so important for the Unix server space. This competition in the RISC server world will provide customer choice and will continue to spur new features and capabilities. However, it is important to note that RISC processors such as POWER are well suited for workloads running on Linux and IBM i. The Linux ecosystems continue to represent a catalyst for market growth, and IBM has made significant investments aimed at driving additional Linux workload onto Power Systems. IDC also notes that ARM processors — a form of RISC — are enabling most of the smartphone and tablet deployments in the world. The rapid expansion of this ARM mobile ecosystem should drive the need for more datacenter transactional capacity, further benefiting Linux and the server architectures that are best optimized to support this workload. This could place additional challenges on the x86 ecosystem going forward.

IBM's charter for Power Systems includes delivering innovations that matter, as witnessed by IBM's recent announcement of real-time analytics and Watson natural language learning on Power Systems. IBM clearly remains committed to Power Systems, and with POWER8, it is doubling down on that commitment. Its $1 billion investment to build an open platform-based ecosystem is a highly credible signal that IBM takes Power Systems very seriously and expects the architecture to be around for technology generations to come.

Power Systems fit squarely into IBM's commitment to delivering innovative solutions to the market, and with its increasing investment in a Linux open platform-based ecosystem, IBM clearly realizes that the success of the platform hinges not only on building the underlying core technologies but also very much on shaping the use cases and ecosystems around Power Systems servers. With these investments, IBM is looking to improve the choices for its customers and better support members of the OpenPOWER Foundation that are looking to deploy hyperscale Linux-based cloud services.

IDC expects that with the coming release of POWER8, IBM will continue to deliver on Power Systems core strengths of providing a highly reliable, scalable, available platform to house mission-critical business applications such as ERP and OLTP, but with new innovations including improved multithreading, greater VM density support, and CAPI acceleration, IBM is now extending Power Systems to encompass the scalability and multitenancy needs of the 3rd Platform — in particular Big Data analytics and cloud. With these investments and advances in its latest generation of POWER technology, IBM clearly intends to continue its positioning as a "safe" technology bet for datacenter infrastructure well into the era of the 3rd Platform.

CONCLUSION

The dawn of the 3rd Platform has brought a number of new challenges to IT organizations across a wide variety of industries — not least of which is putting in place infrastructure to support it. Organizations are moving rapidly to cloud enable their enterprise applications and support Big Data initiatives and require the right platforms for them. And organizations that have traditionally used enterprise servers leveraging RAS features for mission-critical applications such as OLTP must now figure out how to marry those requirements with cloud deployments, especially as 3rd Platform technologies become mission critical and increasingly require the same RAS features — along with the scale-out and virtualization features required for the 3rd Platform.

IDC expects that with the upcoming POWER8 release, IBM will address many of these challenges. While building on its enterprise legacy of providing RAS features for mission-critical OLTP and ERP applications, POWER8 also contains innovations to better support virtualization, multitenancy, and cloud necessary for the 3rd Platform. IDC expects POWER8 to continue to be a viable choice for transaction-oriented, mission-critical deployments as well as Big Data, cloud, and other 3rd Platform technologies.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-insights-community.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data – Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2014 IDC. Reproduction without written permission is completely forbidden.

POL03191-USEN-02