

IBM HPC Management Suite for Cloud

Flexible HPC environment in the cloud

The IBM HPC Management Suite for Cloud is designed to help deploy high performance computing applications in a cloud environment. The comprehensive suite of tools can create the dynamic, flexible infrastructure your business needs today while supporting the performance required by compute-intensive workloads through bare metal dynamic provisioning. A web portal enables ease of access to the HPC environment for administrators and users.

The HPC Management Suite for Cloud allows organizations running the most demanding workloads and applications to take advantage of private cloud computing: application scalability, improved economies of scale, rapid infrastructure and solution deployment, and resource utilization and expandability.

You get high performance in a private cloud because the HPC Management Suite for Cloud gives you access to non-virtualized bare metal, as well as the ability to tailor virtual machines by the number of cores, disk space or the amount of real or virtual memory.

Acting as the resource manager, the HPC Management Suite for Cloud allows you to consolidate scattered and underutilized clusters, increase hardware utilization, and gain access to a larger cluster infrastructure in order to deploy HPC applications in a cloud environment.

The dynamic nature of compute-intensive workloads is accommodated by quickly allocating nodes to standby state or operational mode, according to requirements set by application, department, projects or other policies.

HPC on the cloud

An HPC cloud must provide a flexible runtime environment that will allow HPC applications to achieve petascale performance levels. A private cloud is closest to a conventional HPC cluster in terms of infrastructure ownership and security, but with the added benefits of resource pooling, a flexible infrastructure, reductions in capital expenditure and metered usage. The HPC Management Suite for Cloud provides a private HPC cloud using tools and technologies familiar to the HPC community.

Achieving high performance

While virtual machines may provide sufficient performance for typical applications, such as payroll, using the HPC Management Suite for Cloud, compute intensive workloads can have the power of bare metal at their disposal including access to all cores, memory and a parallel file system in order to complete execution in a timely manner.

Most HPC applications require network-accessible storage, preferably a parallel file system coupled with tape storage. The HPC Management Suite for Cloud will support network file systems (NFSS) and the IBM global parallel file systems (GPFS™) as a common file system for user directories, data staging area, and network storage location for diskless images.

Running on bare metal also helps to eliminate “neighbor noise” and the performance degradation typical when two virtual machines are running on the same physical machine. Even though they are in a cloud environment, the applications that require maximum performance get the resources necessary to complete the job when needed.

When the HPC Management Suite for Cloud is used to create virtual machines, it allows the administrator to set policies such as:

- How many virtual machines are allowed per physical machine.
- How many max CPUs are allowed per virtual machine.
- How much memory can be allocated to a virtual machine.

Whether provisioning for bare metal or virtual machines all applications get enough resources to do their job.

In order to run a highly optimized environment, HPC operations need an infrastructure that is flexible and dynamic. The HPC Management Suite for Cloud supports resource pooling and rapid application deployment capabilities including a reservation system, diskless provisioning, image management, automated cluster configuration and support for InfiniBand high-speed interconnects.

Machines in a traditional cluster can be provisioned and set up to become virtual private clusters running custom software stacks or even become smaller clusters, distinct from the parent. When the application user has finished using the virtual private cluster, the machines return to the parent cluster. The suite also enables elasticity. For example, an HPC cloud user may start with a few physical or virtual machines in their private cluster and then grow or shrink their cluster based on their compute needs.

User and administrator control for flexibility

The HPC Management Suite for Cloud provides on-demand self service to administrators and users through a web portal. The portal can also be used to set up administrative policies, such as power state management policy with the ability for manual overrides or the kind of jobs permissible on certain nodes.

The suite also offers a command line interface that ties into the web portal to do machine requests, usage reports, events, alerts, workload submission, monitoring and more.

While a portion of the infrastructure can remain in a traditional, production HPC cluster, the other portions can be deployed for other uses in virtual private clusters that require customized runtime environments.

Administrators can configure, monitor and administer policies for the installation, and users can log in to make reservations, provision physical or virtual machines, set up virtual clusters and share them with other users. Users have access to physical or virtual machines, a high performance parallel file system and multiple, high bandwidth interconnects. They can request nodes and provision them with the stack they need for their application along with the appropriate tools. When their jobs are done, the nodes are returned to the cloud.

Both administrators and users can access the following services using the HPC Management Suite for Cloud:

- Physical and virtual machine reservation
- Image capture and reuse
- Workload submission and monitoring to a public cluster
- Calendar view of resource allocations
- Usage, audit and billing records
- Events and alerts for the infrastructure based on their responsibilities
- Password management
- Virtual private cluster related services, including rapid provisioning, and energy metering and management for physical machines.

Administrators can access the following additional services for the entire cloud:

- User management
- Role definition and user access policy configuration
- System management and monitoring
- Energy management policies, monitoring and usage records for provisioned nodes
- Usage and billing management and monitoring
- Workload submission and monitoring
- Events, alerts and security audit records

The level of control and visibility provided by the management suite allows administrators to truly plan and manage their capacity. By not over-virtualizing every node, administrators are able to create a platform where applications run with a degree of confidence and quality of service; the application architecture can become more predictable.

High utilization through automation

The HPC Management Suite for Cloud supports the creation of an infrastructure for applications on the fly, based on policies set by the administrator. While reservation, provisioning, configuration and charge back can all be accomplished through straightforward processes, the demands of HPC workloads require that there be pre-planned approaches and flexibility to handle the inevitable requests for greater resources.

From a user perspective, automation can provide fully configured clusters to run their application, rather than loosely provisioned machines. The management suite can already have compute nodes designated, the head node designated, a job queue set up and monitoring established.

All the user needs to know is that the nodes are ready, the reservation has been granted, and everything is set to begin loading jobs into queue.

Configuration and provisioning are essentially simultaneous—there is no lag between when a user is given the nodes and the time to run their application. This means that users and their departments only get charged for productive use of their nodes, not for setup.

With automated setup and configuration for HPC workloads, you get higher resource utilization, along with effective management of multiple clusters within a private cloud, and single-point submission and monitoring of multiple job queues. In addition, this agility enables you to move from a conventional grid to cloud computing if the grid could not provide all the software environments the users needs, or if manual provisioning and configuration was not easy or fast enough to be productive.

Versatile provisioning and cloud workload management

The suite's provisioning tool is a versatile platform to support rapid, diskless provisioning of a wide variety of operating system images to physical and virtual machines. Among physical machines, it currently supports x86 architectures. It can also create and manage virtual machines on hypervisors like a Kernel-based Virtual Machine (KVM).

The suite also supports cloud-specific features such as dynamic setup of virtual private clusters, capture of running images from physical and virtual machines, energy management and monitoring.

In an HPC cloud, the workflow and workload components of the HPC Management Suite for Cloud provide the capabilities that allow optimal utilization of resources, such as reservation, automated provisioning, resource scheduling and resource management. HPC cloud services are implemented using a framework of batch job templates that execute scripts designed for the service.

A unique feature within the management suite is the ability to configure a cluster with workload management capability already set-up by default.

With its existing workflow management capabilities and resource management APIs, the management suite provides extensive options for third-party software and tool integration into the HPC cloud.

Power management without sacrificing performance

Large compute clusters, especially in HPC, can consume large amounts of energy. A goal of many cloud implementations is to reduce operational expenditure. While virtualizing resources could help save energy by reducing infrastructure footprint, virtualization is not typically an option for most HPC applications. The HPC Management Suite for Cloud provides the tools to create and administer energy schemes, power management and consumption policies that are suited to the workloads, including energy monitoring and management of IBM iDataPlex® systems. The suite also provides options for you to charge departments for the energy consumed by taking advantage of the accounting records.

Usage metering and accounting

Metering capability in an HPC cloud could include measuring many types of resources. These usage records can be used as efficiency indicators of an application, user or department. They can also be used as a means for invoicing the user or department based on customized tariffs.

The HPC Management Suite for Cloud provides an open database schema for metering and accounting. This database captures usage records for compute resources, energy usage for iDataPlex hardware and the duration of use on a per-user basis. A sample billing module is provided to show how usage records can be turned into chargeback records, producing a “roll-your-own” tariffs and billing package.

Conclusion

With IBM HPC cloud solutions, the benefits of cloud computing are now available to scientific and technical applications. The HPC Management Suite for Cloud plays a key role in these solutions, providing both bare metal and virtual machine provisioning in a cloud environment and a quality of service assurance when operating on virtual machines to provide the performance necessary for scientific and technical computing workload.

For more information

To learn more about the IBM HPC Management Suite for Cloud please contact your IBM representative or IBM Business Partner, or visit:

ibm.com/systems

© 2015 Lenovo. All rights reserved.

Availability: Offers, prices, specifications and availability may change without notice. Lenovo is not responsible for photographic or typographic errors. **Warranty:** For a copy of applicable warranties, write to: Warranty Information, 500 Park Offices Drive, RTP, NC, 27709, Attn: Dept. ZPYA/B600. Lenovo makes no representation or warranty regarding third-party products or services. **Trademarks:** Lenovo, the Lenovo logo, System x, ThinkServer, TruDDR4 are trademarks or registered trademarks of Lenovo. Microsoft and Windows are registered trademarks of Microsoft Corporation. Intel, the Intel logo, Intel Core, Core Inside, Xeon and Xeon Inside are registered trademarks of Intel Corporation in the U.S. and other countries. Other company, product, and service names may be trademarks or service marks of others. Visit <http://www.lenovo.com/lenovo/us/en/safecomp.html> periodically for the latest information on safe and effective computing.

IBM x86 products are now products of Lenovo in the U.S. and other countries. Learn more at ibm.com/lenovo-acquisition



Please Recycle