# Management += Grid

Sven Graupner, Vijay Machiraju, Akhil Sahai, Aad van Moorsel
Internet Systems and Storage Laboratory
HP Laboratories Palo Alto
HPL-2003-114
June 5th , 2003*

E-mail: {svgr, vijaym, asahai, aad} @hpl.hp.com

management, Grid, SLA, closed-loop management, monitoring, control, resource model

In this article we discuss the need for expanding management capabilities to provide complete life-cycle of resources assigned to applications. We propose a management+ system that uses grid and traditional management technologies to achieve that. A management+ system will be comprised of grid nodes that communicate with each other to coordinate the tasks of resource allocation and management. The management+ system may be used for managing a resource pool inside the enterprise (an enterprise management grid) or one spanning multiple enterprises.

* Internal Accession Date Only
Approved for External Publication

# Management += Grid

Sven Graupner, Vijay Machiraju, Akhil Sahai, Aad van Moorsel
Hewlett-Packard Laboratories
1501 Page Mill Road, Palo Alto, CA 94034
{svgr, vijaym, asahai, aad}@hpl.hp.com

State of the art management products have until now focused on monitoring of resources such as networks, systems, servers and applications. Management usually does not deal with handling the complete resource life-cycle, i.e., of providing resources on-demand to application providers, matching and allocating resources to users so that application requirements are best met, providing guarantees through service level agreements, and monitoring and assuring these SLAs. Grid aims to provide seamless access to computational resources. Most of the Grid based systems till now are based on "cycle-stealing" technology for using available cycles on a cluster of computers. In Grid based systems and the latest initiative of Open Grid Service Infrastructure (OGSI) [3] concepts related to discovery of network topology, allocation based on farms, monitoring towards certain goal (as specified in SLAs), and control to assure these goals are missing. However, there are parallels between grid and traditional management systems. A grid is comprised of grid nodes each of which manages a group of resources. A traditional management system also may be viewed as comprised of managers that monitor and control a group of resources. The management+ [1] system that we propose, will be able to manage the complete life-cycle of resources assigned to applications, provided:

- Grid technologies are used for reserving, allocating, and handing over resources to applications.
- Traditional management functionalities are incorporated that deal with monitoring and control.

A management+ system will be comprised of grid nodes that communicate with each other to coordinate the tasks of resource allocation and management. *The management+ system may be used for managing a resource pool inside the enterprise or spanning multiple enterprises*. In order to do so, the management+ system needs capabilities to discover the resources so as to create a resource pool, maintain & model the resource information, be able to provide resources to application providers on receiving requests, be able to provide guarantees as agreed upon through Service-Level Agreements, and by monitoring and assuring them. The management+ system poses following requirements:

- **Resource discovery:** Under the realm of each grid node are collections of resources that are allocatable and manageable by that grid node. This raises the problem of registration and discovery. The registration information may be maintained in a central repository that may be LDAP based (as in traditional Grid) or a UDDI based registry. The other approach is to let each of the grid nodes manage their own repository without exposing the resources on the grid, but to have protocols for searching for resources amongst grid nodes as in p-2-p systems.

- **Resource modeling:** Resources that are managed by the grid or those even within a grid node are heterogeneous in nature. They are of different types, potentially distributed in different geographic locations and administrative domains (e.g., consolidated in racks vs. located on desktops), and governed by different policies. They could be added, constructed out of other resources, upgraded, or removed over time. Modeling is therefore an important task for the management+ system. A relevant resource model standard is the Common Information Model (CIM).

- **Requesting resources and guarantees: A**pplication providers need to specify their requests to the management+ system in some form. The specification of the request may be done in terms of application level metrics (e.g. throughput, transactions/sec) or at a low-level, an enhanced Resource Specification Language (RSL) (as in grid) may be used. An SLA is typically signed between two parties that have the role of provider and consumer respectively. The Management+ system needs to provide application providers guarantee on their resources usually in terms of reliability, availability, security, timeliness, transactionality [5].

- **Resource Allocation and deployment:** Resource requests are sent to the management+ system that then undertakes resource reservation, allocation and deployment. Resource reservation is done when the resources are not immediately required. The management+ system has to keep a note of all the competing reservations and of ensuring that the resource pool is being utilized according to its policies. As and when the reservations become current, requests for resources have to be satisfied. This leads to match-making [3] of application provider requirements with the resources maintained in the resource pool.

- **Monitoring guarantees:** Traditionally, applications are installed and operated on a fixed set of associated, physical hardware. In management+ system, resources virtualize the physical hardware. This virtualization provides the capability of transparently switching the physical hardware (in case of degradations/failures) while maintaining the transparency of a continually running application. Monitoring SLA guarantees assumes that monitored data at the hardware level can be associated with applications that are operating on resources. The metrics specified on the resources have to be monitored and aggregated into higher-level metrics that business manager(s) may relate to.

- **Assuring guarantees:** Once the SLA violations are detected, it is important for the management+ system to take corrective actions. Analysis tools are employed to analyze the violation data so as to decide on corrective action(s). The corrective action may range from taking one of the control actions like fail-over, reboot/rejuvenation of the resources to transparently switching the failed resources with new set of resources out of the resource pool. These new resources are obtained through the same resource allocation and deployment mechanisms as discussed earlier. This may enable closed-loop management.

## Conclusion

The Management+ system combines traditional management and grid technologies in an innovative manner. It is also a work in progress.

## References


1. Sahai A, Machiraju v, van Moorsel A. A System that combines Grid and Management technologies for Closed Loop Enterprise IT Management. HPL Invention disclosure #200310038 (patent pending)
2. Platform Computing . http://www.platform.org
3. Raman R, Livny M, Sloman. M . Match-Making: Distributed Resource Management for High Throughput Computing. *Proceedings of the Seventh IEEE International Symposium on High Performance Distributed Computing*, July 28-31, 1998, Chicago, IL.
4. Foster I., Kesselman C, Nick J.M., Tuecke S. The Physiology of the Grid - An Open Grid Services Architecture for Distributed Systems Integration, DRAFT, May 2002
http://www.globus.org/research/papers/ogsa.pdf
5. Sahai A, Graupner S, Machiraju V, van Moorsel A. Specifying and Monitoring Commercial Grids through SLA. CCGrid, May 2003.