International Journal of Latest Trends in Engineering and Technology
IJLTET Special Issue- ICRACSC-2016 , pp.012-017
e-ISSN:2278-621X

CLOUD HIGH PERFORMANCE COMPUTING AND PERFORMANCE PREDICTION

Md.Rafeeq[1], Dr.C. Sunil Kumar[2], Dr.N. Subhash Chandra[3]

Abstract: The point of having a high performance computer is so that the individual nodes can work together to solve a problem larger than any one computer can easily solve. HPC users are accustomed to managing directly very complex parallel systems and performing a very fine-tuning of their applications on the target hardware. The matter is to ascertain if it may be convenient to deploy such applications on a cloud, where users "voluntarily" lose almost all control on the execution environment, leaving the management of datacenters to the cloud owner. cloud computing may be exploited at three different levels: IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and AaaS (Application as a Service). In one way or another, all of them can be useful for HPC[2].
Keywords: ***Virtual machine(VM) , Performance, Auto-Scaling, Server Virtualization.***

I. INTRODUCTION

" High Performance Computing most generally refers to the practice of aggregating computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science, engineering, or business**"**. A helpful way to help understand what high performance computers are is to think about the what's in them. You have all of the elements you'd find on your desktop — processors, memory, disk, operating system — just more of them. High performance computers of interest to small and medium-sized businesses today are really clusters of computers.

Each individual computer in a commonly configured small cluster has between one and four processors, and today's processors typically have from two to four cores. HPC people often refer to the individual computers in a cluster as nodes. A cluster of interest to a small business could have as few as four nodes, or 16 cores. A common cluster size in many businesses is between 16 and 64 nodes, or from 64 to 256 cores. The point of having a high performance computer is so that the individual nodes can work together to solve a problem larger than any one computer can easily solve. Clouds in conjunction with traditional HPC and HTC grids provide a balanced infrastructure supporting scale-out and scale-up, as well as capability (HPC) and quick turn-around (HTC) computing for a range of application (model) sizes and requirements[10]. The novelty in resource management and capacity planning capabilities is likely to influence changes in the usage mode, as well deployment and execution management/planning. The ability to exploit these attributes could lead to applications with new and interesting usage modes and dynamic execution on clouds.

MPI (**Message Passing Interface[5]** and 'embarrassingly parallel' problems. I'm using terms I heard at the conference, but will use MPI and EPP (embarrassingly parallel problem) MPI is essentially a programming paradigm that allows for taking extremely large sets of data and crunching the information in parallel WHILE sharing the data between compute nodes. Some times this is also referred to as 'clustering', although that term is frequently overloaded today. Most of them use Infiniband technology for example to effectively turn the entire

[1] *Dept. of CSE, CMR Technical Campus-Hyd, TS, India*
[2] *Dept. of IT,SNIST-Ghatkesar, Hyd, TS, India*
[3] *HITS-Ghatkesar, Hyd, TS, India*

grid into a single '**supercomputer**'. In fact, most of these MPI-based grids are ranked into the Supercomputer **Top500[3]**. EPP workloads, by contrast, have no data sharing requirements. A very large dataset is chopped into pieces, distributed to a large pool of workers, and then the data is brought back and reassembled.

HPC vs. HSC: The reality is that High Scalability Computing is ideal for the majority of EPP grid workloads[8]. In fact, large amounts of this kind of work, in the form of MapReduce jobs have been running on Amazon EC2 since its beginning and have driven much of its growth. HPC is a different beast altogether as many of the MPI workloads require very low latency and servers with individually high performance. It turns out however, that all MPI workloads are not the same. The lower bottom of the top part of that pyramid is filled with MPI workloads that require a great network, but not an Infiniband network*[10]*. In keeping with Amazon Web Service's tendency to build out using commodity (cloud) techniques, their new HPC offering does not use Infiniband, but instead opts for 10Gig Ethernet. This makes the network great, but not awesome and allows them to create a cloud service tailored for many HPC jobs. In fact, this **recent benchmark posting** by CycleComputing shows that AWS' Cloud HPC system has impressive performance particularly for many MPI workloads.

II. GPU-ACCELERATED COMPUTING

GPU-accelerated computing is the use of a graphics processing unit (GPU) together with a CPU to accelerate deep learning, analytics, and engineering applications. Pioneered in 2007 by NVIDIA, GPU accelerators now power energy-efficient data centers in government labs, universities, enterprises, and small-and-medium businesses around the world. They play a huge role in accelerating applications in platforms ranging from artificial intelligence to cars, drones, and robots[8].

2.1 How GPUs Accelerate Software Applications GPU-accelerated computing offloads compute-intensive portions of the application to the GPU, while the remainder of the code still runs on the CPU. From a user's perspective, applications simply run much faster.

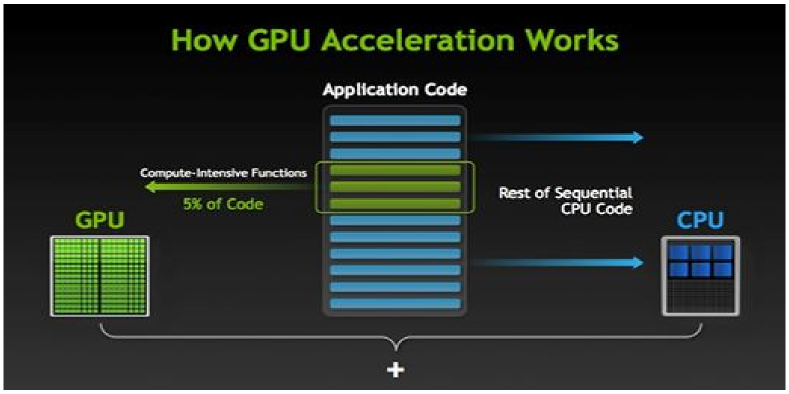

Figure 1: GPU acceleration works

2.2Memory Modes For Increased Performance on Intel Xeon Phi:

The Intel Xeon Phi processor supports different types of memory, and can organize this into three types of memory mode. The new processor from Intel contains two type of memory, MCDRAM and DDR memory. These different memory subsystems are complimentary but can be used in different ways, depending on the application that is being executed. The first, and most familiar is to use the MCDRAM memory as a cache for the DDR memory. The MCDRAM cache in this case is managed by the hardware. In this mode, legacy applications will work just fine and can benefit from the high memory bandwidth of MCDRAM. In the basic case, if information is not found in the MCDRAM, a request is sent to the DDR memory[9]. Another mode is what is called, Flat mode. In this case, the MCDRAM looks like DDR memory and is presented as such. The MCDRAM memory is mapped into the same address space as the DDR memory and acts the same in terms of reading and writing.

2.3 HPC Software: Just like your desktop or laptop, your HPC cluster won't run without software. Two of the most popular choices in HPC are Linux (in all the many varieties) and Windows. Linux currently dominates HPC installations, but this in part due to HPC's legacy in supercomputing, large scale machines, and Unix. Your choice of operating system should really be driven by the kinds of applications you need to run on your high performance computer. If you are using Excel to run option calculations in parallel, you'll want a Windows-based cluster, and so on[1]. In fact the first thing to know when you are considering buying or building an HPC cluster is what you want to do with it.

III. PERFORMANCE PENALTIES

The main question related to the adoption of the cloud paradigm in HPC is related to the evaluation (and, possibly, to the reduction) of possible performance losses compared to physical HPC hardware. In clouds, performance penalties may appear at two different levels:

- **Virtual Engine (VE**). These are related to the performance loss introduced by the virtualization mechanism. They are strictly related to the VE technology adopted.
- **Cloud Environment (CE)**. These are the losses introduced at a higher level by the cloud environment, and they are mainly due to overheads and to the sharing of computing and communication resources.

The actual hardware used in the cloud, along with the losses at the VE and CE levels, will determine the actual performance of applications running in the cloud. As will be discussed later, for HPC users the final perceived performance will be not so much affected by VE and CE levels as by the class of the physical hardware (computing and interconnect) making up the cloud. Even if the computing nodes adopted in cloud are not too different from those making up (economical) HPC clusters, it is a fact that these usually adopt suitable network switches, like Myrinet or Infiniband, which provide high band width and low latency[4].……………………………………………………………….

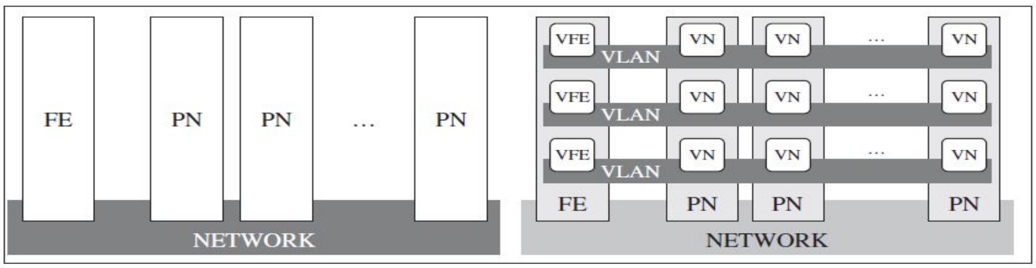

Figure: 2 Physical and virtual cluster.

The configuration and performance analysis of virtual clusters poses problems that are considerably more complex than those involved in the use of physical clusters. The objective of this section is to present the main problems and to introduce a clear and sound terminology, which is still lacking in the literature.

It is essentially made up of a front-end (typically used only for administration purposes, often the only node with a public IP address) and a number of (physical) processing nodes. These are, turn, provided with a single CPU or with multiple CPUs sharing a common memory and I/O resources. The multiple CPUs may be multiple cores on a single processor chip, a traditional single-core CPUs working in SMP mode, a “fictitious” CPU obtained by Hyper threading, or a mixture of all the above. A physical cluster can execute multiple jobs in parallel, by assigning to every job a subset of the total number of CPUs. Usually the choice is to use non overlapping subsets of CPUs, in order to avoid processor sharing among multiple jobs. But, even doing so, the interconnection network (and the frontend) are inevitably shared[7]. Every virtual processing node can host one or several virtual machines(VMs), each running a privateOSinstance. These may belong to the same or to different virtual clusters. At least in theory, the number of VMs is limited only by resource consumption (typically, physical memory). In turn, each VMis provided with several virtual CPUs (VCPUs). A virtual machine manager running in every node makes it possible to share the physical CPUs among the VCPUs defined on the node (which may belong to a single virtual cluster or to several virtual clusters). Typically, it is possible to define VCPUaffinity and to force every VCPU to run on a subset of the physical CPUs available.

It is worth noting that, given a physical node provided with n CPUs, there are two possibilities to exploit all the computing resources available:

- **Using** n VMs (each running its OS instance) with one, or even several, VCPUs;
- **Using** a single VM with at least n VCPUs.

On the other hand, the use in a node of v VCPUs, with v.n, whether in a single or in multiple VMs, leads to a fictitious multiplication of computing resources. In nodes where CPU resources are multiplied, the virtual clusters not only share memory, communication hardware, and the virtual machine manager, but also share CPU cycles, with a more direct effect on overall computing performance[6].

IV. HPC on CLOUD

Depending on the user’s application domain, these two feature sets can make a big difference in performance. For example, applications that require a single node (or threaded applications) can work in a cloud. In this case, the user might have a single program that must be run with a wide range of input parameters (often called parametric processing), or they might have dataflow jobs, such as the Galaxy suite used in biomedical research. These types of applications can benefit from most cloud computing resources. Some applications can utilize highly parallel systems but do not require a high-performance interconnect or fast storage. One often cited example is digital rendering, in which many non-interacting jobs can be spawned across a large number of nodes with almost perfect scalability. These applications often work well with standard Ethernet and do not require a specialized interconnect for high performance. For example: Nimbix is a pure **high performance computing** (HPC) cloud built for volume, speed and simplicity[4].

HPC:R-HPC can provide new 3.4GHz quad core Sandy Bridge-based systems with 16GB of RAM/node (4GB/core), DDR 2:1 blocking InfiniBand, and 1TB of local disk. Additionally, they have dual-socket 2.6GHz eight-core Sandy Bridge with 128GB of RAM/node (8GB/core), QDR non-blocking InfiniBand, 1TB of local storage, and 1TB global storage. These offerings are rounded out by Magny-Cours, Nehalem, and Harpertown systems. GPU-based systems in beta test are provided for dedicated users[6].

Amazon EC2 HPC: Perhaps the most well-known cloud provider is Amazon. Inquiries to Amazon were not returned, so information was gleaned from their web page. Originally, the EC2 service was found not suitable for many HPC applications. Amazon has since created dedicated “cluster instances” that offer better performance to HPC users. Several virtualized HPC instances are available on the basis of users’ needs. Their first offering is two Cluster Compute instances that provide a very large amount of CPU coupled with increased network performance (10GigE). Instances come in two sizes, a Nehalem-based “Quadruple Extra Large Instance” (eight cores/node, 23GB of RAM, 1.7TB of local storage) and a Sandy Bridge-based “Eight Extra Large Instance” (16 cores/node, 60.5GB of RAM, 3.4TB of local storage).

Univa: Univa is known as the place Sun Grid Engine landed after the Oracle purchase. Not only does Univa continue to develop and support Grid Engine, they also offer “one-click” HPC computing with their partner RightScale. The companies offer two distinct ways to build clouds. The first is an internal cloud infrastructure system that provides “raw” virtual machines using one of several methods, including VMware, KVM, Xen, or something more sophisticated like OpenStack or VMware vSphere[3].

Sabalcore Computing : Sabalcore offers HPC systems over the Internet for both on-demand and dedicated solutions. On-demand access is accomplished using an SSH client or secure remote desktop interface from any Windows or Linux laptop, desktop, or workstation. Users are provided with an NFS-exported home directory with persistent storage for results, code, applications, and data.

V. HPC SERVICE’s AND RESULT’s

When projects have critical deadlines and exacting requirements, Sabalcore can provide dedicated High Performance Computing (HPC) environments. Sabalcore’s Dedicated Solutions are for those wanting the convenience of the cloud with a dedicated HPC environment with customized support and service. We will carefully develop a solution tailored to your organization’s specific needs that allows for maximum price-to-performance.

Dedicated Compute Capacity

Sabalcore Dedicated HPC systems are physically isolated. All hardware resides in Sabalcore’s data centers ensuring complete confidentiality.

Complete Solution :

Focus on your core business knowing your HPC infrastructure, administration, and technical support needs are covered. The system is dedicated to your users 24 hours a day, 7 days a week. Sabalcore fully supports the entire system including hardware, operating system, and software. The system is “turn-key” and integrated with your organization’s processes and workflows[3].

Key Advantages of Sabalcore Dedicated HPC CLoud

▪ . System is built, maintained, and supported by Sabalcore .Tailored and configured to meet your company’s requirements

- Production level service
- Custom application and workflow integration
- Scale as needs grow or seamlessly push workloads to Sabalcore’s On Demand service

- Leading hardware and software technology
- Secure, Encrypted, Confidential, and Private.

BEST-OF-BREED HPC SYSTEM

Sabalcore Dedicate HPC Clusters are designed around your application and workflows. We work closely with you during planning and design to come up with the optimal system and service plan. Leverage Sabalcore's years of experience in scientific and high performance computing along with the latest hardware technologies. System options include:

- The newest Intel™ Xeon® processors
- Cutting-edge network technology with the latest Infiniband

Comparison of Cost Optimized vs Performance Optimized for High Value Compute					
		Cost Optimized System - Application Deployment & 3-year Maintenance		Performance Optimized - Application Deployment & 3-year Maintenance	
Equipment		Total Cores		Total Cores	
Equipment		440	$509,600	200	$980,000
3-year maintenance (12%/Year)			$152,880		$352,800
Equipment & Maintenance	Total		$662,480		$1,332,800
Software Costs		Cores		Virtual DB Cores	
Software Licensing	Average Databases & SAS ($24,900/Core)	88	$2,191,200	40	$996,000
Software Maintenance	3yr @ 22%/year list price	88	$2,892,384	40	$1,314,720
Red Hat Virtualized Linux Subscription	3yr @ $2,499/year/cluster list price		$149,940		
Licensing & Maintenance	Total		$5,233,524		$2,460,660
Operational Support		Hours/month		Hours/month	
Database Administration	$200/hour	90	$648,000	54	$388,800
System Administration	$140/hour	48	$241,920	27	$136,080
Operational Support	Total		$889,920		$524,880
			$6,785,924		**$4,318,340**

Table 1: Detailed Comparison of Costs and Differences between Cost Optimized and Performance Optimized Infrastructure for High Value Compute Infrastructure.

- High-capacity high-performance parallel storage systems
- Managed head nodes, web servers, and VPN servers
- Secure VPN (soft or hardwired) access
- Tailored security and Data encryption policies
- Software application workflow optimization
- Programming, scripting, and consulting[7].

VI. CONCLUSION

Given the varied requirements for HPC clouds, it is understandable that the range of options can vary greatly. Solutions range from shared remote clusters to full virtualized systems in the cloud. Each method brings its own feature set that must be matched to the users' needs. Finally, the above items are not intended to be an exhaustive list of HPC cloud providers. Others exist, and given that the market is new and growing, more vendors will be coming online in the near future. Many other factors should also be considered besides the brief analysis offered here. Your best results will come from doing due diligence and testing your assumptions. Perhaps the most important aspect of cloud HPC is the ability to work with your vendor, because having a good working safety net under your cloud might be the best strategy of all.

ACKNOWLEDGEMENT: I , would like to express my sincere thanks to **Dr. C. Sunil Kumar Supervisor** and **Dr. N. Subhash Chandra Co-Supervisor** for constant help and guidance in PhD work and also I thank to my Organization CMRTC for encouragement**.**

REFERENCES

[1] WIKIPEDIA, "High-performance computing",
[2] Cloud Computing Denition, National Insitute of Standards and Technology, Version 15,
[3] Jaliya Ekanayake and Geoffrey Fox, "High Performance Parallel Computing with Cl ouds and Cloud Technologies",
[4] ASF. 2009. Apache Hadoop Core. http://hadoop.apache.org/core.
[5] ASF. 2009. Apache Hadoop Pig . http://hadoop.apache.org/pig/.
[6] C. Vecchiola, S. Pandey, and R. Buyya, "High-performance cloud computing: A view of scientific applications," CoRR, vol. abs/0910.1979, 2009.

[7] Amazon Elastic Compute Cloud (EC2), http://aws.amazon.com/ec2/

[8] Amazon.com, Inc. 2009. Simp le Storage Service (S3). http://aws.amazon.com/s3.

[9] Thomas Sterling, Dylan St ark, A High-Performance Computing Forecast: Partly Cloudy, Computing in Science & Engineering, July/August 2009, pp.42-49.

[10]Rajkumar Buyya, Chee Shin Yeo, and Srikumar Venugopal. Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. In High Perfor- mance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on , pages 5{13, 2008}.