Bridging the Gap Between HPC and IaaS

Andrew J. Younge and Geoffrey C. Fox
Pervasive Technology Institute, Indiana University
2729 E 10th St., Bloomington, IN 47408, U.S.A.
Email: {ajyounge,gcf}@indiana.edu

***Abstract*—With the advent of virtualization and Infrastructure-as-a-Service (IaaS), the broader scientific computing community is considering the use of clouds for their technical computing needs. This is due to the relative scalability, ease of use, advanced user environment customization abilities, and the many novel computing paradigms available for data-intensive applications. However, there is still a notable gap that exists between the performance of IaaS when compared to typical high performance computing (HPC) resources, limiting the applicability of IaaS for many potential users.**

This work proposes to bridge the gap between supercomputing and clouds, specifically by enabling both tightly coupled applications and distributed data-intensive applications under a single, unified framework. Through heterogeneous hardware, performance-tuned virtualization technologies, advanced I/O interconnects, and unique open-source IaaS software, a distributed, high performance cloud computing architecture is formed.

I. Introduction

Scientific computing endeavours have created clusters, grids, and supercomputers as high performance computing (HPC) platforms and paradigms, which are capable of tackling non-trivial parallel problems. HPC resources continually strive for the best possible computational performance on the cusp of Moore's Law. This pursuit can be seen through the focus on cutting edge architectures, high-speed, low-latency interconnects, parallel languages, and dynamic libraries, all tuned to maximize computational efficiency. Performance has been the keystone of HPC since its conception, with many ways to evaluate performance. The Top500 supercomputer ranking system, which has been running for over 20 years, is a hallmark to performance's importance in supercomputing.

Many industry leaders have focused on leveraging the economies of scale from data center operations and advanced virtualization technologies to service two classes of problems: handling millions of user interactions concurrently or organizing, cataloging, and retrieving mountains of data in short order. The result of these efforts has led to the advent of cloud computing, which leverages data center operations, virtualization, and a unified and user-friendly interface to interact with computational resources. This is culminated in the *aaS mentality, where everything is delivered as a service. This model treats both data and compute resources as a commodity and places an immediate and well defined value on the cost of using large resources [1]. Users are able to scale their needs from a single small compute instance to thousands or more instances aggregated together in a single data center. Clouds also provide access to complex parallel resources through simple interfaces, which have enabled a new class of internet applications and tools ranging from social networking to cataloging the world's knowledge. Unfortunately, the ease of use with clouds often have implicit performance impacts that, up until now, have been overlooked.

II. Performance

As with the HPC industry, performance must be a first class function of the new cloud architecture. In the Magellan Project [2], one of the major obstacles identified with virtualized environments is the performance gap when compared to supercomputers. While virtualization can induce overhead to the computation, clouds or virtualization do not intrinsically impose any theoretical limits to performance by design. Rather, the implementation of such technologies leads to the bottlenecks and limitations identified by the Magellan project. Like many new technologies some overhead may be unavoidable, but the goal is to minimize the overhead whenever possible and, in time, consider virtualization's overhead in the same light as we do with language compiler overhead today.

A detailed analysis of the overhead of various virtualization technologies has already commenced [3]. From a computational perspective, the amount of overhead introduced by virtualization needs to be minimized or eliminated entirely. This applies not only to typical floating-point operations common in MPI applications and represented in the Top500 [4], [5], but also with data-intensive calculations as embodied by the Graph500 benchmark [6]. Using HPC benchmarks provide an ideal opportunity to model applications on real hardware implementations. These and other benchmarks can perform well in IaaS when based on a design backed by experimentally-verified practices.

However, CPU and memory utilization is only one aspect of supercomputing application performance. Another key focal point is the I/O bandwidth and interconnect performance. Advanced networking technologies have been relatively unavailable in clouds, which often only support Gigabit Ethernet (the only notable exception is the use of 10GbE in Amazon Cluster Compute Instances). InfiniBand and MMP interconnects, the backbone of the HPC industry, have largely gone unused in clouds. Work on the Palacios project has looked at supporting 10GbE networks in the VMM with notable success [7]. As such, a comprehensive system for connecting and interweaving virtual machines though a high speed interconnects is paramount to the success of a high performance IaaS architecture.

Two technologies have become available to enable virtualzied environments to leverage the same interconnects as many supercomputers; hardware-assisted I/O virtualization (VT-d or IOMMU) and Single Root I/O Virtualization (SR-IOV). With VT-d, PCI-based hardware passed directly to a guest VM, thereby removing the overhead of communicating with the host OS through emulated drivers. As a result, VMs gain direct access to InfiniBand adapters and use drivers without emulation or modification to achieve near-native I/O performance. Second, with the use of SR-IOV, adapter functions can be assigned directly to a given VM. This enables IaaS providers to leverage InfiniBand interconnects for applications that utlize RDMA, IB Verbs, or even IP while simultanously providing a guaranteed QoS based on SR-IOV configuration.

III. Heterogeneity

The heterogeneous nature of the HPC industry has existed since its birth, with a number of unique architectures that enable novel applications within scientific computing [8]. However, a one-size-fits-all supercomputer or cloud that exists to serve the needs of all is unlikely to exist. This can be noted by the current state of the XSEDE project (formally TeraGrid), which hosts almost a dozen different supercomputers that are not directly compatible. While we do not hope to adapt a one-size-fits-all model to HPC, we instead propose to provide these heterogeneous resources through the same unified interface provided by current IaaS. Specifically, supporting a wide array of hardware such as accelerator cards, GPUs, ARM & x86 CPU architectures, and networking resources is now becoming possible through the use of hardware-assisted virtualization or bare-metal provisioning. As a proof of concept, enabling nVidia Tesla GPUs provides an ideal use case, especially as many supercomputing resources look towards utilizing GPU architectures as the community moves through petascale and towards exascale computing.

Providing heterogeneity in IaaS deployments is akin to many Grid computing solutions which saw a swath of meta-schedulers and workflow models to take advantage of heterogeneous resources in an imperfect user environment [9]. Applying similar abstract models of resource federation from Grids while leveraging advanced virtualization technologies and simplified operating environments enables the use of heterogeneous, high performance resources in a system conducive to researchers focusing on the scientific challenges rather than resource specifics.

IV. Data Deluge

There are an increasing number of problems that demand a data-centric model for enabling science in the newfound data deluge of the 21st century [10]. Tackling these problems requires the integration of large scale, high performance storage systems that can span large numbers of resources in a way that integrates naturally with computation. Traditionally, the needs of computational power have outweighed the need for large scale data storage and access. While that may still be the case for many applications, there is an ever-growing array of scientific problems that see data as the largest obstacle to accelerating their research. It is important to instead move computation to the data, a function that virtualization itself enables with the intrinsic ability of live-migration of VMs within cloud infrastructure. Leveraging this potential function in order to increase the performance and ability of the next generation of scientific problems will be key to the success of a unified HPC Cloud.

V. Direction

Currently implementation and experimentation is under way to evaluate the applicability of a high performance IaaS architecture. Research is currently leveraging the OpenStack IaaS framework, as it represents an open-source, scalable, community driven software stack with a wide consortium of users [11]. Utilizing the Xen hypervisor to obtain near-native performance of supercomputing resources with experimentally-validated deployments and hardware-assisted virtualization, we hypothesize VM performance overhead becoming negligible. Furthermore, we plan to leverage advanced hardware such as InfiniBand and nVidia GPUs available within today's supercomputing resources within an OpenStack IaaS. The FutureGrid project [12], a distributed grid and cloud testbed, provides an idea test-bed to build an experimental IaaS based on real-world resources, eliminating the need for simulation or emulation.

While using supercomputing benchmarks provides a way to evaluate experimental computational resources, we will also look for novel scientific applications in Bioinformatics to provide *in-situ* evaluation of the proposed architecture. Looking at large scale analysis of newly sequenced and divergent genomes and applying computationally intensive analysis techniques provides an ideal use case to evaluate a distributed high performance IaaS system. A successful evaluation could lead to substantial impacts in Bioinfromatics and many other fields looking to better utilize advanced computational resources.

References

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–58, Apr. 2010.

[2] K. Yelick, S. Coghlan, B. Draney, and R. S. Canon, "The magellan report on cloud computing for science," U.S. Department of Energy, Tech. Rep., 2011.

[3] A. J. Younge, R. Henschel, J. T. Brown, G. von Laszewski, J. Qiu, and G. C. Fox, "Analysis of Virtualization Technologies for High Performance Computing Environments," in *Proceedings of the 4th International Conference on Cloud Computing (CLOUD 2011)*. Washington, DC: IEEE, July 2011.

[4] P. Luszczek, D. Bailey, J. Dongarra, J. Kepner, R. Lucas, R. Rabenseifner, and D. Takahashi, "The hpc challenge (hpcc) benchmark suite," in *Proceedings of the 2006 ACM/IEEE conference on Supercomputing*. Citeseer, 2006, pp. 11–17.

[5] H. Meuer, "The top500 project: Looking back over 15 years of supercomputing experience," *Informatik-Spektrum*, vol. 31, no. 3, pp. 203–222, 2008.

[6] M. Anderson, "Better benchmarking for supercomputers," *Spectrum, IEEE*, vol. 48, no. 1, pp. 12–14, 2011.

[7] L. Xia, Z. Cui, J. R. Lange, Y. Tang, P. A. Dinda, and P. G. Bridges, "Vnet/p: bridging the cloud and high performance computing through fast overlay networking," in *Proceedings of the 21st international symposium on High-Performance Parallel and Distributed Computing*, ser. HPDC 12. New York, NY, USA: ACM, 2012, pp. 259–270.

[8] S. Graham, M. Snir, and C. Patterson, *Getting up to speed: The future of supercomputing*. National Academies Press, 2005.

[9] I. Foster, Y. Zhao, I. Raicu, and S. Lu, "Cloud Computing and Grid Computing 360-Degree Compared," in *Grid Computing Environments Workshop. (GCE'08)*, 2008, pp. 1–10.

[10] A. Hey, S. Tansley, and K. Tolle, *The fourth paradigm: data-intensive scientific discovery*. Microsoft Research Redmond, WA, 2009.

[11] K. Pepple, "Revisiting openstack architecture: Essex edition," Webpage, Feb 2012. [Online]. Available: http://ken.pepple.info/openstack/2012/02/21/revisit-openstack-architecture-diablo/

[12] "FutureGrid," Webpage. [Online]. Available: http://portal.futuregrid.org