



# High Performance Computing in the Cloud

Preparing for the Inevitable

A Platform Computing Technology White Paper  
November 2010

## Table of Contents

|   |   |
|---|---|
| 1. Introduction                                 | 3 |
| 2. The Cloud: Yesterday and Today               | 4 |
| 3. HPC Applications Differences from Enterprise | 5 |
| 4. HPC in the Cloud                             | 6 |
| 5. Conclusions                                  | 7 |

# 1.Introduction

When did the cloud become a resource for high performance computing? Probably long before you even knew it did. You may find your high-performance applications are currently running in a cloud-like infrastructure and, with a few changes, you can take full advantage of the power of the cloud.

High-Performance Computing (HPC) has a long tradition of using dedicated, homogeneous, and fast resources connected via an extremely high speed network. Therefore, many HPC users don't believe that cloud computing can be used as an HPC resource.

Why? Because implementations of cloud computing are generally a loose set of commodity servers in an infrastructure that is not designed for speed. Furthermore, cloud computing often utilizes virtualization to enable better flexibility and higher utilization at the expense of application performance. Another factor that further reduces application performance is that there are not customized storage facilities that are tuned for HPC workloads. A typical cloud infrastructure is designed for transactional processing, and does not provide the high bandwidth, low latency connections or extremely low hop counts needed for high performance computing..

But clock speeds have stalled as processors have reached their limit on speed. To compensate, the number of cores in one compute node are growing and applications are increasingly parallel. Applications must run faster for organizations to keep up competitively. As HPC looks for alternatives to keep and grow their competitive advantage, the public cloud is becoming both technically and economically feasible.

This white paper discusses the history and market forces that are driving the growth of the cloud, the difference between enterprise cloud infrastructures and cloud infrastructures designed for high-performance compute environments, and how an HPC cloud environment will ultimately work for your organization.

## 2.The Cloud: Yesterday and Today

### Proliferation of the Cloud

What has led to the proliferation of the cloud?

Pressure to move workload to cloud resources comes from two sources: the users and the resources.

The users have been driving a grass roots movement to cloud based resources. Often frustrated with central IT departments who control access to compute resources, users found that either the resources were being used by another department when they needed them or they were subject to complicated processes to access central IT assets. Looking for options for more control over their compute needs, they turned to the departmental mini-computer. This option was short-lived. Typically their application outgrew the power of the mini-computer or sometimes central IT would just absorb the asset into their management domain. Users then looked to the cloud to regain control over their workflow and have access to compute resources when they need them.

Central IT wants to meet user needs. The issue is oversubscribed compute resources and insufficient compute power to meet demand. Add to this the burst nature of HPC workloads where large projects require many compute cycles in a short amount of time. Organizations do not want to build for peak workloads where assets could sit idle during periods of low activity. The result – sufficient compute resources are not always available.

IT budgets and power/cooling issues in the data center drive the resource plan. Data center spending has significantly decreased in the past couple of years as organizations tighten their belt and control spending. Power and cooling costs must also be considered as data centers not only go greener, but are forced by their municipality to control their power consumption. Both these factors play a role in forcing IT Departments to consider cloud computing.

### Factors limiting applicability of HPC in the Cloud

For a long time, high performance computing operations did not need cloud based resources. IT departments could regularly count on Moore's law to predict the doubling of transistors every 18 months<sup>1</sup>, thus faster compute processors. Moore's law has stalled resulting in single core (thread) speeds at around 3GHz. To increase compute speed, transistors are being used to add more and more cores and threads into the same CPU socket. Expanding compute power by refreshing compute resources is no longer a viable option. HPC organizations must therefore purchase additional physical infrastructure or find a viable path to the cloud.

Another reason that cloud computing isn't a viable option for HPC is that HPC, applications are generally not written to take advantage of more than one thread of execution. This is changing as ISV's are creating parallelized versions of their applications. These parallelized versions take advantage of the multi-thread and multi-core design of current compute resources to initiate multiple job processes concurrently. This further reduces the dependency on a high bandwidth, low latency physical infrastructure.

Virtualization was once touted as the savior of IT. Virtualization enabled multiple applications to run on the same server independently in isolated virtual OS instances and move those OS's around the datacenter from physical server to physical server **without stopping the application**. However there was a price to pay, which primarily included performance penalties. This penalty was often acceptable for enterprise transactional applications, but it was deemed unacceptable for high performance computing. Because cloud computing uses virtualization to enable better flexibility and higher utilization, the low application performance in virtualized environments created a huge barrier for cloud adoption in HPC.

---

<sup>1</sup> This equates to the doubling of single threaded performance every 1.5 years.

However, virtualization technology has advanced in recent years and performance is becoming less of an issue. Processor support for virtualization as well as para-virtualized operating system device drivers have improved, removing the performance penalty once felt by the virtualized device driver layer. Many applications now have no more than a 3-6% performance penalty on a virtual machine as compared to directly running on the hardware.<sup>2</sup> Additionally programming environments, compilers, libraries, etc., are able to take advantage of the virtualized hardware architecture. As a result, some users have successfully run HPC applications in a virtualized environment.

### In Summary

The IT industry has reached a cost-per-compute inflection point via hosted grids and clouds which is changing how we think about HPC overall. The cloud is becoming a viable alternative for users who are demanding access to compute resources and organizations that are looking for ways to increase the return on investment of their assets. Performance penalties from virtualization are solved through improved programming environments. As more applications are parallelized, virtualized resources will become increasingly more viable. Public and private clouds bring technology assets into one pool, not only solving the utilization issue but removing the need to build for peak workloads, thus lowering infrastructure costs.

The roadblocks are down; HPC applications are moving to the cloud solving lack of resource and burst issues, organizations are lowering capital costs and improving the utilization of existing assets. Public and private clouds are proliferating as they are becoming a viable alternative for the organization's compute needs. So how do HPC applications differ from Enterprise Applications?

## 3.HPC Applications Differences from Enterprise Applications

Enterprise applications have been successfully used in the cloud for many years. By employing this infrastructure, these applications have been able to achieve better utilization rates on the hardware with some data center enterprise applications running at utilization rates around 20-40%<sup>3</sup>. This is good considering the unpredictable workloads from transactional workflows and short transaction jobs that are typical in enterprise environments.

HPC application workloads are more predictable resulting in cluster utilization rates in the 80-90% range. However, HPC workloads are often spiky in that the demand for additional compute resources can spike well above the running average demand and are dependent on project deadlines and needed application runs. When IT departments buy, build, and maintain clusters to handle peak loads it can be expensive, time consuming, and wasteful. Compute environments designed for peak loads often see utilization rates drop with idle compute resources when the project that created the spike is complete.

Data is a huge consideration and HPC applications generate a significant amount of data. How the application interfaces with the storage resource, whether streaming or high IOPS, can affect the performance of the system. The processes that handle data can slow down a system and cause an application to take more time to finish. Enterprise systems don't have the data generated from application runs and they have a simpler infrastructure for managing the data resulting from enterprise applications, therefore cloud infrastructures tuned for enterprise applications can quickly limit HPC applications.

<sup>2</sup>Platform Computing Whitepaper "Could the C in HPC stand for Cloud?"

<sup>3</sup>Tata centers that employ virtualization through post server consolidation have experienced higher utilization rates.

## Differences in Enterprise and HPC applications

| Characteristic                   | HPC                 | Enterprise       |
|----------------------------------|---------------------|------------------|
| Utilization rates                | 80-90% typically    | 20-40% typically |
| Workload type                    | Computational       | Transaction      |
| Workload queues                  | Often to continuous | None             |
| Unpredictable spikes in workload | Rarely              | Often            |
| Data growth rates                | Extremely high      | High             |

A well designed HPC environment can take advantage of a cloud infrastructure and achieve the performance, high utilization rates, and the ability to dynamically flex the size of the cluster to handle peak workloads.

Platform Computing helps improve utilization of HPC applications running in a cloud infrastructure. Our tools, which include Platform LSF for dynamic host capability, the Platform MultiCluster orchestrator, and Platform ISF for infrastructure sharing, facilitate the creation of a cloud environment that meets the demands of high performance computing applications. It is now possible to have the best of all worlds: performance, high utilization rates, and the ability to dynamically flex the size of the cluster to handle peak workload requirements.

## 4.HPC in the Cloud

It's all about performance. Every HPC user thinks about performance because the types of problems they are solving with their application often take too long to run in a non-distributed environment. For organization that use highly parallelized applications and have the compute resources available to run the parallel processes, job output time is significantly decreased. Applications run faster.

What if resources were temporarily available when they are really needed, like cloud resources? This is commonly referred to as cloud bursting.

Platform Computing offers three powerful and flexible HPC cloud solutions that enable you to to achieve cloud bursting from an HPC datacenter configuration.

- 1. Workload Scheduling  
Using Platform LSF dynamic host capabilities in a cloud environment, the cloud resources appear to be operating as an on-site HPC datacenter with local IP addresses and host names, etc. When a job is submitted, Platform LSF daemons built into the software allow the cloud resource to recognize the image request and prepare the resources for running the workload by adding the valid image. Workloads sent to the cloud can be refined by allowing the user to tag a job to ensure the application, data, and cloud resources are configured to run in the cloud.
- 2. Multi-cluster Management  
Using the Platform MultiCluster orchestrator solution with Platform LSF, users can start a new cluster with any cloud or hosting provider without a dedicated link. The cloud cluster becomes available and users can use MultiCluster to handle distributing the workload between the internal HPC cluster and the external "cloud" based cluster resources Policies driven by security, regulations, SLA/ SLO, etc must be considered.
- 3. Cloud Management  
Platform ISF has resource kits to handle different types of systems:
  - Physical or virtual
  - Linux/Unix/Windows
  - Virtual machines such as VMware, XEN, HyperV, KVM
  - Cloud resources such as Amazon, Rackspace

By using a combination of Platform ISF with Platform LSF, users can automatically flex up or down a cluster. Platform ISF manages both physical and virtualized resources in addition to the ability to scale out to other servers - either internal within the company infrastructure or external to a cloud hosting provider, cloud computing firm, or other type of service provider. The physical resources can also be automatically re-provisioned on demand to meet application needs such as specific operating system requirements.

Virtual machine instances are more flexible as they provide the ability to change CPU or memory resources, operating system, or even dynamically move the application to a server with more or less capabilities. ISF has a policy management capability for which each application can be setup specifically for the type of virtual resource required to get the job processed and meet SLA requirements. This can be defined independently based on the application and whether it can dynamically flex either internally, externally in the cloud or not at all.

## 5. Conclusions

Since clock speeds have stalled, applications are becoming more parallel to speed up job process time. A combination of cloud based resources and parallel applications add a level of complexity to workload and resource scheduling, but this is minimized with new technologies.

Cloud capabilities are considered and used on a regular basis in the life sciences, education, government, media/digital content creation and manufacturing arenas. Other industries are also starting to do testing and verification such as financial services, oil and gas, and EDA. This transition is happening now.

*Platform Computing is dedicated to enhancing and supporting cloud environments to help customers speed up the transition to the private and public cloud to help organizations achieve their business goals.*

Can high performance be achieved in a cloud? Just as in the enterprise datacenter where it is unlikely that all mission critical applications will be virtualized, the same is true in with HPC applications; not all HPC workloads and applications can run on virtual systems or in a hosted grid. However, as technology improves, HPC cloud usage will expand with the primary dependency on the ability to parallelize application workload to run on the cloud. Many companies today are now using one or more of the above method implementations. Their results are a powerful testament to achieve the stated goals: To be both technically feasible and economically beneficial.



Platform Computing is the leader in cluster, grid and cloud management software - serving more than 2,000 of the world's most demanding organizations for over 18 years. Our workload and resource management solutions deliver IT responsiveness and lower costs for enterprise and HPC applications. Platform has strategic relationships with Cray, Dell™, HP, IBM®, Intel®, Microsoft®, Red Hat®, and SAS®. Visit [www.platform.com](http://www.platform.com).

**World Headquarters**

Platform Computing Corporation  
3760 14th Avenue  
Markham, Ontario  
Canada L3R 3T7  
Tel: +1 905 948 8448  
Fax: +1 905 948 9975  
Toll-free Tel: 1 877 528 3676  
[info@platform.com](mailto:info@platform.com)

**Sales - Headquarters**

Toll-free Tel: 1 877 710 4477  
Tel: +1 905 948 8448

**North America**

New York: +1 212 888 6270  
San Jose: +1 408 392 4900

**Europe**

Bramley: +44 (0) 1256 883756  
London: +44 (0) 20 3206 1470  
Paris: +33 (0) 1 41 10 09 20  
Düsseldorf: +49 2102 61039 0  
[info-europe@platform.com](mailto:info-europe@platform.com)

**Asia-Pacific**

Beijing: +86 10 82276000  
Xi'an: +86 029 87607400  
[asia@platform.com](mailto:asia@platform.com)  
Tokyo: +81(0)3 6302 2901  
[info-japan@platform.com](mailto:info-japan@platform.com)  
Singapore: +65 6307 6590  
[wliaw@platform.com](mailto:wliaw@platform.com)

