Smart IBM Solutions for High Performance Cloud Computing: Supercomputing Performance for Complex Problems Delivered in a Flexible, Secure and On-Demand Mode through Public, Private, and Hybrid Clouds.

Sponsored by IBM

Srini Chari, Ph.D., MBA

September, 2010

mailto:chari@cabotpartners.com

Introduction – Why High Performance Computing is Turning to Clouds

In today's climate, organizations must innovate with flexibility and speed in response to customer demand, market opportunity, regulatory changes, or a competitor's move. The current economic downturn and the escalating energy and people costs for information technology (IT); however, will force companies to reevaluate how they can maximize their returns on investments. They will need smarter approaches to reduce costs, manage complexity, improve productivity, reduce time to market, and enable innovation. Simply put, companies must and will carefully examine the business value and cost of IT investments.

As high performance computing (HPC) mainstreams, its business innovation impact expands: High performance computing (HPC) uses supercomputers and clustered computers in the teraflops to petaflops performance range to solve computational and data intensive problems. HPC helps enterprises achieve the speed, agility, insights, and sustained competitive advantage to deliver innovative products, increase revenues, and improve operational performance. Scientists, engineers, and analysts in leading-edge enterprises rely on HPC to solve challenging problems in fields ranging from engineering, manufacturing, finance, risk analysis, revenue management, to life and earth sciences to name a few. This mainstreaming of HPC has put considerable pressure on solution providers to provide scalable, reliable, and secure solutions while reining in costs and complexity.

Cloud computing provides effective approaches to reduce costs and complexity…but enterprise class security and reliability remain concerns: Cloud computing -- in which large amounts of data and computing resources can be accessed remotely over the Internet using a personal computer, cell phone or other device -- holds great promise in the IT market. The cloud model has the potential to cut the costs, complexity and headaches of technology deployment for companies, universities, and government agencies.

The potential benefits to enterprise clients are immense. Cloud computing turns the economics of enterprise IT on its head. Delivery of IT services (including infrastructure, platform, and applications) from the cloud has both capital expense and operation expense advantages. The ability to pool resources, virtualize them, and then dynamically provision from the resource pool yields a much higher utilization rate and thus better economics -- sometimes improving system utilization from 15 percent to 90 percent!

Already, companies such as Amazon.com, Google and Salesforce.com, among others, offer cloud-based Web services including e-mail, computer storage, and customer management software. But many enterprises and government agencies have been wary of cloud computing because of traditional IT concerns like data security, reliability of service, and regulatory compliance. Those concerns are now being fully addressed as trusted computer solution providers like IBM apply their fundamental strengths in deploying enterprise computing level security and bringing reliability to cloud computing.

Cloud computing will dramatically impact the way IT services are consumed and delivered in the future. Our broad workload-driven approach strongly supports the view that cloud computing is a rapidly growing IT opportunity throughout the value chain; analyst estimates range from the tens of billions of dollars to even over hundred billion dollars in the next 3-5 years[1]. Regardless of the precise estimate, we believe:

- The market opportunity is large with growth rates much faster than the overall IT industry,
- Linux and open source will remain pervasive throughout the cloud ecosystem; becoming even more dominant with the emergence of cloud standards to drive interoperability between clouds,
- Private and hybrid clouds will become the dominant cloud delivery models as enterprise workloads begin to leverage the promise of clouds and security concerns persist with public clouds,

[1] Markus Klems, "Merrill Lynch Estimates Cloud Computing to be $100 Billion Market", http://cloudcomputing.sys-con.com/node/604936

- Before making substantial new cloud investments, businesses will carefully examine the business case that will be primarily driven by their current and future workload needs, and lastly,
- Customers will rely on cloud providers who have the deepest insights into their workloads and can deliver a broad portfolio of cloud services and systems optimized to these workloads.

IBM offers a rich portfolio of HPC cloud computing solutions with substantial business benefits: Through in-depth research and interviews, this paper describes IBM's rich portfolio of HPC cloud computing offerings. IBM has been the leader in clustered high performance computer systems for many years and consistently dominates the TOP500 (www.top500.org) list of the world's most powerful supercomputers. In 2008, building upon prior initiatives in service-oriented architectures (SOA), Linux, and energy-efficient dynamic infrastructure, IBM announced a significant company-wide cloud computing initiative, tying its systems, software, and services businesses; driven and governed by a central cross-brand group. Since then, IBM outlined its cloud computing vision (www.ibm.com/cloud), strategy, and a detailed roadmap of specific HPC cloud solutions deployed at IBM customer sites with the following bottom-line benefits that are described in greater detail later in this paper:

- ***24x7 access to general and HPC cloud computing for students, researchers, and faculty*** at North Carolina State University,
- ***democratizing HPC through cloud computing*** at the Rocky Mountain Supercomputing Center in Montana,
- ***robust and flexible hybrid HPC cloud platform*** using IBM partner Gridcore's Gompute service,
- ***a private cloud for global research collaboration*** at SciNet, Canada's largest supercomputing site,
- ***making HPC accessible to smaller manufacturers*** using Nimbis, a cloud services provider,
- ***cutting costs and streamlining infrastructure with the IBM iDataPlex***, a large financial services firm.
- ***enhancing high-resolution climate modeling on the Discover supercomputer cloud with the IBM iDataPlex,*** NASA Goddard Space Flight Center.

IBM HPC cloud computing offerings deliver the performance and value of supercomputing to multiple users in an inexpensive, efficient, flexible, reliable, and secure manner on public, private, and hybrid clouds.

As Cloud Computing Mainstreams, Multiple Delivery Models Emerge

In many ways, cloud computing is the next logical evolutionary step, building upon the industry's rapid adoption of Linux, open source solutions, high performance cluster computing, SOA, and more recently virtualization. While many definitions and analogues to the electric utility industry exist, we believe that most do not completely capture the cloud's broader concepts and business benefits. Here, we present a perspective that motivates this discussion with cloud business models that parallel the passenger airline transportation industry.

The value of cloud computing: Cloud computing promises to provide dynamically scalable and often virtualized IT resources (hardware, software, and applications)- as a service transparently to a large set of users who may possess a broad but differing range of knowledge, expertise in, or control over the technology infrastructure. The concept incorporates software as a service (SaaS), Web 2.0 and other popular, recent, internet computing trends such as SOA, and also builds upon recent IT infrastructure solution concepts such as grid/cluster computing, utility computing, and autonomic computing.

With a spectrum of flexible offerings and pricing models, cloud service providers are well-poised to provide secure, affordable, elastic, often automated with "self-service" access to IT resources for companies that need to quickly scale-up or scale-down their IT needs to adapt to their business demands. Cloud computing can transform companies of all sizes to become more agile and develop sustainable competitive advantage while reining in costs.

With cloud computing solutions, smaller companies – that typically face steep entry cost barriers to access IT resources - such as internet companies, service providers, or Independent Software Vendors (ISVs)- will no longer need large capital outlays in hardware or facilities to deploy their services or the labor to operate these IT facilities. On the other hand, larger organizations benefit from the increased business value resulting from the added capability and flexibility to rapidly deploy standardized yet customizable "self-service" solutions that automate and scale business processes end-to-end while minimizing escalating labor and infrastructure costs.

A broad range of business and delivery models: Over the last two years, with the increasing interest in cloud computing, many excellent articles[2, 3, 4, 5] have characterized or defined cloud computing. Briefly, following typical IT architectural stacks, cloud services are delivered as: Infrastructure – servers, storage, etc. - as a Service (IaaS), Platform - a software development environment – as a Service (PaaS), Software – typically applications – as a Service (SaaS), or even Business processes as a Service. These services can be implemented and delivered as a private (within an enterprise) cloud or as a public (accessible through the internet) cloud, or as a secure hybrid (extending private) cloud. While analogues to the electric utility industry are popular, we believe, from a customer perspective, it is more illustrative (and realistic) to depict the wide spectrum of cloud business model choices with an analogy from the passenger airline industry as summarized in the following figure.

IT Industry	Transportation	Cost Model	Business Impact
In-House Data Center or Private Cloud (not-outsourced)	Custom built planes	CapEx, OpEx, long term	Total control and responsible for entire infrastructure including service levels
Private Cloud (outsourced)	Lease planes and service	CapEx and OpEx in medium term lease	Control of infrastructure but not fully responsible for service levels
Hybrid	Charter planes	Pay for duration, short to medium term/trip	Some control of infrastructure and depend on provider for service levels for duration
Public Cloud	Commercial airlines	Pay as you use, very short term/trip/user	Almost no control of infrastructure or service levels

While the evolution of public cloud adoption has been rapid, particularly with smaller businesses and individual developers, early adopters at larger enterprises are increasingly turning to private and hybrid clouds to address concerns (with public clouds) of security, regulatory compliance, governance, reliability, and IP protection. These larger enterprises, who already have substantial in-house IT investments, are implementing private or hybrid clouds to improve utilization levels, reduce operational expenses, and can - through "self-service" portals - dynamically provision IT services in minutes or hours instead of months. With open systems management and workload scheduling solutions powering their cloud implementations, these enterprises will be able to rapidly develop, customize, test, and roll-out new business services and applications to their users and clients while reining in costs.

Workload Classification: Making the Case for Cloud Computing

Accurately forecasting application workload requirements is a major challenge for IT managers and planners. This challenge becomes even more acute as businesses depend increasingly on high performance analysis and web applications which have more variability than traditional enterprise business applications. Using a simple workload characterization and classification model developed here, we examine workload trends that are poised to fundamentally transform the delivery of IT solutions through cloud computing.

First, we examine a wide range of workloads typical in many IT applications across several dimensions. These applications range from traditional enterprise business and transactional applications to more compute intensive High Performance Computing (HPC) applications and web/business analytics. These applications are classified according to their typical workload characteristics with compute-intensive/job on the x-axis and workload variability[6], $V_{W,}$ on the y-axis. The bubble size in this chart is indicative of the total server capacity deployed globally to execute these applications. The arrows indicate the size of workload growth in the future across the two primary dimensions. For

[2] Wikipedia
[3] Tim Jones, "Cloud Computing with Linux", http://download.boulder.ibm.com/ibmdl/pub/software/dw/linux/l-cloud-computing/l-cloud-computing-pdf.pdf
[4] Jeffrey Rayport and Andrew Hayward, "Envisioning the Cloud: The Next Computing Paradigm", March, 2009.
[5] Ashar Baig, "A Cloud Guide for HPC", May 2009.
[6] Srini Chari, "Confronting the Data Center Crisis: A Cost Benefit Analysis of IBM's Computing on Demand Cloud Offering", http://www-03.ibm.com/systems/resources/tcocodpaperfinal.pdf

example, traditional transactional applications and ERP comprise a large part of the workload today but they are not very compute-intensive and exhibit low variability. On the other hand, web analytics is an emerging area that is currently small but expected to grow rapidly. HPC applications are normally very compute intensive often requiring 100s of CPUs to execute one job and since these applications are often used for complex analyses, workload variability is large and frequently difficult to predict *a-priori*. Web searching capability is becoming deeper and more complex with multi-modality capability beyond simple text searches. With more users using complex search, transactional web applications, and web analytics, we expect the web workload to become more compute-intensive with increasing variability. Business analytics is a large portion of the workload and is becoming more variable and compute-intensive. The large development and test environments have cyclical patterns with increasing variability to adapt to severe pressures to deliver new products and services at ever reducing cycle times. And finally, with the interconnected and mobile nature of today's business environment, collaboration workloads are large, will grow, and become more variable.

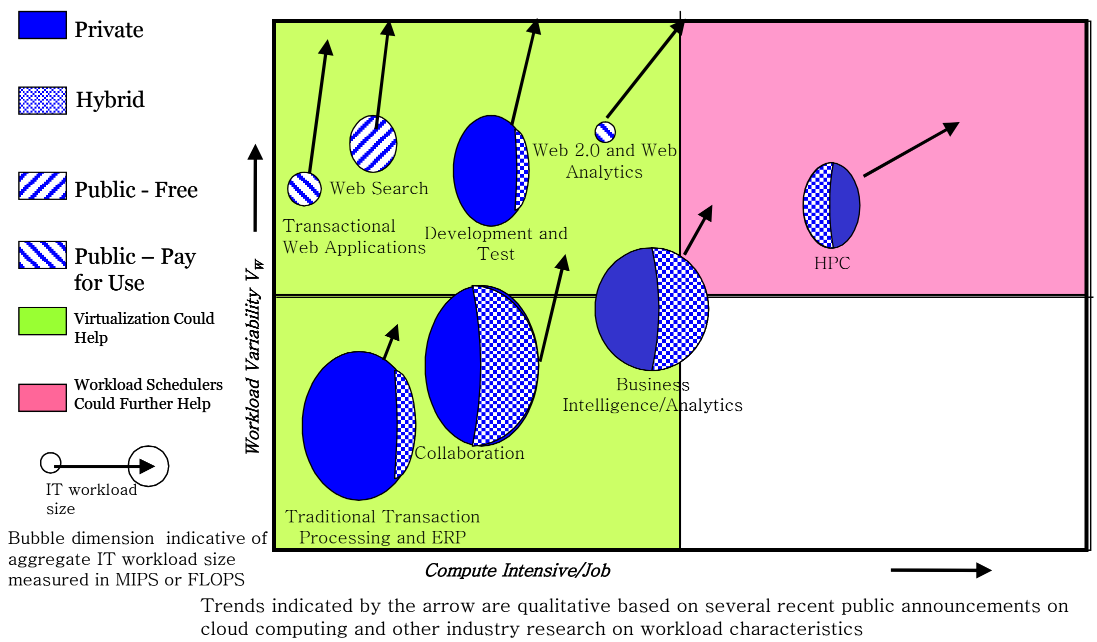


Figure 1: High Level Workload Classification by Application Domains

Another important dimension in the workload analysis is the way these applications are typically delivered and executed today and what may be expected in the future. Again, in a non-prescriptive manner, we break up the typical delivery models into Private (In-House), Public (public access through the Internet), and Shared Hosted (Internet access through Virtual Private Networks (VPNs) contracted and customized to the needs of the end-users). Public is further divided into Public-Free which is a free access to the end-user (e.g. Google search) and Public-Pay for Use (e.g. Amazon Web Services) which is a pay-for-service infrastructure utility model. As depicted in Figure 1 and consistent with the opinion of most analysts, we believe that most of the future cloud opportunity is in private or hybrid clouds.

Virtualization and workload scheduling and management are key software solutions that often help increase system utilization and overall data center efficiency. With virtualization and consolidation, many low-to-moderate compute intensive workloads can be mapped onto fewer physical systems without adversely impacting service levels. Hence, workloads depicted in the left quadrants in Figure 1 could benefit, substantially resulting in efficiency gains for customers. VMware is one prominent example of a virtualization solution. The IBM System z mainframe is another excellent example. However, at the other end of the spectrum, with HPC and other compute intensive workloads often requiring several CPUs per job, workload scheduling and management solutions from Platform Computing, Adaptive Computing, Gridcore, and the IBM Tivoli LoadLeveler have been used to increase overall system utilization and throughput in cluster configurations. Virtualization solutions that usually consolidate several jobs onto one CPU may actually adversely impact the scalability and performance of HPC and analytics jobs especially parallel batch applications that use data partitioning. However recently, holistic systems designs and reliability-aware software solutions based on virtualization and scheduling solutions have been used to enhance reliability and availability and to maximize HPC application uptime[7] while delivering the quantum increase in performance and scale needed for tomorrow's HPC applications. IBM led Extreme Cloud Administration Toolkit (xCAT) solutions have been used effectively in this regard for very large scale-out cloud environments.

IBM Delivers a Broad Portfolio of Workload Optimized Cloud Computing Offerings

IBM announced a broad vision and strategy for cloud computing with an initial set of offerings spanning its major businesses – software, systems, and services. IBM then extended these initial offerings with a broader set of offerings with increased capability optimized for specific high-opportunity workloads. Branded as IBM Smart Business services and IBM Smart Business systems (www.ibm.com/cloud), these offerings are tailored for development and test, collaboration, business analytics, and information-intensive (compute and storage) workloads behind a company's firewall or on the IBM Cloud. Of special mention are the IBM Compute Cloud, IBM HPC Cloud, and the IBM Smart Analytics Clouds which are relevant for HPC workloads. The following figure (courtesy IBM) summarizes key elements of the current – as of February 2010 - IBM cloud computing offering portfolio.

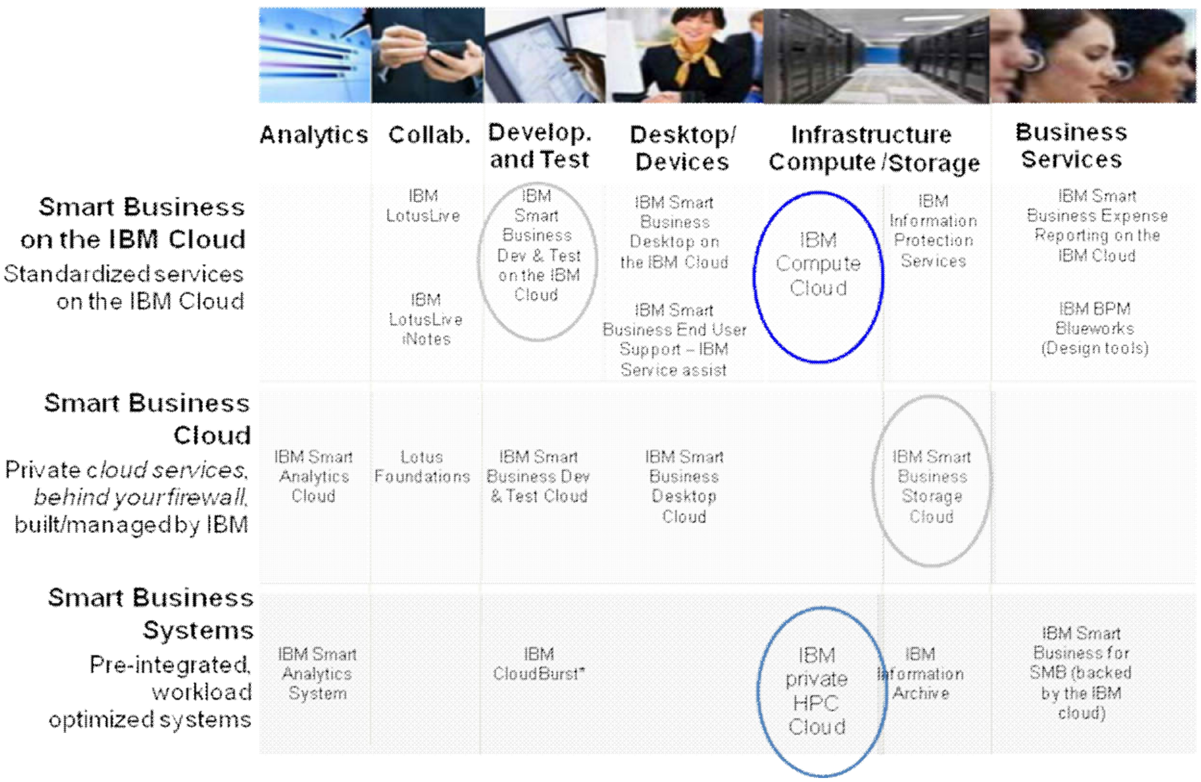


Figure 2: IBM's Broad Portfolio of Cloud Computing Offerings with a Focus on HPC Clouds

IBM has been leveraging cloud computing for internal transformation for at least three years and working with early cloud adopter customers for "in-market" experiments at any one of its twenty cloud laboratories through-out the world. These efforts and its global enterprise data center outsourcing business have formed the foundation of IBM's current cloud initiative.

---

[7] Raju Gottumukula, *et. al*, "Reliability-Aware Resource Allocation in HPC Systems", IEEE International Conference on Cluster Computing, 2007, pp 312-321.

IBM High Performance Cloud Computing Solutions

Since users consume high performance computing in different ways, IBM delivers HPC in all the three different cloud architectures – private, shared hosted, and hybrid clouds, and services to help clients with deployment.

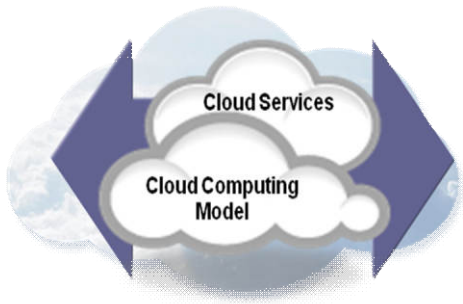
Private Cloud

Client owned. Managed by client or services provider.

Access defined by client.

Advantages

Security, privacy, customization & control



Shared Hosted Cloud

Standardized services on the provider's cloud.

Access by subscription.

Advantages

Standardization, capital preservation, flexibility and time to deploy

Hybrid Cloud

A mix of models

Figure 3: IBM Delivers HPC Cloud Solutions as Public, Private, and Hybrid Clouds

IBM Shared hosted Cloud -- 'Infrastructure as a Service'

The IBM Cloud can be easily adapted for the HPC community. Clients leverage a security-rich supercomputing environment that can be used like in-house hardware, but without the capital commitment. If computing demands exceed in-house capacity, users can shift the excess workload to the IBM Cloud and use the processing power they need to meet demand. With the IBM Cloud, users can get the extra power they need, but they only have to pay for the capacity they use. Peaks and valleys may still exist, but IBM can help smooth them out. Users can implement an optimal combination of in-house fixed capacity or fixed cost, and hosted variable capacity or variable cost HPC infrastructure.

IBM Private HPC Cloud

When a cloud environment is created inside an enterprise firewall, it can provide users with the same rapid access to IT as the public model, but with less exposure to Internet security risks. This can often make private clouds more appropriate for specialized programs and systems unique to organizations with extremely sensitive data that must be protected.

In November 2009, IBM announced a new solution, the IBM Smart Analytics Cloud, for clients to build their own private cloud environments. This is based on the same cloud infrastructure that IBM uses internally. IBM's internal cloud, the world's largest private cloud computing environment for business analytics, provides IBM sales teams and developer's new levels of insight to better meet the needs of clients worldwide. The cloud launched with more than a petabyte of data, the equivalent of more than 300 billion ATM transactions.

An IBM HPC private cloud typically begins with an IBM Intelligent Cluster or an IBM Power Systems Cluster coupled with a choice of one or more components from the IBM HPC Cloud Software stack and key business partners:

- **xCAT (Extreme Cloud Administration Toolkit –** www.xcat.sourceforge.net**).** An open source scalable distributed computing management and provisioning tool that provides a unified interface for hardware control, discovery, and operating system diskful/diskfree deployment. This robust toolkit can be used for the deployment and administration of AIX or Linux clusters and Microsoft Windows. Its features are based on user requirements, and take advantage of IBM System x and Power Systems hardware.
- **Moab Cluster Management Software –** www.clusterresources.com/moab**.** Available as part of the System x Integrated Solutions/ Intelligent Cluster portfolio, or as a stand-alone software option, Moab Adaptive HPC Suite and Moab Adaptive Computing Suite are intelligent workload management software offerings by

Adaptive Computing, Inc. The key benefits of Moab include the ability to dynamically adjust resources with flexible management policies for user, group and workloads needs and also enforces Quality of Service (QoS) and Service Level Agreements (SLAs) based on organizational objectives.

- **Tivoli Provisioning Manager –** www.ibm.com/software/tivoli/products/prov-mgr**.** Helps organizations optimize efficiency, accuracy, and service delivery by automating best practices for data center/cloud provisioning activities on Microsoft Windows servers and clients, as well as Linux and AIX servers.
- **Tivoli Workload Scheduler LoadLeveler –** www.ibm.com/systems/software/loadleveler**.** Used for dynamic workload scheduling, Tivoli Workload Scheduler LoadLeveler is a distributed network-wide job management facility designed to dynamically schedule work to maximize resource utilization and minimize job completion time. Jobs are scheduled based on job priority, job requirements, resource availability and user-defined rules to match processing needs with resources. LoadLeveler provides consolidated accounting and reporting and supports IBM servers including IBM Power Systems and IBM System x environments.
- **GPFS (General Parallel File System –** www.ibm.com/systems/software/gpfs**).** A high-performance cluster file management infrastructure for AIX, Linux, Windows and mixed clusters (x86 and Power) that provides performance, scalability and availability for file data. It is designed to optimize the use of storage, to support scale-out applications and to provide a high availability platform for data intensive applications. GPFS provides online storage management, scalable access and tightly integrated information lifecycle tools capable of managing petabytes of data and billions of files. GPFS can help clients move beyond simply adding storage to optimizing data management.

IBM Hybrid HPC cloud (Private/Shared Hosted)

Companies can also use a mix of cloud models, depending on how critical data is for particular applications and uses. Public clouds are often used for normal activities and situations, such as word processing, document management, scheduling and social media. More sensitive applications can be shared on a cloud behind a firewall. A typical hybrid cloud can comprise an IBM Intelligent Cluster based private cloud with some tasks spilling over to the IBM Cloud.

Private cloud scenarios can also enable smaller companies with data-intensive processes to expand their infrastructure when needed, and not worry about vital information getting into the wrong hands.

IBM System Intelligent Cluster for HPC Clouds

IBM System Intelligent Clusters, based on IBM System x x86 servers and related storage and file management resources, offer leading edge technology, flexibility, and high performance at an attractive price point, energy efficiency, and ease of deployment for clients looking to deploy HPC cloud computing solutions. (For more on IBM System x, see www-03.ibm.com/systems/x/).

Servers	+	**Switches & storage**	+	**Cluster software**	=	**IBM Intelligent Cluster**
IBM System x - iDataplex -Rack servers - BladeCenter		Ethernet, InfiniBand, Fibre Channel, IBM Systems Storage, etc.		Linux/ Windows Management software - xCAT - Moab - Tivoli IBM cluster file system (GPFS)		

IBM HPC Cloud Services Offerings

IBM offers services to simplify and expedite the planning and implementation of a cloud environment. The IBM HPC Open Software Stack is intended to ease the deployment of supercomputing clusters. The Open Software stack includes several distinct software tools that have been tested and integrated by IBM. These include xCAT, Advance Toolchain for Power Systems; install scripts, a resource-management tool and a cluster-administration toolkit. The Open Software Stack makes it easy for universities and academic researchers and others– who need an open source community development environment with full access to source code. IBM's commercial stack for enterprise customers is production ready, secure, and fully supported with the ability to bill on demand.

- **IBM Deep Computing Services for HPC Clouds.** The IBM Deep Computing Services offering is a market driven response to customers seeking highly specialized skills and expertise from the IBM research and development teams. These services can be provided as an element of the total opportunity, or the service offering can be used to facilitate a more formal collaboration. In either case, the appropriate skills will be identified and made available to meet the customer's need to design, install and optimize a private or hosted private cloud with spill-over, if desired, to the IBM Cloud.

- **IBM University Delivery Services – for Virtual Computing Lab support.** Cloud computing for education and healthcare programs based on virtual computing lab (open source). Includes installation, configuration, instructions, and support.

- **IBM Cloud Infrastructure Workshop.** Feasibility assessment for implementing advanced IBM technologies for a private cloud with estimates for the expected cost and operational benefits.

Customer Examples Highlighting the *Benefits* of IBM HPC Cloud Implementations

IBM and its business partners serve a variety of customers who use cloud computing environments to access HPC capabilities. These HPC cloud customers represent a range of industries and application uses. In all cases, IBM's technology, expertise, resources, and support are critical to the customer.

North Carolina State University – *24x7 access to general and HPC cloud computing for students, researchers, and faculty*

IBM and North Carolina State University have collaborated to establish the Virtual Computing Lab (VCL)[8], a cloud computing-based technology, to provide students around the state and the University of North Carolina system campuses access to advanced educational materials, select software applications and computing and storage resources on demand. Today, VCL is open to 30,000+ NCSU students and faculty. At any given time, 1300 to 1800 IBM BladeCenter blades are in use of which 500 to 700 work in non-HPC mode and the rest in HPC mode. VCL delivers over 460,000 CPU hours annually to general workloads, over 7,000,000 HPC CPU hours annually, distributed over four data centers.

How it works: The VCL infrastructure consists of three tiers: a web server, a database server, and one or more management nodes. At the heart of VCL is a web-based service for scheduling and provisioning remote access to high-end computational resources. These resources consist of blade computers located in multiple data centers and other specialized University lab computers. The IBM BladeCenter compute resources are dynamically loaded on demand with a choice of operating system images and predefined application sets in either bare metal or hypervised environments. The blade servers also provide flexibility between High Performance computing and academic computing by easily repurposing during low use times between clusters used for batch processing or single seat use for less compute intensive work. IBM University Delivery cloud services are also available to other institutions and their students/faculty to replicate the VCL cloud. This service facilitates remote collaboration and access to servers and comes with additional support for installation and configuration, instruction and training, and systems/image management.

Challenges: VCL began in 2004 with a simple idea of providing dedicated remote access to a range of computing environments for students and researchers to access from any networked location either on or off campus. In a shared

[8] http://vcl.ncsu.edu/

computing resource environment, where many students run high-end applications or experiment with computer science coding assignments on the same computer, the level of service degrades very quickly. Additionally, software media distribution for distance education students was prohibitive. Depending on the application and the related license agreement, it was limited to university owned computers and was not allowed to be installed on the student's personal computer.

Why IBM: VCL exemplifies the goals of the Virtual Computing Initiative (VCI), launched by IBM and NC State University in 2006, to improve the quality of education through the application of technologies that include virtualization, cloud computing, hosted client-server models, and robust, energy efficient IBM Systems, etc.

Rocky Mountain Super Computing Center, Butte, Montana – *democratizing HPC through cloud computing*

Rocky Mountain Supercomputing Centers, Inc. (RMSC)[9] is a private non-profit company established in 2009 by the state government of Montana with support from IBM to make High Performance Computing and Supercomputing accessible and affordable for commercial, academic and governmental customers in the state and beyond. The pay-as-you-go RMSC computing model has already attracted a wide range of customers including university professors performing astrophysics and climate modeling; a biotech company; a company performing massive pattern recognition for the U.S. Navy; financial companies analyzing real-time stock feeds; and an Indian reservation looking at terrestrial carbon sequestration in tribal land.

How it works: IBM provided consulting and implementation services to help RMSC establish their HPC cloud based on IBM BladeCenter and IBM System p Power clusters and IBM Deep Computing Visualization (DCV) for remote 3D visualization. The Extreme Cloud Administration Toolkit (xCAT) and Adaptive Computing Moab are used to dynamically allocate the IBM supercomputer resources, simplify management, and schedule a wide range of HPC applications and workloads. RMSC has a total performance capability of 3.8 teraflops. Though RMSC has plans to scale up to 20 to 50 teraflops, it is not in any hurry to grow -- RMSC's goal is to democratize HPC and make it available to users in the Montana region who otherwise had no easy way to solve problems requiring HPC.

Challenges: Before the RMSC facility, Montana was in the words of its governor Brian Schweitzer, "in the middle of a supercomputing desert". This hampered Montana from providing a plank to attract high-tech, high paying jobs. The government wanted to provide affordable high performance computing through an on-demand distributed system.

Why IBM: RMSC leveraged IBM Deep Computing Services and expertise to establish and support the initial implementation of the center. Users have access to RMSC computing resources providing supercomputing consumers a seamless way to scale and out-perform the competition while transacted on a pay-as-you-use market-based fee model.

Gompute – *robust and flexible hybrid HPC Cloud platform*

Gompute[10] is Gridcore's HPC on Demand Service. Gridcore AB, an IBM business partner, is a Swedish company founded in 2002 focused on delivering complete solutions for technical and scientific computing. Divided in three business areas -- systems integration, software development and on-demand computing solutions -- Gridcore offers a complete set of services, hardware and software solutions to help accelerate the development of clients' products and use of virtual prototyping. Leveraging Gompute, Gridcore offers customers a complete ready to use HPC Cloud platform to set-up and manage an in-house private cloud and/or seamless access to their public cloud.

How it works: Users of Gompute get a ready to use HPC resource with the needed software and hardware. They access these resources by either a simplified web interface or by a thin client which allows them to run applications using their native graphic interface. Gompute supports many applications, some of them are distributed as open source and others are commercial. For commercial applications, users may either start a private license server at Gompute or enable a connection to their local license server. Users get real time information about their consumption for each job, thus enabling them to optimize their compute spends. Customers include Westinghouse Sweden, SAAB Automobile, ABB, and Sandvik.

---

[9] Rocky Mountain HPC center brings supercomputing to the people, Jon Brodkin, Network World, 19 Nov 2009, http://www.networkworld.com/news/2009/111909-nonprofit-supercomputing.html
[10] www.gompute.com

Why IBM: Gridcore can help client's establish private HPC clouds based on IBM Systems, storage and middleware integrated with Gompute, optimized for performance, ease of use and team collaboration. With its built-in portal, accounting and billing tools, and Gompute platform as a service cloud, clients can scale their HPC infrastructure as required, paying only for incremental resources as required. The Enterprise Gompute software stack and service are made available for customers in computer aided engineering (CAE) and other HPC scientific and technical domains. This enables Gompute to offer their customers access to larger HPC infrastructure including IBM BladeCenter, System x, System p, Infiniband, scalable IBM System Storage, and remote visualization to help improve insight and optimize performance.

SciNet –*Canada's most powerful supercomputer - a private cloud for global research collaboration*

SciNet[11], a consortium of the Canadian university and government bodies, has a mandate to provide high-performance computing resources to their own academic researchers as well as other users across the country and international collaborations. Together with Compute Canada and IBM[12], SciNet has collaborated to create Canada's most powerful supercomputer and one of the most powerful and energy-efficient supercomputers in the world. The facility enhances SciNet's competitive position in globally important research projects.

How it works: With peak performance of more than 300 trillion calculations per second, the IBM System x iDataPlex system ranked 22nd in the Nov 2009 TOP500 list of world's most powerful supercomputers. It uses a total of 30,240 Intel processor 5500 series 2.53 GHz processor cores and is entirely water cooled. Adaptive Computing Moab and xCAT are used to administer, manage, and schedule a wide range of HPC applications and workloads on this system. The IBM Supercomputer is being used for ground-breaking research in aerospace, astrophysics, bioinformatics, chemical physics, climate change prediction, medical imaging and the global ATLAS project, which is investigating the forces that govern the universe.

Why IBM: The IBM System x iDataPlex server is specifically designed for data centers that require high performance, yet are constrained by floor space, power and cooling infrastructure. This system provides up to five times the compute density versus competitive offerings and a unique water cooled technology -- IBM's Rear Door Heat Exchanger -- extracts more heat than the systems actually generate. This new iDataPlex system adds to SciNet's existing supercomputing capability, which includes an IBM water cooled Power 575 supercomputer with 3,328 POWER6 cores with peak performance of more than 60 trillion calculations per second. This, combined with additional energy efficiency technologies, including dynamic provisioning software in xCAT that automatically turns off processors not currently in use, and the state-of-the-art data center design at the University of Toronto saves enough energy to power more than 700 homes yearly.

Large Financial Services Customer - *cutting costs and streamlining infrastructure with IBM iDataPlex*

A large financial services customer in the US faced to address several challenges – internal test and certification requirements were increasing rapidly, disk storage requirements needed to be stateless, running costs for file systems and applications were rising, and applications workload needed to be streamlined across the organization.

How it works: The customer implemented an IBM cloud solution comprising 84 IBM iDataPlex servers configured to be used in virtualization mode and connected through 10 GB Ethernet. The software to manage this configuration included Moab for workload management, xCAT for provisioning, VMWare for virtualization and Linux OS.

Why IBM: With IBM, the customer obtained a comprehensive rack-based solution comprising compute power, storage and networking based on open standards and IBM cloud implementation services. As a result, the customer now runs 50% to 95% of the bank's applications on an internal private cloud resulting in an improvement in system utilization from about 20% to 90%. Further, the cloud solution eliminates a large portion of individual test and certification requirements, saves on the cost of redundant hardware through virtualization and greatly reduces disk storage space through stateless technology. This has resulted in a highly streamlined IT infrastructure.

---

[11] www.scinet.utoronto.ca

[12] http://www-03.ibm.com/press/us/en/pressrelease/27755.wss

NASA Center for Computational Sciences – *enhancing high-resolution climate modeling on the Discover supercomputer cloud with the IBM iDataPlex*

The NASA Center for Computational Sciences (NCCS) based at the Goddard Space Flight Center in Maryland, is one of two organizations that support the NASA High-End Computing (HEC) Program. The HEC Program provides more than 77,000 computer processors and peak processing power greater than 800 teraflops to the NASA user community. Enhanced computational capabilities allow NASA climate scientists to run high-resolution simulations that reproduce atmospheric features not previously seen in their models, supporting vital global research.

<u>How it works:</u> Aiming to keep pushing the envelope and create ever more detailed climate simulations, the NCCS has added more than 1,000 IBM iDataPlex servers to its Discover cluster, providing a total of 8,256 new cores of computational power and over 24 TB of high speed memory.

<u>Why IBM:</u> The NCCS generates vast quantities of data from climate and weather models. The IBM iDataPlex solution brings together sensor data from numerous sources and numerical models to produce sophisticated and highly detailed climate simulations. The new computational capabilities enable NASA climate scientists to create more detailed simulations and reproduce atmospheric features previously not visible in their models.

Conclusion

High performance computing helps academic and enterprise users achieve the speed, agility, and insights to lead the market. But handling, analyzing, and visualizing the exploding datasets resulting from HPC is straining data center capacity. Pressure on IT budgets, the need to improve utilization and return on compute dollars, and escalating energy and operational costs of building and maintaining data centers are forcing enterprises to adopt cloud computing models. HPC users are thus naturally drawn to cloud computing with its promise of pay-as-you-go supercomputing.

But apart from the traditional corporate concerns of data security of computing over a cloud, the main problem with running HPC tasks on conventional clouds is that conventional clouds are geared toward supporting general-purpose applications and services -- short transactional workloads such as Web applications and database tasks. HPC tasks, on the other hand, are mostly complex, long-running algorithms processed in parallel, with the result of one task not dependent on the outcome of another. Processing threads are brought together at the end of the activity.

Today, most HPC Clouds today run on bare metal (i.e. without a virtualization layer), often with high-performance interconnects to maximize absolute performance. However, in the future, we envision HPC clouds to be powered by computational clusters with software optimized for performance, flexibility, reliability, and availability via a thin light-weight virtualization layer. [13]

IBM, with its long history of enterprise computing, addresses these security concerns and the specific computing needs of the HPC community through server (x86 and Power), storage, and software (IBM and partner) that is custom built for HPC requirements. These solutions are delivered over public, private, and hybrid clouds and implemented at many clients worldwide.

[13] Arun Babu Nagarajan, et. al., "Proactive Fault Tolerance for HPC with Xen Virtualization", ICS'07, June 18-20, 2007.