A Grid Infrastructure for Utility Computing *

Ignacio M. Llorente[1], Rubén S. Montero[1], Eduardo Huedo[1], Katia Leal[2]

[1] Univ. Complutense de Madrid
Dep. Arquitec. de Comp. y Automática
Facultad de Informática 28040, Spain

[2] Universidad Rey Juan Carlos
Dep. de Ing. Telemática y Tec. Electrónica
Escuela Sup. de Ciencias Experimentales y Tec.
Tulipán SN, Mósteles, Madrid, Spain

Abstract

Utility computing is a service provisioning model, which will provide adaptive, flexible and simple access to computing resources, enabling a pay-per-use model for computing similar to traditional utilities such as water or electricity. The deployment of a utility computing solution involves a full separation between the provider and the consumer. The consumer requires a uniform, secure and reliable functionality to access the utility computing service and the provider requires a scalable, flexible and adaptive infrastructure to provide the service. The solution should be based on standards and allow a gradual deployment in order to obtain a favourable response from the application developers and the information technology staff. Grid technology overcomes such challenges by means of its standard functionality for flexible integration of diverse distributed resources. This position paper proposes an innovative solution for utility computing which can be deployed on a Grid infrastructure based on Globus toolkit and GridWay components

1 Introduction

Utility is a computing term related to a new paradigm for an information technology (IT) provision. A utility computing solution should provide access to the latest computing platform and technology and still be flexible enough to adjust capacity as required without needing to purchase costly hardware. In doing so information technology costs are transferred from fixed to variable as the organization gets charged for how much they use [8]. Such pay-as-you-use paradigm exhibits several potential benefits for an organization: reducing fixed costs, treating IT as a variable cost, providing access to unlimited computational capacity and improving flexibility, thereby making resource provision more agile and adaptive. Such valuable benefits may bring the current fixed-pricing policy of IT provision to an end, where computing is carried out within individual corporations or outsourced to external service providers [3].

The deployment of a utility solution requires a full separation between the provider and the consumer. As is argued in Section 2, a feasible and realistic solution should meet the following requirements: a uniform, secure and reliable functionality to access the utility, a scalable, flexible and adaptive infrastructure to provide the service and a solution based on standards and gradually deployable. Section 3 describes the grid as the infrastructure that offers a common layer to integrate non-interoperable computational platforms. Today's utility solutions do not use a grid infrastructure. Our thesis is that utility solutions will spread provided that they are based on grid technology. Grid technology overcomes utility computing challenges by means of its standard functionality for flexible integration of diverse distributed resources. In Section 4, we propose an innovative solution for utility computing which can be deployed on a Grid infrastructure based on Globus toolkit and GridWay components. Finally, Section 5 presents some conclusions, and the plans for future work.

---

*This research was supported by Consejería de Educación of Comunidad de Madrid, Fondo Europeo de Desarrollo Regional (FEDER) and Fondo Social Europeo (FSE) through BIOGRIDNET Research Program S-0505/TIC/000101 and by BSCH/UCM through research grant PR27/05-14035-BSCH.

Proceedings of the 15th IEEE International Workshops on Enabling Technologies:Infrastructure for Collaborative Enterprises (WETICE'06)
0-7695-2623-3/06 $20.00 © 2006 IEEE

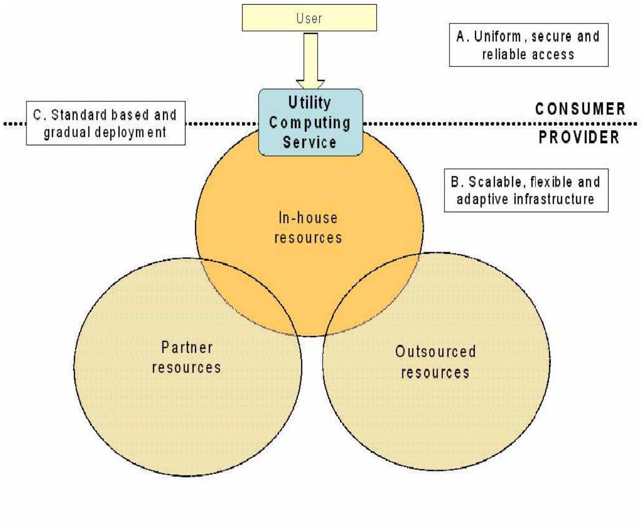

Figure 1. Requirements of a utility computing solution.

2 Requirements of a utility computing solution

The deployment of a utility solution involves a full separation between the provider and the consumer. Therefore, a feasible and realistic solution should meet some requirements, as shown in figure 1, that we will see in the following sections.

2.1 Secure and reliable functionality to access computing resources

Application developers, portal builders, and Independent Software Vendors (ISVs) should be able to delegate the execution of computing demanding tasks by means of a single uniform interface that provides the high level functionality required to define, submit, monitor and synchronize single, complex and array jobs. Such an interface should allow end user applications to access computing resources in a secure and reliable way through interaction with a utility computing service. The utility computing service implements the separation between the consumer and the provider, acting as a proxy to access utility resources. The end user should be able to monitor and control the state of the tasks submitted and to access and manage its own resource usage accounting and statistics. However, as soon as a secure and reliable execution is guaranteed, the end user is no longer interested in the underlying resources where the tasks are finally executed.

2.2 Scalable, flexible and adaptive infrastructure to provide computing resources

Resources should be dynamically provisioned when addressing the changing service needs in an organization. The utility computing service owner, or computing provider, may be the organization's IT department or an external service provider. The administrators must maintain total flexibility in the utility computing deployment, from in-house provision to fully outsource, or hybrid approaches in which in-house resources can be supplemented by outsourced resources or partner resources to satisfy peak or unusual demands. Some companies, such as IBM[1] or Sun Microsystems[2], are currently offering computing utility services. The separation between consumers and providers gives administrators new ways of managing workloads and resources that may even involve the adaptation of business and research processes.

2.3 Standard based and gradual deployment

Standardization generates an environment in which final users, ISVs and technology providers can undertake investments with greater confidence. On the other hand, the real requirements for deploying utility solutions should also be considered. The Utility Accounting Working Group of the Enterprise Grid Alliance[3] is creating use cases to help define utility computing requirements in enterprise grid computing environments. On the other hand, a solution should allow for gradual deployment in order to deal with the obstacles for adoption such as enterprise skepticism and IT staff and management resistance [2]. As a first step the utility service should be managed by the IT department, the return of investment should be demonstrated and a low number of applications ported to the utility model. In fact, in the worst case, cultural limitations would restrict utility computing to an in-house private utility.

3 Grid Infrastructures

Several Distributed Resource Management (DRM) systems have been developed to provide workload management of applications. These systems first appeared in the late eighties, coinciding with the advent of parallel and distributed platforms, such as high performance computing servers and clusters. There are a number of

---

[1] http://www.gridtoday.com/03/0120/100974.html
[2] http://www.sun.com/service/sungrid/index.html
[3] http://www.gridalliance.org

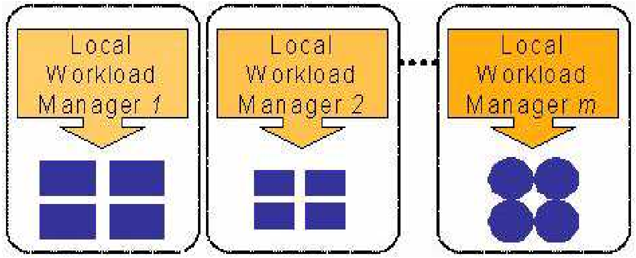

Figure 2. Distinct local workload managers are used for efficient harnessing of different computing platforms (*vertical silos*).

commercial and open source local workload management systems available today, such as PBS[4], SGE[5], LSF[6] or Condor[7], and each one is used for different underlying computer architectures and execution profiles. In any case their benefits in cost minimization and performance maximization are mainly due to greater utilization of underlying resources.

Even though DRM systems share many capabilities, mainly batch queuing, job scheduling and resource management; they do not provide a common interface and security framework, and so their integration is not possible. As shown in figure 2, such lack of interoperability involves the existence within an organization of independent computational platforms (vertical silos) responsible for distinct functions that require specialized administration skills and generates overprovisioning and global load unbalance. Moreover, such technologies are also unsuitable to build computational infrastructures where resources are scattered across several administrative domains, each with its own security policy and DRM system.

A grid infrastructure offers a common layer to integrate these non-interoperable computational platforms by defining a consistent set of abstraction and interfaces for access to, and management of, shared resources [5]. The Globus Toolkit [4], a de facto standard in grid computing, is open source software that implements a collection of high level services at the grid infrastructure layer. These services include, among others, resource monitoring and discovery services (MDS), resource allocation & management (GRAM), a security infrastructure (GSI), and file transfer services (RFT).

The Globus Toolkit meets most of the abstract requirements set forth in Open Grid Services Architecture (OGSA)[8]. OGSA, under development by the Global Grid Forum (GGF)[9], aims to define a common, standard, and open architecture for the grid. The Web Services Resource Framework (WSRF)[10], developed by OASIS[11], defines a family of specifications based on standard Web Services to access the stateful resources that OGSA needs. In fact, the Globus Toolkit is implemented on top of the WSRF standard.

Several open-source software components[12] are available to deploy Globus-based grid solutions that address the requirements of the new grid projects. The deployment of a grid infrastructure for utility computing is just one case, and in it the meta-scheduler plays a key role. The Globus layer provides a uniform interface to many different DRM systems, allowing the development of grid workload managers that optimize the use of the underlying computing platforms. Grids are loosely coupled systems that inherently present the following characteristics: scalability, autonomy of the multiple administration domains, dynamism, and heterogeneity [7]. These characteristics determine how scheduling and execution must be performed on grids. Scalability and autonomy prevent the deployment of centralized resource brokers, which maintain total control over client requests and resource status. In addition, job scheduling and execution must be adaptable to resource dynamics, such as changing availability, capacity, and cost.

In the grid infrastructure shown in figure 3, the GridWay[13] [6] workload manager performs job execution management and resource brokering on a grid consisting of distinct computing platforms managed by Globus services. GridWay allows unattended, reliable, and efficient execution of single, array, or complex jobs on heterogeneous and dynamic grids. GridWay performs all the job scheduling and submission steps transparently to the end user and adapts job execution to changing grid conditions by providing fault recovery mechanisms, dynamic scheduling, migration on-request and opportunistic migration. GridWay on Globus provides decoupling between applications and the underlying local management systems.

The evolution of grid computing is expected to follow this pattern [9]. Initially, the Globus toolkit supports the deployment of enterprise grids that enable diverse resource sharing to improve internal collaboration and achieve a better return from their information technology investment. Secondly, partner grids of sev-

---

[4] http://www.openpbs.org
[5] http://www.sun.com/software/gridware
[6] http://www.platform.com
[7] http://www.cs.wisc.edu/condor/
[8] http://forge.gridforum.org/projects/ogsa-wg
[9] http://www.gridforum.org
[10] http://www.globus.org/wsrf/
[11] http://www.oasis-open.org
[12] http://www.globus.org/grid_software/computation/
[13] http://www.gridway.org

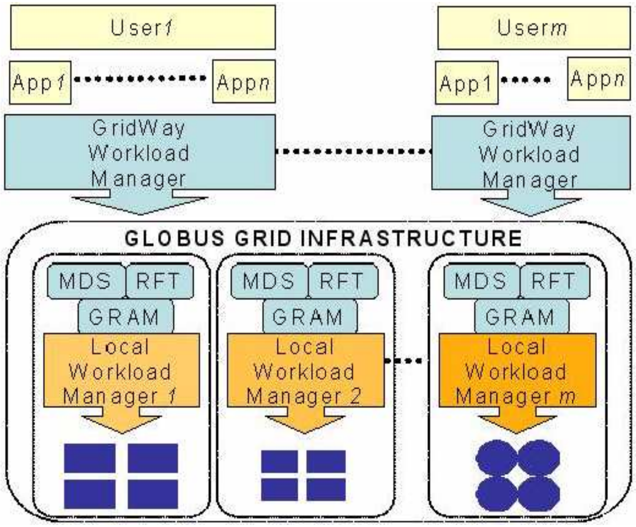

Figure 3. Globus services provide a uniform interface for secure discovering, monitoring (MDS) and accessing (GRAM) distinct local workload managers; and file transferring (RFT). GridWay workload manager performs job execution management and resource brokering on top of Globus services.

eral scales will be mainly deployed within the context of different research projects, whose final goal is to provide large-scale, secure and reliable sharing of resources among partner organizations and supply-chain participants. Such partner grids allow access to a higher computing performance to satisfy peak demands and also provide support to face collaborative projects. Finally, it has been predicted that outsourced grids, managed by dedicated service providers, will supply resources on demand over the Internet. Different studies suggest that increased network capacity will allow businesses and consumers to draw their computing resources from outsourced grids apart from enterprise grids

4 A Compute Utility Solution based on the Globus Toolkit

Grid technology overcomes utility computing challenges by means of its standard functionality for flexible integration of diverse distributed resources. The technological feasibility of the utility model for computing services proposed above is established by using a novel grid infrastructure based on Globus Toolkit components and the GridWay workload manager. It is well known that the WS-GRAM[14] Globus Toolkit service provides a uniform, secure and reliable interface to heterogeneous computing platforms managed by different DRM systems. The main innovation of our model is the use of the WS-GRAM service to recursively interface to the services available in a federated Globus based grid. A WS-GRAM service hosting a GridWay workload manager provides the standard functionality required to implement a gateway to a federated grid. Such a combination allows the required virtualization technology to be created in order to provide a powerful abstraction of the underlying grid resource management services

4.1 Uniform, secure and reliable functionality to access computing resources

A grid gateway, i.e. a WS-GRAM service hosting GridWay, may act as the utility computing service, providing a uniform standard interface for the secure and reliable submission and control of jobs on grid resources. The security requirement at the user level is addressed by the GSI component of the Globus Toolkit. At the beginning of a working session, the user specifies the grid gateway address and authenticates with an X509 identity. The access to grid resources by means of a grid gateway also minimizes the firewall configuration, which is a welcome advance for security administrators.

Application developers, portal builders, and ISVs will interface with the utility computing service by using the Distributed Resource Management Application API (DRMAA)[15]. DRMAA is a GGF standard that constitutes a homogenous interface to different DRM systems to handle job submission, monitoring and control, and retrieval of finished job status. The GridWay job execution management functionality is also required in the end user's system to provide support for the DRMAA functionality. GridWay is the only grid workload manager that complies with the DRMAA standard.

4.2 A scalable, flexible and adaptive infrastructure to provide computing resources

A WS-GRAM service hosting a grid workload manager, such as GridWay, may also act as a grid gateway between two grid infrastructures. The grid gateway is managed as a common resource in the first grid and acts as a proxy of the users of the first grid in the second grid. The second grid appears in the information system (MDS service) on the first grid as a

---

[14] http://www.globus.org/grid_software/computation/gram.php

[15] http://www.drmaa.org

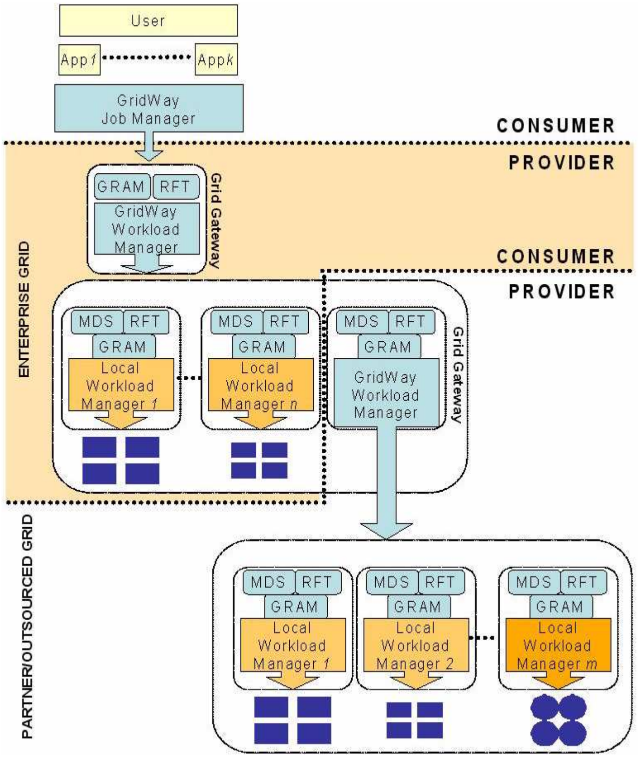

Figure 4. A utility computing solution based on Globus services and the GridWay workload manager.

normal WS-GRAM publishing an aggregation of grid resources; and enabling submission, monitoring and control of jobs across the grids. Such a federation of grids meets the scalability, flexibility and adaptability requirements of the infrastructure for a compute utility solution.

A similar approach has been previously applied to meet LCG and GridX1 infrastructures [10], hosting a GridX1 user interface in a LCG computing element. However, this solution is not based on standards. A scalable solution should follow the hourglass approach in the Globus architecture, in which the bottom of the hourglass represents resources, the middle is the core Globus services, and the top refers to higher-level Globus services and applications. The grid workload manager in a grid gateway should have access to a wide range of resources provided through a limited, standardized set of protocols and interfaces. The Globus core grid middleware provides this set. Just as in the Internet, the protocols and interfaces are provided through TCP/IP.

The grid hierarchy in our utility computing model is clear. An enterprise grid, managed by the IT Department, includes a grid gateway to an outsourced grid, managed by a utility computing service provider. The outsourced grid provides pay-per-use computational power when local resources are overloaded. This hierarchical grid organization may be extended recursively to federate a higher number of partner or outsourced grid infrastructures with consumer/provider relationships. Figure 4 shows one of the many grid infrastructure hierarchies that can be deployed with grid gateway components.

The access to resources, including user authentication, across grid boundaries is under control of the grid gateway service and is transparent to end users. In fact, different policies for job transfer and load balancing can be defined in the grid gateway. The user and resource accounting and management could be performed at different aggregation levels in each infrastructure.

4.3 Standard based and gradual deployment

Today's utility solutions do not use a grid infrastructure. A grid involves standardization, so utility solutions will spread provided that grid technology is available. On the other hand, the cultural and business model changes required for adopting the utility model should be gradual, starting with access to a local workload manager, followed by an in-house enterprise grid and finally moving onto outsourced services. These adoption steps are transparent to the end user in the psed solution since he always interfaces with a given grid gateway (WS-GRAM service) using DRMAA standard.

5 Conclusions

The innovative utility solution for computing provision, which can be deployed on a grid infrastructure based on existing Globus toolkit components and related tools, will allow companies and research centers to access their in-house, partner and outsourced computing resources via automated methods using grid standards in a simpler, more flexible and adaptive way. New components for negotiation, service level agreement, accounting, and billing are being developed in the context of the Grid4Utility project[16]. Our solution can be treated as a prototype in two areas. Firstly, it is a useful tool in the development of extended solutions to additionally access file replica services and data bases, and secondly, as a testbed for new research lines [1] under development which result from the business and technological challenges of this computing model.

---

[16] http://www.grid4utility.org

References

[1] "Utility Computing". *IBM System Journal*, 43(1), 2004.

[2] A. Bednarz and D. Dubiem. "How to Get to Utility Computing". *TEchWorld*, 2004.

[3] N. G. Carr. "The End of Corporate Computing". *MIT Sloan Management Review*, 46(4), Summer 2005.

[4] I. Foster. "Globus Toolkit Version 4: Software for Service-Oriented Systems". In *IFIP, International Conference on Network and Parallel Computing*, volume LNCS 3779, pages 2–13. Springer-Verlag, 2005.

[5] I. Foster and S. Tuecke. "Describing the Elephant: The Different Faces of IT as Service". *Enterprise Distributed Computing*, 3(6), July-August 2005.

[6] E. Huedo, R. S. Montero, and I. M. Llorente. "A Framework for Adaptive Execution on Grids". *Journal of Software - Practice and Experience*, 34:631–651, 2004.

[7] I. M. Llorente, R. S. Montero, and E. Huedo. "A Loosely Coupled Vision for Computational Grids". *IEEE Distributed Systems Online*, 6(5), 2005.

[8] R. Murch. "Introduction to Utility Computing: How It Can Improve TCO". *Prentice Hall*, 2005.

[9] P. Plaszczak and R. Wellner. *"Grid Computing: The Savvy Manager's Guide"*. Morgan Kaufmann, 2005.

[10] R. Walker, M. Vetterli, R. Impey, G. Mateescu, B. Caron, A. Agarwal, A. Dimopoulos, L. Klektau, C. Lindsay, R. Sobie, and D. Vanderster. "Federating Grids: LCG Meets Canadian GridX1". *GlobusWorld*, 2005.