HPC and HPDA for the Cognitive Journey with OpenPOWER

Sponsored by IBM

Srini Chari, Ph.D., MBA

May, 2016

mailto:chari@cabotpartners.com

Executive Summary

Big Data has become characteristic of every computing workload. From its origins in research computing to use in modern commercial applications spanning across industries, data is the new basis of competitive value. The convergence of High Performance Computing (HPC), Big Data Analytics, and High Performance Data Analytics (HPDA) is the next game-changing business opportunity. It is the engine driving a Cognitive organization with Data as its fuel.

But the volume, velocity and variety of data are creating barriers to performance and scaling in almost every industry. To meet this challenge, organizations must deploy a cost-effective, high-performance, reliable and agile infrastructure to deliver the best possible business and research outcomes. This is the goal of IBM's data-centric design of Power Systems and solutions from the OpenPOWER Foundation for HPC and HPDA.

Businesses are investing in HPDA to improve customer experience and loyalty, discover new revenue opportunities, detect fraud and security breaches, optimize research and development, mitigate financial risks, and more. HPDA also helps governments respond faster to emergencies, improve security threat analysis, and more accurately predict the weather – all of which are vital for national security, public safety and the environment. The economic and social value of HPDA is immense. It is also integral to the journey towards a Cognitive and Learning business—a business that utilizes hardware and software designed to learn from its own information, continuously evolve, and return the most insightful, actionable results.

A key underlying belief driving the OpenPOWER Foundation is that focusing solely on microprocessors is insufficient to help organizations overcome performance barriers. System stack (processors, memory, storage, networking, file systems, systems management, application development environments, accelerators, workload optimization, etc.) innovations are required to improve performance and cost/performance. IBM's data-centric design minimizes data motion, enables compute capabilities across the system stack, provides a modular, scalable architecture and is optimized for HPC and HPDA.

Real world examples of innovations and performance enhancements resulting from IBM's data-centric design of Power Systems and the OpenPOWER Foundation are discussed here. These span financial services, life sciences, oil and gas and other HPC/HPDA workloads. These examples highlight the need for clients (and the industry) to evaluate HPC systems performance at the solution/workflow level rather than based on narrow synthetic point benchmarks such as LINPACK that have long dominated the industry's discussion.

Clients who invest in IBM Power Systems and high-value offerings from the OpenPOWER Foundation could lower their total cost of ownership (TCO) with fewer, more reliable servers compared to alternatives. More importantly, these customers can accelerate performance and time to insight in their journey to become a Cognitive business.

Systems Innovation Needed to Extract Value from Data Deluge

The relentless rate and pace of technology-enabled business transformation and innovation are astounding. Several intertwined technology trends in Social, Mobile and the Internet of Things (IoT) are making data volumes grow exponentially. In 2018, about 4.3 exabytes (10^{18} bytes) of data is expected to be created daily – over 90% will be unstructured.[1]

In many industries, High Performance Data Analytics (HPDA) use cases are rapidly emerging – galvanizing growth in the High Performance Computing (HPC) market which is expected to reach $31 billion by 2019[2]. The lines between HPC and Big Data Analytics are blurring as HPDA continues to grow 3 to 4 times the rate of traditional HPC.[3]

These HPDA use cases combine Systems of Records (structured data) with Systems of Engagement (unstructured data – images, videos, text, emails, social, sensors, etc.) to produce new *High Value Systems of Insights* (Figure 1) to spur organizational learning and innovation – on the road to a Cognitive business.

Figure 1: High Value Insights from Integration and Analysis of Structured and Unstructured Data

But extracting insights from the growing data is very challenging. The journey towards a Cognitive and Learning organization requires investments in high-performance systems and solutions: servers, storage, networking, accelerators, software, etc. As these systems deliver greater performance and capabilities, they can progressively turn data to predictive, actionable intelligence and the best outcomes, maximizing business value.

Many traditional HPC applications (seismic analysis, life sciences, financial services, climate modeling, design optimization, etc.) are becoming more data-intensive with higher fidelity models and more interdisciplinary analyses. These applications are also increasingly leveraging unstructured data to glean deeper insights. There is also a need for near-real-time analytics in many industries, such as telecommunications, retail and financial services.

Newer HPDA applications are also being used for cyber-security, fraud detection, social analytics, emergency response, national security, and more. Many high value use cases automating iterative reasoning and Machine Learning continue to emerge rapidly. Deep Learning (Unsupervised Machine Learning leveraging HPDA) and Cognitive Computing are rapidly growing applications that can significantly benefit from HPC infrastructure.

---

[1] https://storageservers.wordpress.com/2016/02/06/how-much-data-is-created-daily/
[2] http://www.idc.com/getdoc.jsp?containerId=259211
[3] [3] Earl Joseph, et. al., "IDC's Top Ten HPC Market Predictions for 2015, January, 2015

But this requires a flexible and modular architecture that minimizes costs and enables innovations to accelerate computing at all levels of the systems hierarchy. The OpenPOWER Foundation and its many (and growing) collaborating members continue to innovate and provide a range of solutions to accelerate performance and reduce costs across the entire iterative workflow: from data to analytics to learning to best business outcomes.

Cognitive Computing: From Data to Analytics to Learning

The majority of data doesn't offer much value unless iteratively and progressively analyzed by the user and the system to produce powerful insights with recommended actions for the best outcome(s). In fact, IBM Watson (IBM's leadership Cognitive system) constantly sifts through data, discovers insights, learns and determines the best course of action(s).
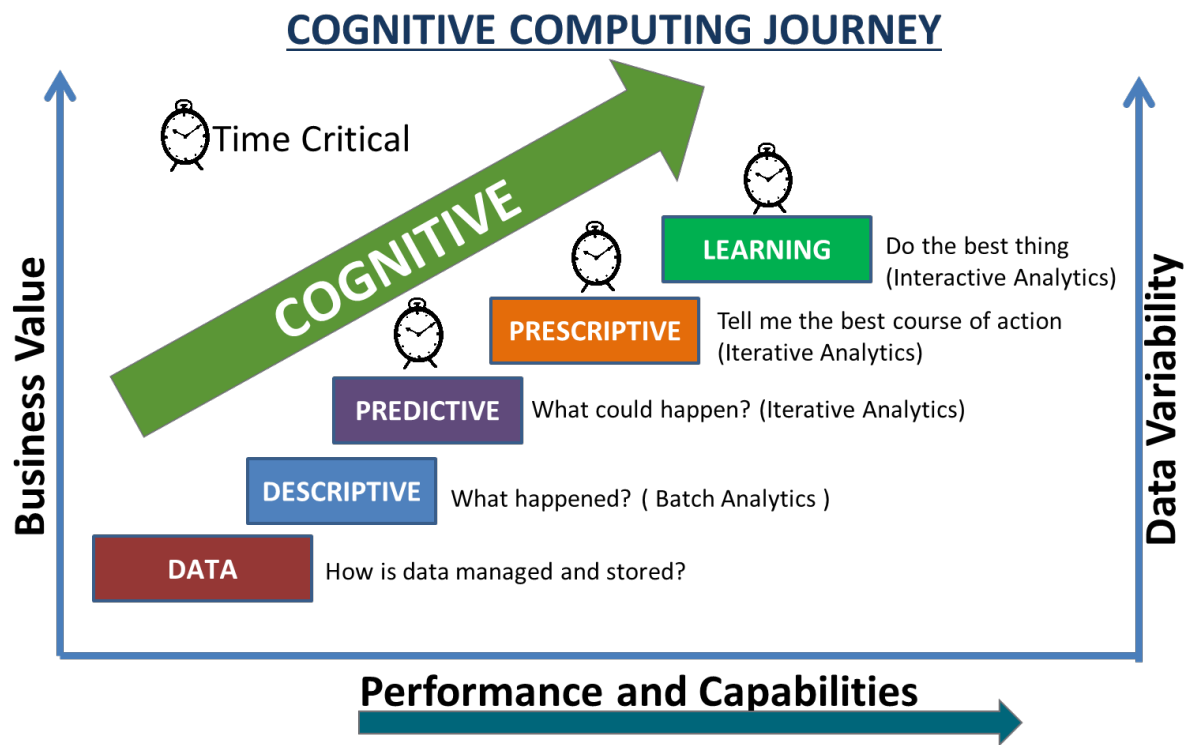
COGNITIVE COMPUTING JOURNEY

Time Critical

COGNITIVE

LEARNING — Do the best thing (Interactive Analytics)

PRESCRIPTIVE — Tell me the best course of action (Iterative Analytics)

PREDICTIVE — What could happen? (Iterative Analytics)

DESCRIPTIVE — What happened? (Batch Analytics)

DATA — How is data managed and stored?

Business Value

Data Variability

Performance and Capabilities

Figure 2: The Cognitive Computing Landscape with Increasing Business Value

The Cognitive Computing landscape continues to evolve rapidly; giving clients unique capabilities to progressively solve complex problems for higher value. Figure 2 depicts this landscape starting with data at the foundation, and then by analytics type, business value, data variability and time-criticality:

Data is fundamental in any Analytics initiative. A data warehouse is typically built to capture, store, secure, retrieve and manage the raw and processed data. Today, data warehousing is widely used by clients and traditional implementations are usually routine activities. Modern warehouses with newer technologies such as Hadoop and Apache Spark are being implemented rapidly. But unless data is converted to insights, there is little reward.

3

Descriptive batch analytics is dominant today and with low to medium reward. It condenses data into nuggets of insights summarizing what happened. Social media analytics is one prominent descriptive analytics example.

Predictive iterative analytics uses a combination of several statistical, modeling, data mining, and machine learning techniques to analyze data to make probabilistic time-critical forecasts about the future. Weather prediction and customer sentiment analysis are some noteworthy predictive analytics examples.

Prescriptive iterative analytics goes beyond descriptive and predictive analytics. It recommends one or more courses of action and the likely outcome of each action, including the usually time-critical "next best action". Predictive customer intelligence is a key example.

Learning (Cognitive and Deep Machine Learning) interactive analytics systems that continuously build knowledge over time by processing natural language and data. These systems learn a domain by experience just as humans do and can discover and suggest the "best course of action"; providing highly time-critical valuable guidance to humans or just executing this "next best action". IBM Watson is the premier cognitive system in the market.

The underlying technologies for Deep Learning include Artificial Neural Networks (ANN) – neural networks inspired by and designed to mimic the function of the cortex, the thinking matter of the brain. Driverless autonomous cars, robotics and personalized medical therapies are some key disruptive innovations enabled by Deep Learning.

A performance-optimized infrastructure is critical for the Cognitive Computing journey.

Must Optimize Performance/Value Across All Analytics Workflows

To maximize value from infrastructure investments, clients must evaluate the performance and costs of HPC systems holistically. Evaluations based on a narrow focus on point benchmarks such as High Performance LINPACK (HPL)[4], which primarily depend on processor Flops (floating point operations per second)/core and number of cores, are inadequate for many real life workloads, particularly in Analytics and Learning.

In addition to processor performance, workflow performance also depends on other system attributes such as memory, networks and storage – larger data sets make this dependency greater. Clients who rely solely on point benchmarks to make purchasing decisions are at risk of deploying an ineffective HPC environment for their specific workflows. In addition, they may not have the necessary systems infrastructure for the HPDA, Machine Learning and Deep Learning (Unsupervised Learning from Big Data) workloads of the future.

Prospective clients must use a framework of inter-related drivers and associated metrics that examine the total costs incurred and the value delivered for *their-specific workflows*. Figure 3 depicts a typical iterative data-centric seismic survey workflow in the Oil and Gas industry.

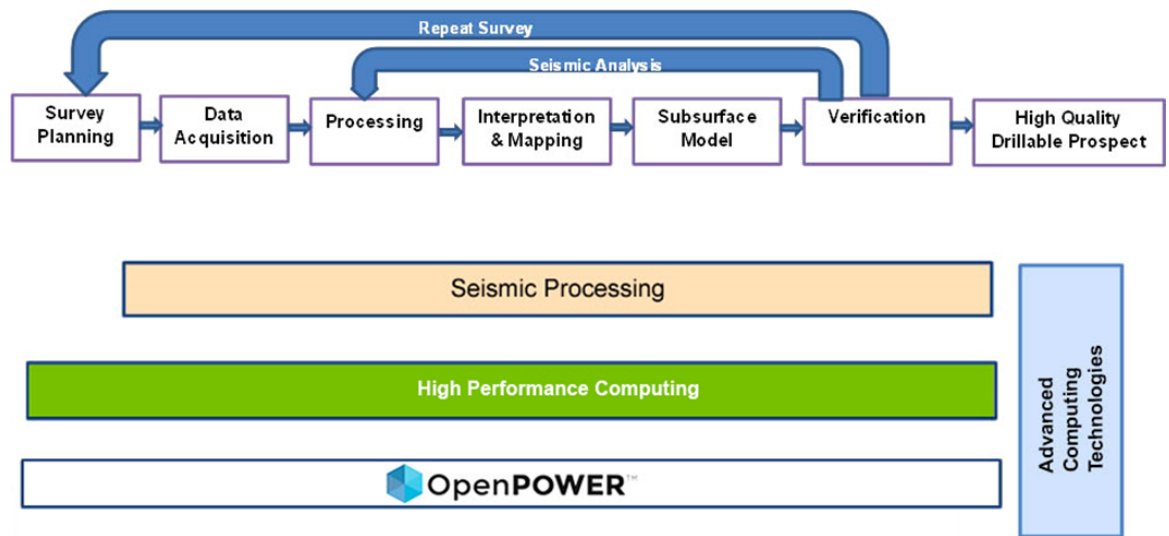[4] http://www.netlib.org/benchmark/hpl/

Figure 3: Iterative Seismic Survey Workflow with Embedded High Performance Data Analytics

Some key value and cost drivers that should be carefully considered include:

Value Delivered

- Business Value: e.g. customer revenues, new business models, compliance regulations, better products, increased business insight, faster time to market, and new breakthrough capability
- Operational Value: e.g. faster time to results, more accurate analyses, more users supported, improved user productivity, better capacity planning
- IT Value: e.g. improved system utilization, manageability, administration, and provisioning, scalability, reduced downtime, access to robust proven technology and expertise.

Costs Incurred

- IT /Data Center Capital e.g. new servers, storage, networks, power distribution units, chillers, etc.
- Data Center Facilities e.g. land, buildings, containers, etc.
- Operational Costs: e.g. labor, energy, maintenance, software license, etc.
- Other Costs: e.g. system management, deployment and training, downtime, migration, etc.

Beyond specific workflow and cost-benefit analyses, clients must invest in Cognitive systems to further maximize business value. These systems are designed to deliver outstanding performance and handle the most time-critical Deep Learning applications.

System Attributes Impacting HPC Application Performance

In the last several decades, application developers, academic institutions, government laboratories and other HPC solution providers have made substantial investments to optimize and tune HPC applications to take advantage of clusters and parallel systems. Substantial performance improvements have been achieved by careful load balancing, maximizing single socket performance, maintaining data locality, minimizing cache misses, improving I/O performance, and maximizing the computation-to-communication ratio.

End-to-end HPC workflow performance is further enhanced with scalable networking of servers and storage, accelerators - Graphics Processing Units (GPUs) and field-programmable gate arrays (FPGAs), workload and cluster management software and parallel file systems.

Figure 4 depicts performance characteristics for a range of HPC applications mapped to seven key cluster system features that influence performance: Flops/core, number of cores, node memory capacity, memory bandwidth at each node, I/O performance, interconnect latency and interconnect bandwidth.

Four major application categories (Figure 4) representing today's mainstream use of HPC are considered: Computer Aided Engineering (CAE), Life Sciences, Financial Services, and Energy and Environmental Sciences. Each category is further divided into several segments. Within each segment, the relative contributions (normalized to one) of system features that typically impact performance are indicated by the colored polygonal lines.

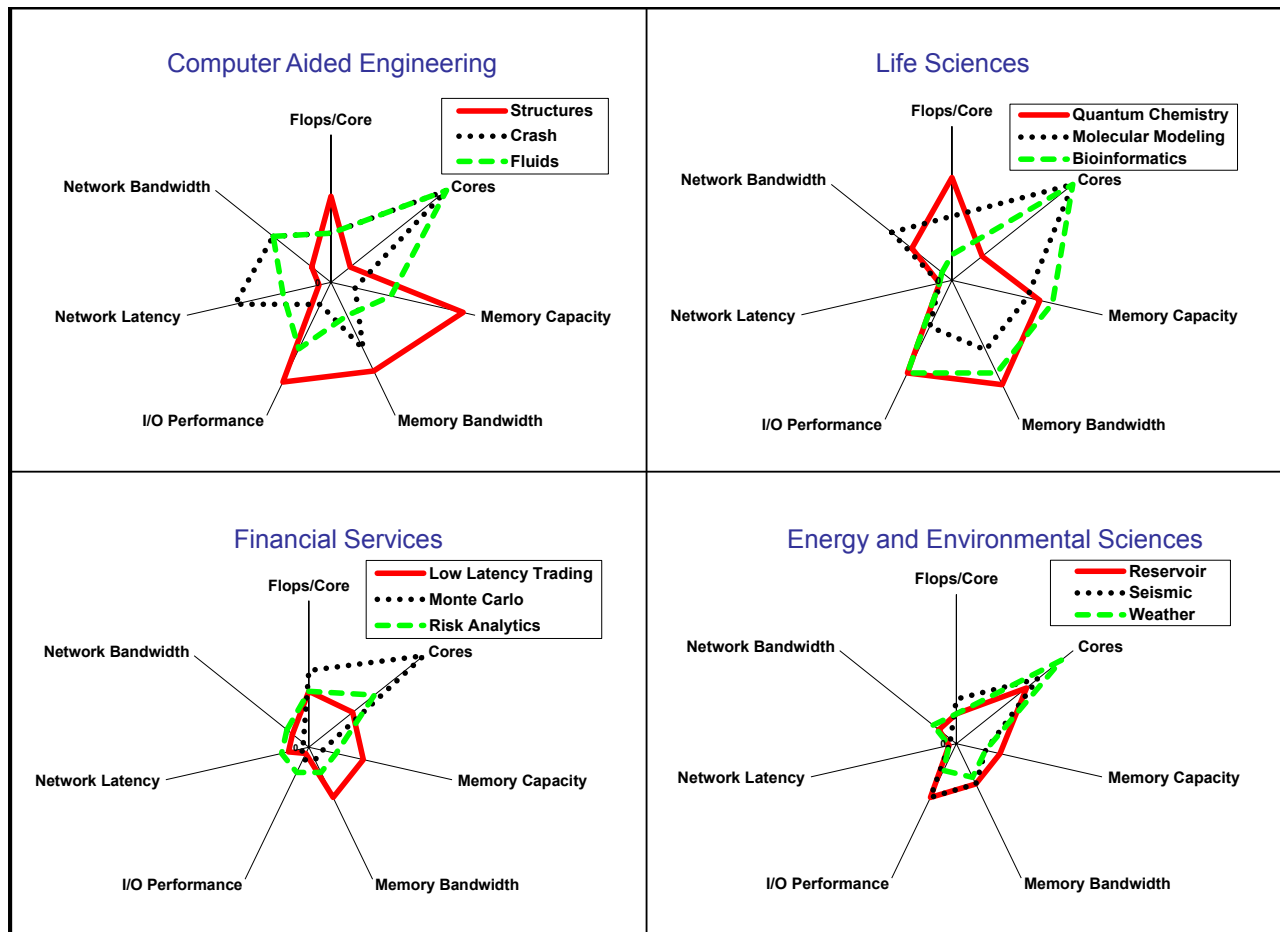**Figure 4: Application Performance Characteristics Mapped to Seven Key Cluster System Features**

For our high-level conceptual workload analysis (Figure 4), we have used actual performance data obtained from several published and observed benchmarks[5] together with other expert

---

[5] Swamy Kandadai and Xinghong He, "Performance of HPC applications over InfiniBand, 10 Gigabit and 1 Gigabit Ethernet", 2010

analyses[6] for each application within each category. In addition, we have interviewed many application experts who have an intimate knowledge of the underlying applications.

Figure 4 summarizes this analysis for typical HPC workloads that are common for each application category assuming these workloads are run on clusters with the widely available components for processors, storage, and networks.

It is clear that the performance of most practical HPC applications also depend on memory, I/O and network and not exclusively on Flops/core and the number of cores. Yet, for over two decades, the High Performance LINPACK (HPL) benchmark is the most frequently metric used to evaluate the performance of HPC systems. HPL is even less appropriate for Analytics.

System Attributes Impacting Analytics and Learning Applications

Analytics workloads types span a wide range with topologically variable data access patterns:

- Descriptive analytics are typically **batch** workloads with predictable and topologically regular (but not necessarily uniform) data access patterns.
- Predictive and Prescriptive analytics are typically **iterative** analytics with a greater number of unpredictable and regular/irregular data access.
- Learning analytics are typically iterative or **interactive** with almost continuous regular/irregular data access.

Analytics infrastructure demands and time-criticality needs also grow significantly from Descriptive to Predictive to Prescriptive and Learning. Further, if the Analytics involves streaming data, the time-criticality needs are even greater. Highly variable unstructured data access (typical in Learning) is extremely challenging for many "commodity" systems.

Why Analytics on Unstructured Data is Challenging: A common way to represent the growing unstructured data stream is through a graph: vertices denote users or events, edges their relationships. For example, the number of vertices (Facebook users) could be billions but the edges (friends) could be only hundreds. So the connectivity matrix is typically very sparse and unstructured.
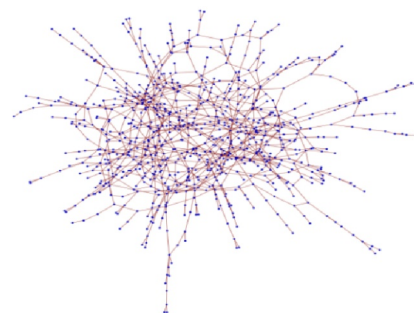
Analytics (e.g. social media analytics) using algorithms that manipulate these sparse matrices to glean insights, have irregular and fine-grained memory access patterns, limited cache line locality and reuse, and frequent misses for very large data sets. The performance of these sparse matrix algorithms is limited by the costs of moving data across the memory hierarchies (processors to cache levels to on-processor memory to remote memory (via the network) to storage through controllers, and so on). Many recent Analytics applications involve similar sparse matrix operations.

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

---

[6] Wayne Joubert, Douglas Kothe, and Hai Ah Nam, "Preparing for Exascale: ORNL Leadership Computing Facility; Application Requirements and Strategy", December 2009.

Figure 5 depicts performance characteristics for Analytics (Descriptive, Predictive, Prescriptive and Deep Learning) mapped to the same seven key cluster system features that influence performance: Flops/core, number of cores, node memory capacity, memory bandwidth at each node, I/O performance, interconnect latency and interconnect bandwidth.

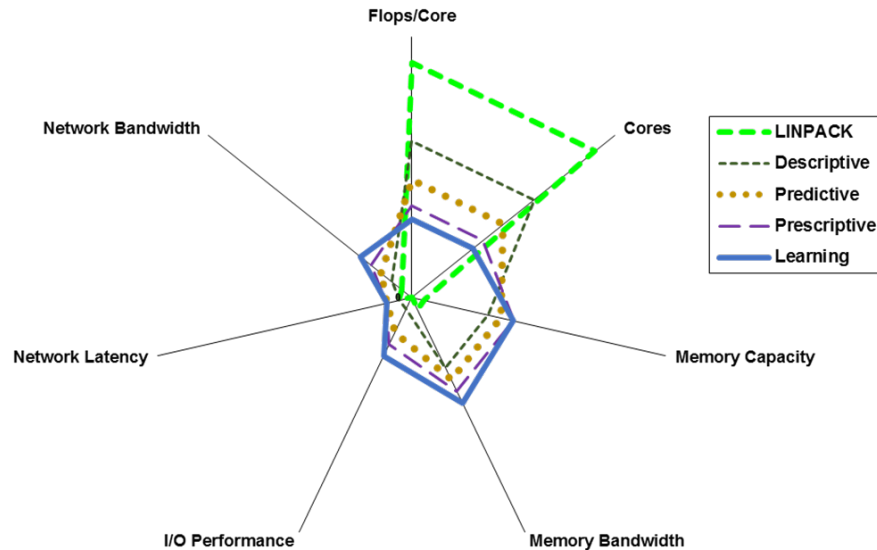## High Performance Data Analytics vs. LINPACK

Figure 5: HPDA Performance Characteristics Mapped to Seven Key Cluster System Features

For each Analytics type, the relative contributions (normalized to one) of system features that typically impact performance are indicated by the colored polygonal lines. The High Performance LINPACK (HPL) benchmark is also superimposed on Figure 5.

The well-tuned HPL benchmark measures the performance of systems to solve a dense matrix system of equations. Here, the memory access is very structured with significant cache-data reuse and negligible cache misses. So data movement costs are negligible. HPL performance is primarily determined by the processor performance (number of cores and Flops/core).

However, Analytics and Cognitive Computing require more balanced data-centric HPC systems. The emphasis on the capabilities of memory, network and I/O performance relative to processor performance increases from Descriptive to Predictive to Prescriptive to Learning. The OpenPOWER Foundation's goal is to provide this balanced and flexible system to ensure clients can seamlessly transform to a Cognitive business.

OpenPOWER: Re-imagining System Design for HPC and Cognitive

As data volumes grow exponentially, the costs of moving the data in and out of a central processor becomes prohibitive. To move 1 byte from storage to the central processor, it could cost 3-10 times the cost of one floating point operation (flop).[7] So why not move data less by running workloads where the data resides? This requires "computing" at all levels of the

---

[7] https://www.nersc.gov/assets/NERSC-Staff-Publications/2010/ShalfVecpar2010.pdf

system stack including network, memory and storage (Figure 5 – left side). This is the core of the OpenPOWER Foundation's approach with the following key architectural principles:

1. **Build modularity** with a balanced, composable architecture that is scalable from a sub-rack to hundreds of racks.

2. **Enable compute in all levels of the systems hierarchy** with "active" system elements throughout the stack including network, memory, storage, etc.

3. **Minimize data motion** by providing hardware and software to support and enable compute in data and schedule workloads to run where they run best.

4. **Optimize for HPC and Cognitive** by using real workloads/workflows to drive design points optimized for client business value.

5. **Accelerate innovation** and provide clients flexibility and choice to deploy well-integrated, best-of-breed HPC and HPDA software solution components.
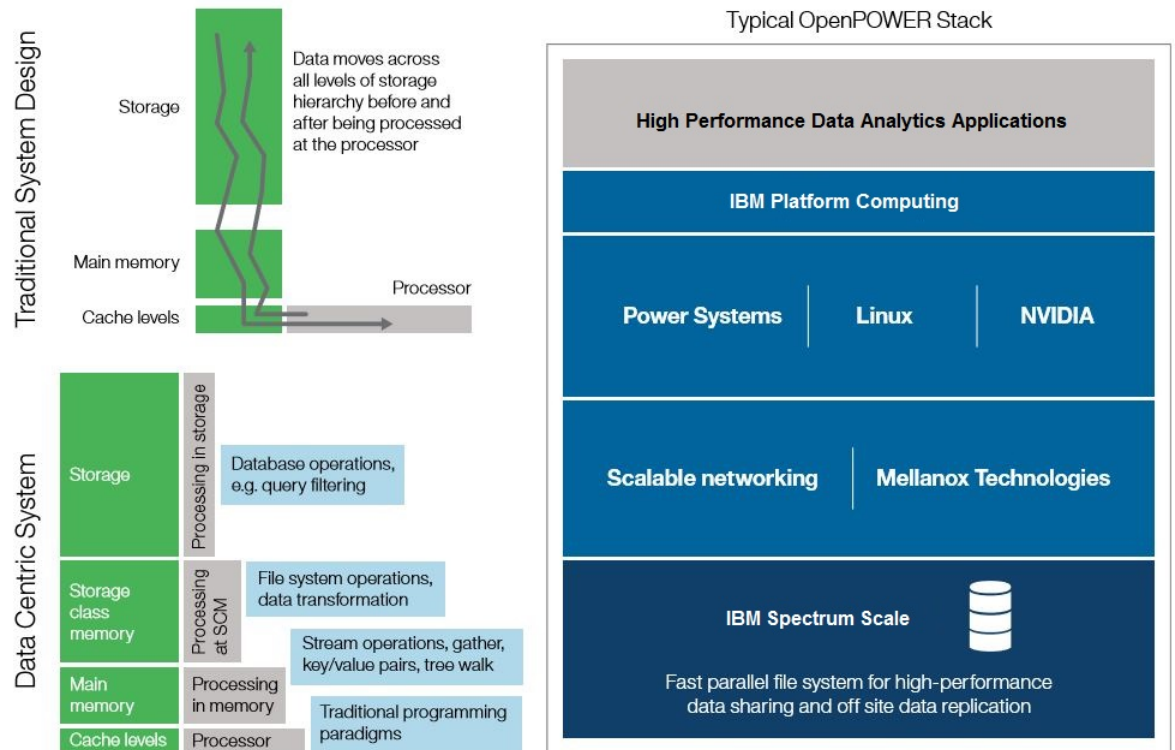
Figure 6: Traditional and Data Centric System Design   Typical OpenPOWER HPC/HPDA Stack

In addition, IBM offers a wide array of High Performance Computing solutions (Figure 6 – right side) including high-performance systems, clusters, software and HPC cloud services. Featured systems include: IBM Power Systems and IBM System Storage on Linux and IBM AIX. Key software includes powerful and intuitive workload and resource management software from IBM Spectrum Computing, a high-performance shared-disk clustered file system – IBM Spectrum Scale and optimized scientific and engineering libraries.

Additional innovative offerings include IBM InfoSphere BigInsights – a comprehensive, enterprise-grade full-featured Hadoop platform for Analytics; middleware and business

9

partner applications and service providers with deep proven expertise in HPC. In addition, IBM (particularly IBM Research) has a worldwide technical staff of domain experts to collaborate with clients to migrate and optimize their applications on IBM systems and software to solve their largest and most challenging HPC and HPDA problems.

Across many industries, clients are deploying OpenPOWER based HPC and HPDA solutions to accelerate their entire workflow and gain unprecedented levels of insights and learning.

Performance Examples Highlighting Advantage of OpenPOWER

IBM and OpenPOWER members are working extensively on a range of performance benchmarks studies (see Appendix for details on specific configurations) including on specific industry applications.

Industry Standard Benchmarks: Figure 7 covers some results of standard benchmarks that are good indicators of data-centric HPC performance.
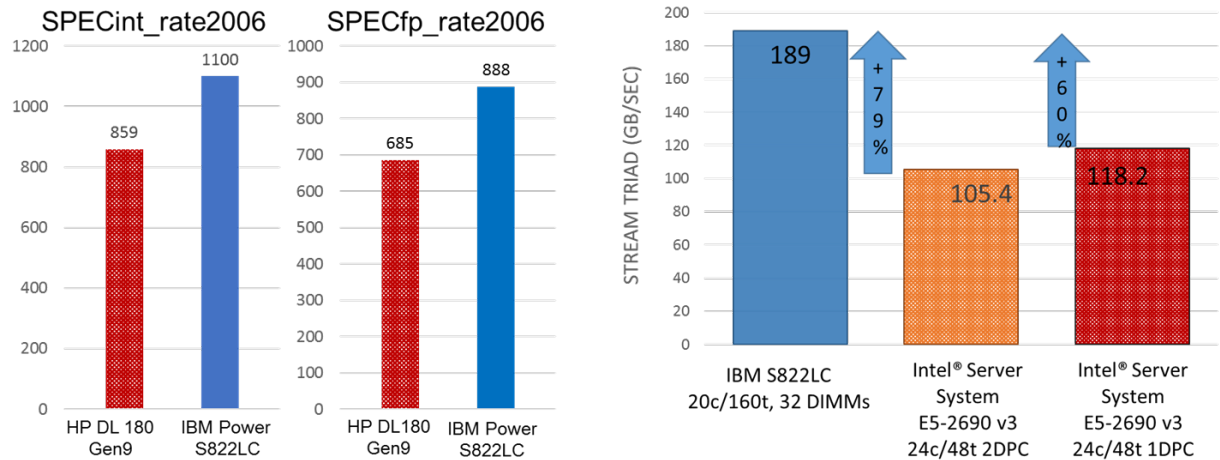
Figure 7: Better Performance on SPECint_rate2006, SPECfp_rate2006 and STREAM TRIAD

SPECint_rate2006 measures integer performance which is critical for Bioinformatics and Graph Analytics. SPECfp_rate2006 is a balanced average over many floating point intensive HPC applications across several industries. The IBM Power S822LC is over 1.28X faster than the HP DL 180 Gen9 system.

STREAM TRIAD accurately measures memory bandwidth. The IBM Power S822LC delivers 60-79% greater memory bandwidth than lightly populated Intel systems.

HPC/HPDA Application Benchmarks: A wide range of HPC/HPDA application benchmarks are summarized. These range from Computational Fluid Dynamics (CFD), Molecular Dynamics (MD), Genomics Workflows and Financial Services.

Figure 8: Better Performance on Computational Fluid Dynamics and Molecular Dynamics

OpenFOAM is a free, open source Computational Fluid Dynamics (CFD) application with a large global user base. The IBM Power System S822LC d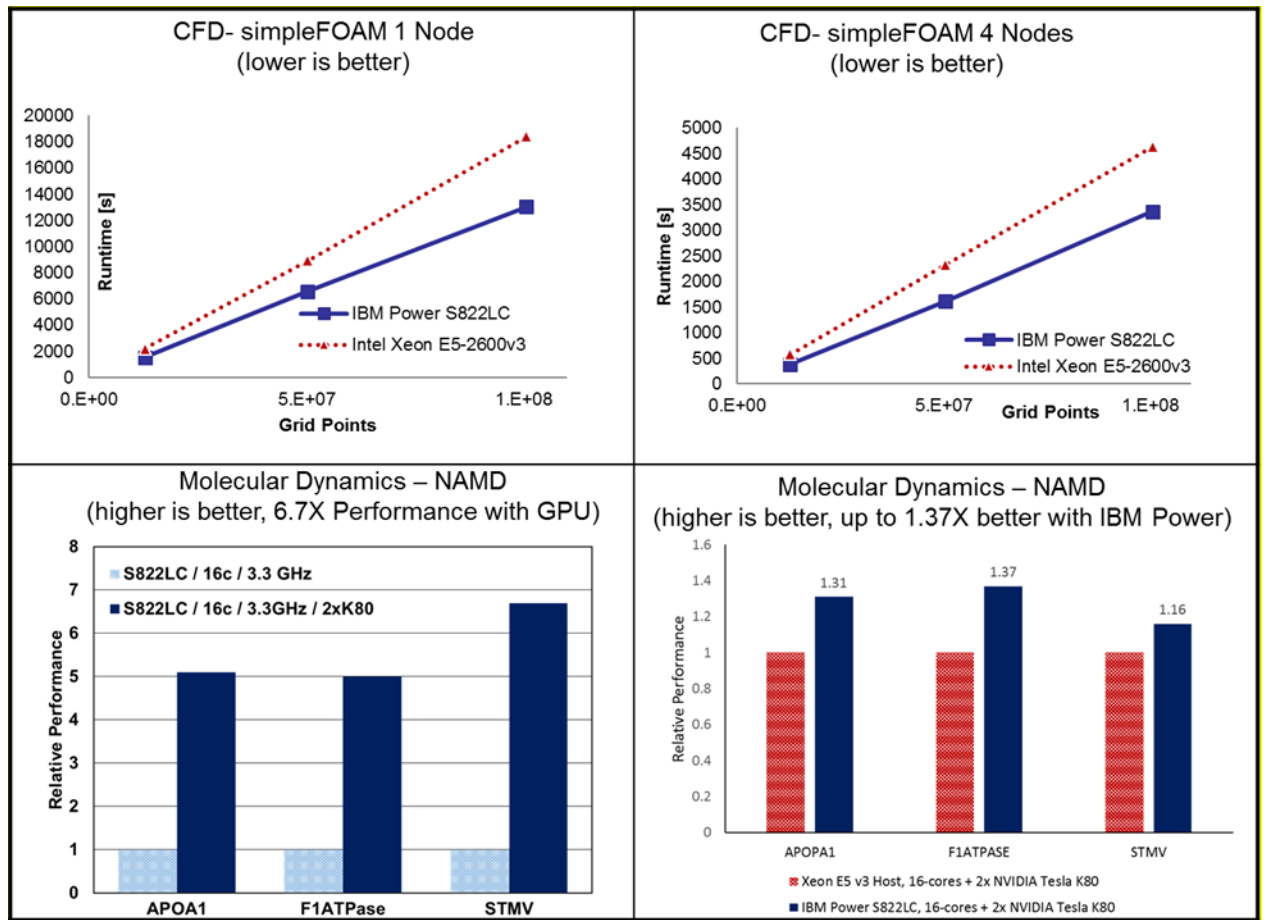elivers excellent throughput for the largest meshes with 40% faster results compared to the Intel Xeon E5-2600v3 System. The IBM Power System's higher memory bandwidth, increased L3 cache and more simultaneous multi-threading were key to this increased performance (Figure 8)

NAMD is a prominent Molecular Dynamics application that leverages single thread performance, highlighting core performance. Adding 2 NVIDIA Tesla K80 GPUs to the IBM Power S822LC delivers up to 6.7X better performance, and it outperforms the Intel Xeon E5-2600 v3 with NVIDIA Tesla K80s by up to 37% (Figure 8).

tranSMART (www.transmartfoundation.org/) is an open-source data warehouse and knowledge management system for integrating, accessing, analyzing, and sharing clinical, genomic, and gene expression data on large patient populations. The platform has already been widely adopted by the pharmaceutical industry and large-scale public-private initiatives for Genomic Medicine research. tranSMART v1.2 data loading and analytics loading performance was evaluated on the IBM POWER8 / Elastic Storage Server (ESS) System.

Loading Data into PostgreSQ: Test dataset 'TGCA_OV' containing clinical and genomic information on ovarian cancer patients were obtained from The Cancer Genome Atlas: http://tcga-data.nci.nih.gov/docs/publications/ov_2011

11

*transSMART
Loading and
Analytics in
Minutes
instead of
hours with IBM
Power and ESS*

Analytics: The 'TGCA_OV' dataset was also used to test the performance of the following built-in analytics functions within tranSMART:

- Marker-Assisted Selection (MAS)
- Hierarchical Cluster Analysis (HCA)
- Principal Component Analysis (PCA)

Complete loading and analytics were achieved in minutes as opposed to hours (Figure 9).

Loading 'TCGA_OV' into PostgreSQL

Elapsed time, in seconds
5,789,362 records

Data Type	File Size	Time
Clinical	Up to 0.5 MB	13s
Gene Expression	Up to 32 MB	227s (3min 47s)

tranSMART Analytics using 'TCGA_OV'

Elapsed time by Analysis Type, in seconds, for single node

Type	Acquire Data	Run Analysis	Total
MAS	157s	55s	212s (3min 32s)
HCA	262s	83s	345s (5min 45s)
PCA	253s	558s	811s (13min 31s)

Figure 9: tranSMART Complete Loading and Analytics in Minutes instead of Hours

The Genome Analysis Toolkit (GATK) is Broad Institute's software package for analyzing high-throughput sequencing data. These wide variety of tools focus on variant discovery and genotyping as well as on data quality assurance. On the IBM Power S822LC with ESS, a complete 65X coverage of whole human genome using Broad's best practice pipeline on single node took about 20 hours – more than 2X faster than the best published Intel result.

*Complete 65X
coverage of
whole human
genome
pipeline on
single IBM
Power node
took only 20
hours*

Short Oligonucleotide Analysis Package (SOAP) provides a full solution to Next Generation Sequencing (NGS) data analysis. SOAP3-dp (part of SOAP) leverages both CPU and GPU with optimized algorithms, delivering high speed and sensitivity simultaneously. The IBM Power S822LC with 2 NVIDIA K80 GPUs can process genomics data at 92 Million genetic bases per second with near linear scaling from 1 to 4 instances. Moreover, CPU utilization on the Power System is just 30% vs. 43% for the Intel system (Figure 10). This provides greater available system capacity to complete the other components of the genomics workflow.

Car-Parrinello Molecular Dynamics (CPMD) is an *ab initio* electronic structure and molecular dynamics program using a plane wave/pseudopotential implementation of density functional theory (DFT). The IBM Power S822LC is 20% faster than Intel Xeon E5-2698 v3 (Fig. 10).

*Record setting
performance
on STAC-M3
Financial
Services
benchmark*

The Securities Technology Analysis Center (STAC)-M3 Benchmark suite (Figure 10) is the industry standard for testing solutions that enable high-speed analytics on time series data, such as tick-by-tick market data (aka "tick database" stacks). STAC-M3 Benchmarks of McObject's eXtremeDB Financial Edition 7.0 on a 2-socket IBM System S824L server with HGST Ultrastar SN150 NVMe PCIe SSDs.

This solution stack set new records in 6 of 17 mean response-time benchmarks, including 1.88X the performance of the previously published best result for 10T.STATS-AGG (SUT ID: KDB150612) and 1.77X that of 100T.VWAB-12D (SUT ID: KDB130603). The system also had more consistent response times (lower standard deviation) than the previous best results for 5 of the 17 operations. Compared to other publicly reported 2-socket systems, this system set records in 11 of the 17 benchmarks. Five of the remaining records are still held by the previous eXtremeDB/Power8 system (SUT ID: XTR141023).
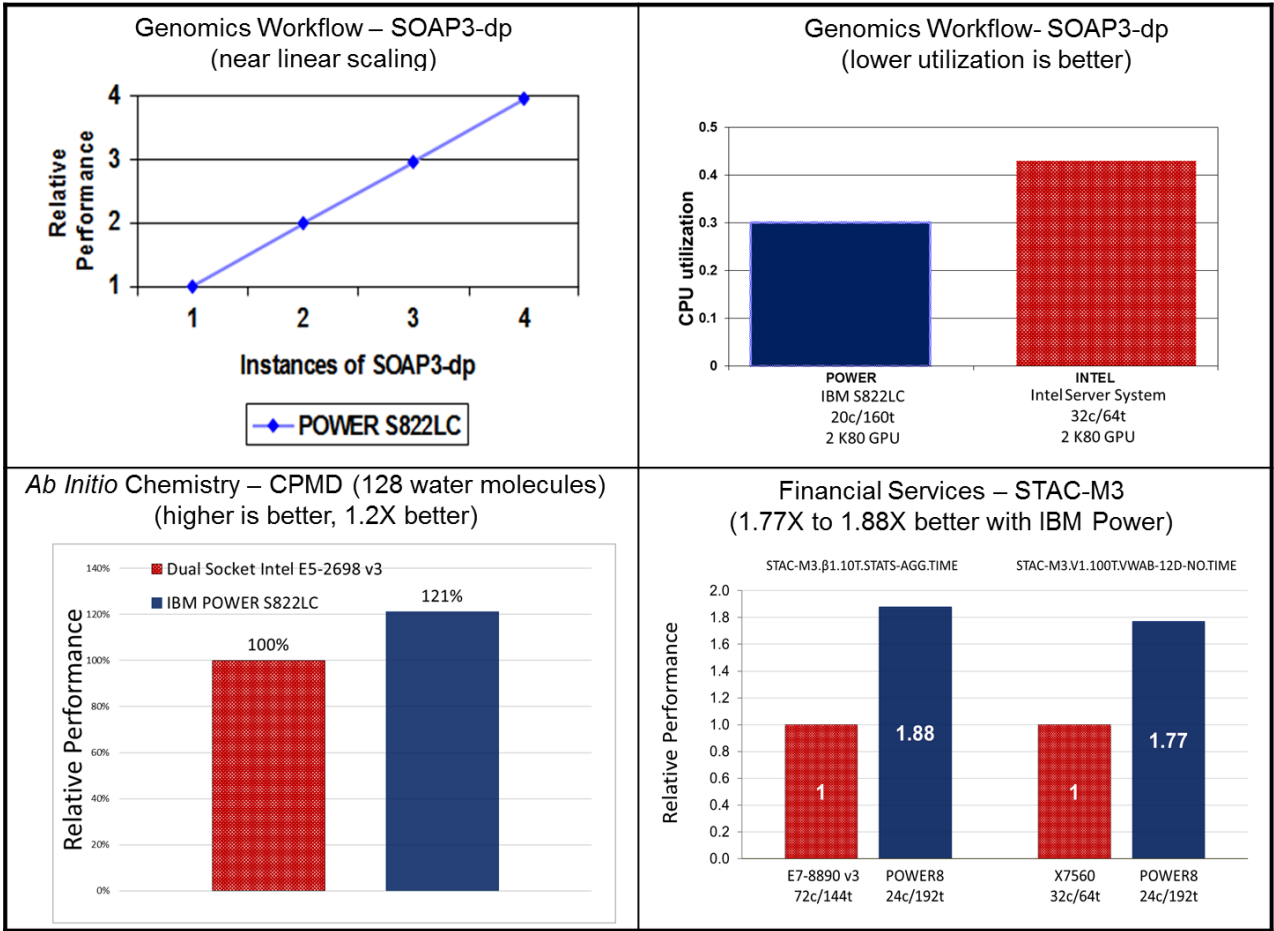
Figure 10: Better Performance for Genomics, *Ab Initio* Chemistry and Financial Services

Machine Learning and Deep Learning: The OpenPower Deep Learning Software Distribution gives developers and data scientists a platform on which to develop new machine learning-based applications and/or analyze data with an emphasis on immediate productivity, ease of use and high performance.

This Software Distribution integrates software and hardware components that have been co-optimized and tested for ease of use and immediate productivity, scalability, stability and performance.   The first release consists of the most advanced and popular deep learning frameworks in the research community:

- Caffe, a dedicated artificial neural network (ANN) training environment developed by the Berkeley Vision and Learning Center at the University of California at Berkeley
- Torch and Theano, two frameworks consisting of several ANN modules built on an extensible mathematics library.

These frameworks can take millions of inputs and desired outputs, and perform the task of training. When combined with the rich, high-performance innovations (including GPU acceleration) available from the OpenPOWER Foundation, these Machine and Deep Learning frameworks can accelerate the Cognitive journey.

13

OpenPOWER Foundation: Driving Collaborative Innovation[8]

OpenPOWER Foundation with over 200 members and growing key to drive stack innovation

In 2014, IBM opened up the technology surrounding Power Systems architecture offerings, such as processor specifications, firmware and software. IBM offers this on a liberal license basis and uses a collaborative development model with partners. The goal is to enable the server vendor ecosystem to build their own customized server, networking and storage hardware for HPC, Analytics, Cloud and future data centers.

The Foundation – representing over 200 global technology leaders and growing – was founded by NVIDIA, Mellanox, IBM, Google and Tyan. The group continues to deliver on an innovation roadmap with over 50 new infrastructure and software innovations, spanning the entire system stack, including systems, boards, cards and accelerators. These new products build upon 30 OpenPOWER-based solutions already in the marketplace, including IBM Power Systems. Wistron, Inspur, Iventec, Supermicro, Penguin Computing and Tyan are some of the other server manufacturers taking advantage of the OpenPOWER technology.
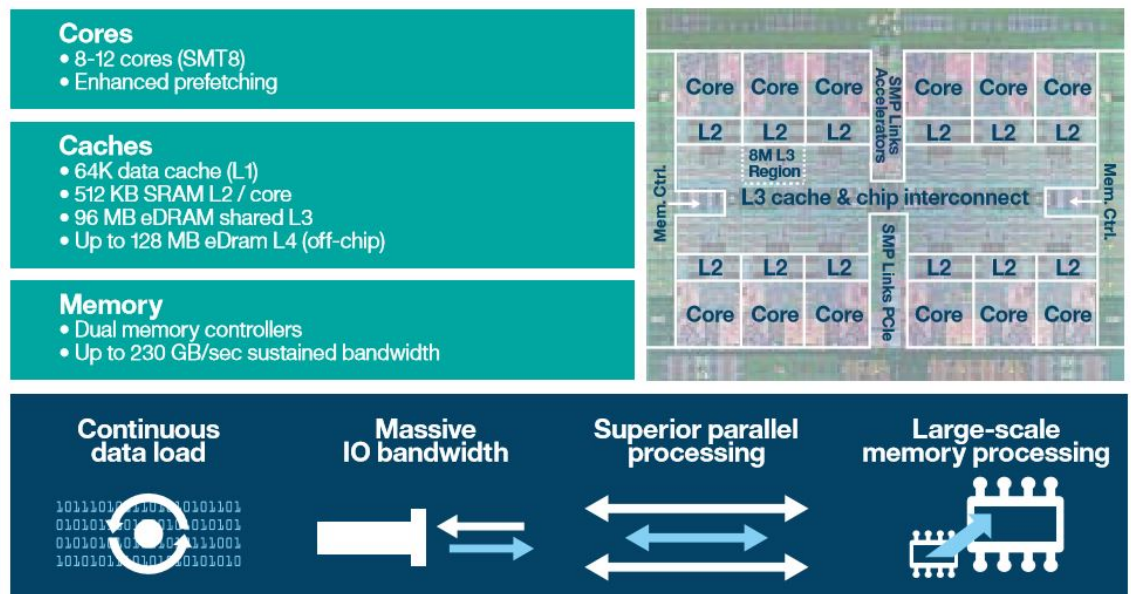
Many server manufacturers leverage OpenPOWER technology

Figure 11: The POWER8 Processor

IBM Power provides massive threads, large memory bandwidth and excellent RAS for HPC

IBM Power Systems (with POWER8 processors): These systems offer a tightly-integrated and performance-optimized infrastructure for HPC /HPDA with the following benefits:

1. **Massive Threads:** Each POWER8 core is capable of handling eight hardware threads simultaneously for a total of 96 threads executed simultaneously on a 12-core chip.
2. **Large Memory Bandwidth:** Very large amounts of on- and off-chip eDRAM caches and on-chip memory controllers enable very high bandwidth to memory and system I/O.
3. **High Performance Processor:** POWER8 is capable of clock speeds around 4.15GHz, with a Thermal Design Power (TDP) in the neighborhood of 250 watts.
4. **Excellent RAS:** IBM Power Systems are engineered to typically perform better than x86 systems in Reliability, Availability and Serviceability (RAS).[9]

---

[8] http://openpowerfoundation.org/

Power Systems Optimized for HPC/HPDA:

The **IBM Power System S822LC** includes up to 2 NVIDIA Tesla GPU Accelerators and Mellanox InfiniBand. The Power S822LC delivers faster time to insight by pairing the built-for-big-data architecture of POWER8 and accelerator performance. It is also available without GPU accelerators.

The **IBM Power System S812LC** is designed for new analytics workloads such as Hadoop and Spark. It delivers performance far exceeding competing offerings for these workloads.

IBM Elastic Storage Server is a software defined cluster storage, combining IBM Spectrum Scale with POWER8 servers and disk arrays. It can be used to deploy petascale, high-speed storage quickly with pre-assembled and optimized servers, storage and software.

The **IBM Power System S812L and S822L, and S824L** are scalable single and dual-socket 2U and 4U servers with POWER8 that can be deployed as single-nodes or clusters.

NVIDIA: The NVIDIA Tesla Accelerated Computing Platform is the leading platform for accelerating deep learning and scientific HPC workloads. The platform combines the world's fastest GPU accelerators, the widely used CUDA parallel computing model, and a comprehensive ecosystem of software developers and Independent Software Vendors (ISVs).

With over hundreds of GPU-accelerated applications, NVIDIA is collaborating with IBM and Mellanox through the OpenPOWER Foundation on a joint roadmap (Figure 9) to accelerate time-critical HPC/HPDA workloads.
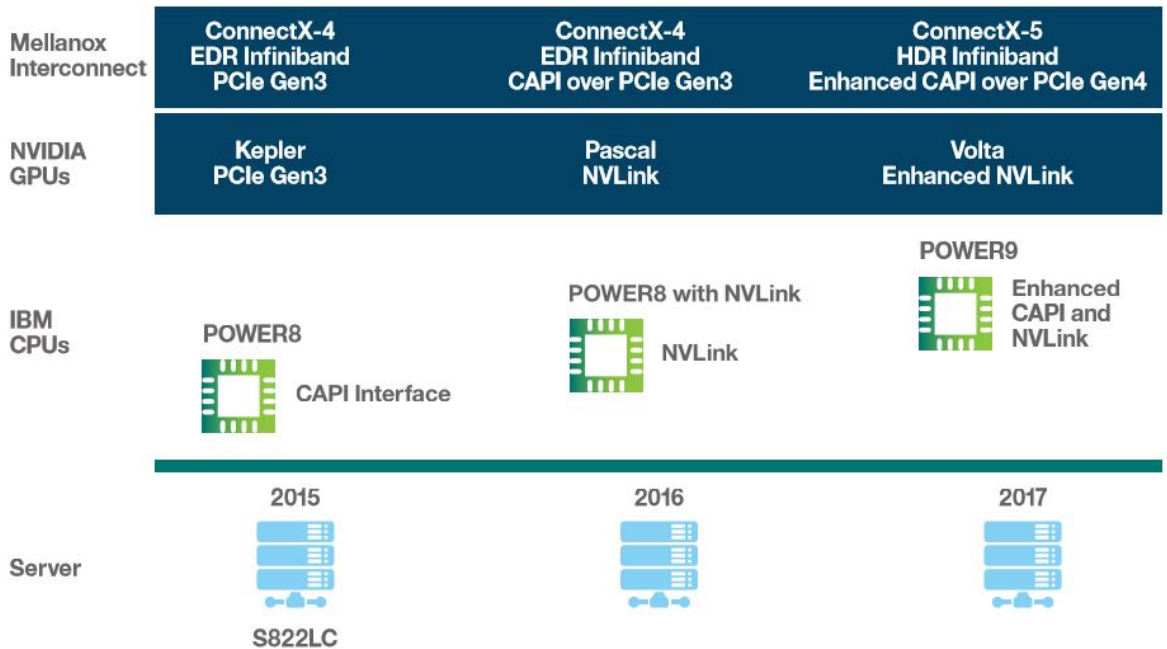
Figure 12: IBM Power Accelerated Computing Roadmap

---

[9] http://www.ibm.com/common/ssi/cgi-bin/ssialias?subtype=WH&infotype=SA&appname=STGE_PO_PO_USEN&htmlfid=POW03133USEN

Mellanox: Server and storage connectivity solutions are designed to deliver very high networking and system efficiency capabilities related to bandwidth, latency, offloads, and CPU utilization for HPC. With networking solutions that deliver 100Gb/s throughput at less than 1 microsecond server-to-server latency, efficient networking hardware offload capabilities, and innovative acceleration software, the Mellanox high-performance networking solution accelerates most HPC workflows and delivers excellent ROI.

Extended use of copper infrastructure from Mellanox versus proprietary fiber optic technology from other providers, results in more reliable connectivity, much lower cost at initial investment and reduced OPEX, as copper cable technology does not require additional power. Mellanox solutions also provide industry standard interoperability and protection of investment with guaranteed forward and backwards compatibility across generations.

Differentiated Acceleration: The IBM Power Accelerated Computing Roadmap offers
choice and flexibility for hardware acceleration of HPC and HPDA workloads. Two different options for differentiated acceleration are available (Figure 10):
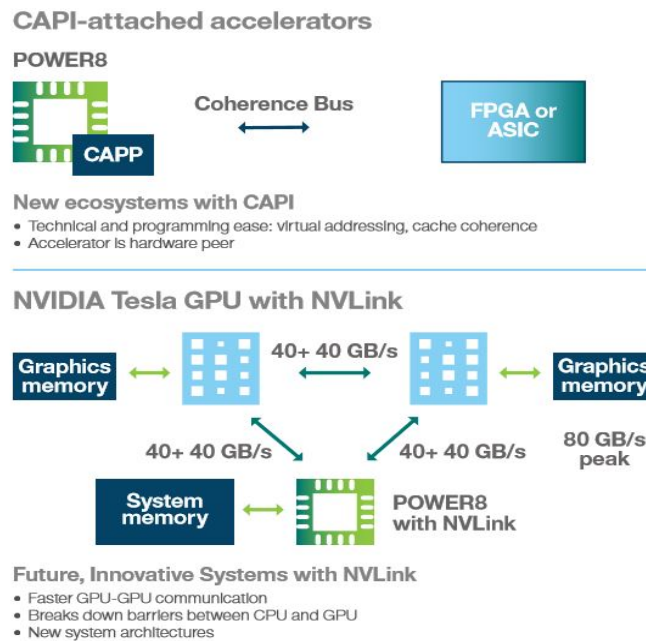

Figure 13: Differentiated Acceleration Interfaces: CAPI and NVLink

Coherent Accelerator Processor Interface (CAPI): CAPI, a direct link into the CPU, allows peripherals and coprocessors to communicate directly with the CPU, substantially bypassing operating system and driver overheads. IBM has developed CAPI to be open to third party vendors and even offers design enablement kits. In the case of flash memory attached via CAPI, the overhead is reduced by a factor of 24:1. More importantly though, CAPI can be used to attach accelerators like FPGAs — directly to the POWER8 CPU for significant workload-specific performance boosts.

NVLink: A broader, fatter pipe to GPUs, enabling the faster host-device, device-device communication many HPC applications require. It also enables new system architectures. In 2016, POWER8 with NVLink will provide faster GPU-to-GPU and CPU-to-GPU communications, accelerating performance even further.

16

Google and Rackspace are working on server architecture specification based on the upcoming IBM POWER9 microprocessors. There are several other real world client examples of innovations and performance enhancements resulting from the OpenPOWER Foundation.

OpenPOWER Client Cases – Accelerating Computing and Learning

Many leading clients benefit from the OpenPOWER Foundation's broad portfolio of HPC and HPDA infrastructure solutions to boost performance, utilization and efficiency. Often, fewer Power Systems are required to address stringent performance and capacity requirements; translating to lower operating costs for facilities, electricity and labor while providing leadership Cognitive Computing capabilities. Here are some real life deployments:

LSU can execute an entire Genome pipeline in few hours vs. multiple days

Louisiana State University (ASC)

Description/ Challenges	• Genome input datasets are growing rapidly— far outpacing the capabilities of the existing Intel-based computing systems • A large metagenome dataset could not be analyzed in a reasonable period of time because of severe memory and I/O constraints.
Solution	• 16 IBM Power System S824L with Ubuntu Linux, 512 TB of Spectrum Scale Storage with IBM Spectrum Computing software.
Benefits	• New innovative scientific methods for bio science analytics. • A total solution for Genome processing, streamlined to complete the analysis in 6.5 hours vs. multiple days. • Reduced server footprint from 120 commodity x86 servers to 40 Power Systems. • Coherent Accelerator Processor Interface (CAPI) enabling terabytes of less expensive flash to appear as main memory to genome assembler application. • Efficient parallelization of the Genome workflow and SMT8 Parallel threading of Hadoop methodology.

University of Michigan[10]

Description/ Challenges	• Accelerate scientific learning in diverse fields such as aircraft and rocket engine design, cardiovascular disease treatment, materials physics, climate modeling and cosmology. • Enable real-time predictive analytics for scientists and engineers using HPC applications integrated with unstructured big data.
Solution	• IBM Power Systems LC servers, designed collaboratively with OpenPOWER Foundation members including Mellanox, NVIDIA and Tyan. • NVIDIA Tesla P100 GPU accelerators with the NVLink high-speed interconnect technology. • IBM Elastic Storage Server, IBM Spectrum Scale software and

---

[10] https://record.umich.edu/articles/u-m-ibm-collaborate-data-centric-high-performance-computing

	IBM Spectrum Computing software.
Benefits	• The POWER8 system significantly outperformed a competing architecture by providing low latency networks and a novel architecture that allows for the integrated use of central and graphics processing units. • Enables large-scale data-driven modeling of complex physical problems, such as the performance of an operating aircraft engine, which consists of trillions of molecular interactions. • Accelerates scientific learning and discovery through cognitive computing by combining noninvasive imaging such as results from MRI and CT scans with a physical model of blood flow. This could help doctors estimate artery stiffness within an hour of a scan, serving as an early predictor of diseases such as hypertension.

OpenPOWER: Accelerating Computing and Cognitive Learning

With the ever increasing volume of data, the boundaries between HPC and Analytics continue to blur. High Performance Data Analytics (HPDA) is growing about 3 to 4 times the rate of traditional HPC. It is the engine of a Cognitive and Learning organization.

By 2019, global spending in Cognitive Computing Systems is expected to reach $31 billion with 17% in hardware.[11] This hardware must perform and scale for Cognitive applications.

Prospective clients must evaluate performance holistically since traditional performance projections based on compute-intensive LINPACK benchmarks are inadequate for HPDA and many HPC workloads. These workloads perform best on data-centric systems that minimize data motion, enable compute capabilities across the system stack and provide a modular, scalable architecture optimized for the *entire workflow*.

Across many HPC and HPDA workflows, OpenPOWER solutions are performing much better than traditional alternatives. Consequently, the membership of the OpenPOWER Foundation continues to grow with newer capabilities and more proof points in HPC, HPDA, Deep Learning and Cognitive Computing.

Organizations should actively consider investing in IBM Power Systems and offerings from the OpenPOWER Foundation to get the following benefits:

• Provides flexibility and choice to deploy well-integrated, best-of-breed HPC/HPDA solution components in an open architecture.
• Minimizes costly data-motion across the entire workflow; accelerating compute and data intensive tasks several fold.
• Lowers total cost of ownership (TCO) with fewer servers with improved utilization, lower application software license costs and less storage and data bottlenecks because of consolidation, fewer redundant copies, novel compression algorithms and efficient data-aware scheduling.

---

[11] http://www.idc.com/promo/thirdplatform/RESOURCES/ATTACHMENTS/CognitiveSystemsInfographic.pdf

- Protects current IT investments in people, processes, platforms and applications while providing a seamless and cost-effective path to scale throughout the Cognitive Computing journey from Data to Analytics to Learning.

With the high-performance offerings provided by the OpenPOWER Foundation, prospective clients can accelerate their Cognitive journey.

Appendix: Performance and Configuration Details

Individual results will vary depending on individual workloads, configurations and conditions.

SPECint_rate2006, SPECfp_rate2006 and STREAM TRIAD:

IBM Power S822LC SPECint_rate2006 results based upon published results at http://spec.org/cpu2006/results/res2015q4/cpu2006-20151002-37621.html
IBM Power S822LC SPECfp_rate2006 results based upon published results at http://spec.org/cpu2006/results/res2015q4/cpu2006-20151002-37620.html
HP DL 180 Gen9 SPECint_rate2006 results based upon published results at http://spec.org/cpu2006/results/res2014q4/cpu2006-20141006-32106.html
HP DL 180 Gen9 SPECfp_rate2006 results based upon published results at http://spec.org/cpu2006/results/res2014q4/cpu2006-20141006-32118.html

IBM Power System S822LC results are based on IBM internal measurements of STREAM Triad; 20 cores / 20 of 160 threads active, POWER8; 3.5GHz, up to 1TB memory,
Intel Xeon data is based on published data of Intel® Server System R2208WTTYS running STREAM Triad; 24 cores / 24 of 48 threads active, E5-2690 v3; 2.3GHz. For more details, see http://www.intel.com/content/www/us/en/benchmarks/server/xeon-e5-2600-v3/xeon-e5-2600-v3-stream.html

OpenFOAM: Results are based on IBM internal testing of systems running OpenFOAM version 2.3.0 code benchmarked on POWER8 systems. Individual results will vary depending on individual workloads, configurations and conditions.

IBM Power System S822LC, POWER8; 3.5 GHz, 512 GB memory; 2x 10 core processors/ 4 threads per core.
BULL R424-E4, Intel Xeon E5-2680v3, 2.3 GHz, 256 GB memory, 2x 12 core processors / 1 thread per core.

NAMD: Results are based on IBM internal testing of systems running NAMD version 2.10 APOA1, F1ATPASE, STMV code benchmarked on POWER8 systems installed each with 2 NVIDIA Tesla K80 GPUs. Individual results will vary depending on individual workloads, configurations and conditions.

IBM Power System S822LC; 16 cores / 128 threads, POWER8; 3.3GHz, 128 GB memory
IBM Power System S822LC; 16 cores / 128 threads, POWER8; 3.3GHz, 128 GB memory, 2 NVIDIA K80 GPUs
Supermicro 2028GR-TRT, 16 cores, x86, 2.3 GHz, 128GB memory, 2x Tesla K80 GPUs

tranSMART: Results are based on IBM internal testing and optimization using:
IBM Power S822LC 16-core, 3.32 GHz 8335-GTA with SMT=2, Turbo Mode 512 GB memory, IBM XL Compiler Suite, IBM ESSL and IBM Elastic Storage Server Model GL4

GATK: Results are based on IBM internal testing of systems with:
IBM POWER8 Firestone: 16-core, 3.32 GHz 8335-GTA with SMT=8, Turbo mode, 1024 GB memory and IBM ESS with Spectrum Scale Model GL4
Best known Intel result at 44 hours (without variant discovery by HaplotypeCaller)
http://library.wolfram.com/infocenter/Conferences/9045/Intel_LifeSciences_Personalized_Medicine_Wolfram%202014_Paolo%20Narvaez.pdf

SOAP3-dp: IBM internal measurements with:
IBM Power System S822LC, POWER8; 3.5 GHz, 256GB memory; 2x 10 core CPUs, 2 Tesla K80 GPUs
Intel Xeon E5-2698 v3, 32 cores, 2.3GHz, 256 GB, 2 Tesla K80 GPUs

CPMD: Results are based on IBM internal testing of systems.
IBM Power System S822LC; 10 cores, POWER8; 2.93 GHz, 256 GB memory, #MPI =20, #OMP=1
Intel Xeon E5-2698 v3, 16 cores, 2.3GHz, 128 GB, #MPI=32, #OMP=1

STAC-M3: All IBM Data is found in STAC Configuration Disclosure for this SUT: www.STACresearch.com/XTR151016
Based on STAC-M3. 10T.STATS-AGG. STAC Configuration Disclosure for this SUT: www.STACresearch.com/KDB150612
Based on STAC-M3. 100T.VWAB-12D. STAC Configuration Disclosure for this SUT: www.STACresearch.com/KDB130603

Cabot Partners is a collaborative consultancy and an independent IT analyst firm. We specialize in advising technology companies and their clients on how to build and grow a customer base, how to achieve desired revenue and profitability results, and how to make effective use of emerging technologies including HPC, Cloud Computing, and Analytics. To find out more, please go to www.cabotpartners.com.