



# Nuevos modelos para la provisión de servicios comerciales basados en Grid

## New Models for Commercial Provisioning of Grid Based Services

◆ E. Jacob, J. J. Unzilla, M<sup>a</sup> V. Higuero y P. Saiz

◆  
El concepto de Grid que se utiliza originalmente hace difícil plantear una comercialización exitosa del mismo por la especificidad de los clientes potenciales

### Resumen

En este artículo se va a presentar un trabajo en curso, en el que se plantean varios esquemas para la comercialización de servicios basados en el paradigma Grid (en la acepción más amplia, el uso de recursos distribuidos de una manera colaborativa) diseñados para ser ofrecidos con la participación de un ISP. Se estudiarán las características que tiene que tener un servicio de este tipo y se describirá más a fondo el tercero de los esquemas propuestos y el servicio escogido denominado de "almacenamiento perdurable". Se discutirá brevemente su arquitectura y los criterios de diseño a utilizar.

**Palabras clave:** Grid, recursos distribuidos, almacenamiento perdurable.

### Summary

In this article, we are going to present a work in progress, in which we will present several schemas for commercial provisioning of services that are based in the GRID paradigm (In its broadest sense, the cooperative use of distributed processing power, disk space or communication resources), through an ISP. We will study the characteristics that a service of this kind must have and it will be described the third model and the selected service to be provided: "durable storage". Finally we will briefly present architecture and the criteria to be used in its design.

**Keywords:** Grid, distributed processing power, durable storage.

## 1.- Introducción

Las tecnologías Grid se desarrollaron inicialmente con el objetivo de conseguir usar recursos distribuidos de forma masiva, básicamente potencia de procesamiento y espacio de almacenamiento. Con respecto a otras tecnologías anteriores (PVM, MPI, etc.) que ofrecían prestaciones similares, hay que destacar la automatización en el acceso seguro a esos recursos. El toolkit para la implementación de Grid por antonomasia es Globus, basado en las ideas de Ian Foster [1]. Aun cuando hay diversas tecnologías que permiten ofrecer servicios Grid, se está realizando un esfuerzo importante de estandarización a través de OGSA (Open Grid Service Architecture).

El concepto de Grid que se utiliza originalmente (entornos con necesidades de procesamiento, comunicación y almacenamiento de masivo de datos) hace difícil plantear una comercialización exitosa del mismo por la especificidad de los clientes potenciales. Por otro lado, uno de los entornos naturales de comercialización de servicios son los proveedores de servicio Internet. Los ISP sirven de punto de interconexión entre multitud de clientes y suelen disponer de varios CPD distribuidos geográficamente.

Es en este entorno cuando surge un proyecto de investigación Universidad-Empresa financiado por la Universidad del País Vasco y la empresa Contec, S.L. denominado "Definición de modelos de provisión de servicios telemáticos basados en tecnología Grid Computing" con una duración de dos años (2003-2004). Este artículo va a presentar las conclusiones del primer año de trabajo y las líneas para el segundo.

## 2.- El ISP como una organización virtual

Podemos considerar este conjunto de usuarios y CPD de ISP como una nueva "organización virtual". Esto nos permite diseñar diversos escenarios de utilización de la tecnología GRID que consideramos novedosos que representan una evolución del modelo de ASP.

Antes de nada trataremos de caracterizar a esta "organización". En gran cantidad de casos de conectividad corporativa, la conexión entre clientes e ISP está basada en enlaces permanentes con velocidades que van desde los 128Kbit/s a los 2Mbit/s; en muchos casos el consumo de ancho de banda sigue un modelo clásico de pico/valle superpuesto a la jornada laboral y los usuarios tienen en este momento en su gran mayoría pocas necesidades adicionales de servicio distribuidos, alternativos a los de conectividad básica. La heterogeneidad de las aplicaciones empleadas, si exceptuamos las ofimáticas y el número relativamente pequeño de usuarios con grandes necesidades de proceso, hacen que en este entorno no sea demasiado factible, al menos en este momento, el aprovechamiento de CPU. De hecho, es un atractivo para los clientes poder aprovechar el enlace infrautilizado fuera de las horas laborables. El ISP en muchos casos dispone de diversos CPD, segregados físicamente, aunque las comunicaciones entre ellos siguen un patrón de comunicaciones distinto. Trabajos previos realizados nos hacen ver que uno de los pocos servicios que podían calar en la comunidad de usuarios sería el de backups remotos.

Los servicios que son vistos con buenos ojos tanto por ISP como por clientes de los mismos son los relativos al almacenamiento remoto

### 3.- Paradigmas de comercialización de servicios Grid

A través de proyectos de investigación relacionados con la provisión de servicios ASP, hemos llegado a la conclusión de que los servicios que son vistos con buenos ojos tanto por ISP como por clientes de los mismos son los relativos al almacenamiento remoto (copias de seguridad o discos remotos).

Los aspectos a tener en cuenta en el diseño de un servicio comercial son los siguientes:

- Tiene que haber una conectividad periódicamente desaprovechada, normalmente nocturna.
- Se debe ofrecer algo de lo que el cliente no dispone, en este caso se ofrece redundancia y disponibilidad de los datos.
- Hay que tener en cuenta las dudas y las percepciones de los usuarios sobre la confidencialidad y seguridad de los datos.
- Tiene que haber un criterio de facturación claro o, visto de otra manera, hay que saber cómo se define y granulariza la calidad del servicio.

En este momento y a la hora de caracterizar lo que puede ser un Grid al margen del concepto directamente vinculado a Globus vamos a requerir dos características: que estén basados en sistemas distribuidos y que se empleen sistemas colaborativos.

La primera característica es natural, pero también común a muchos servicios comerciales ofrecidos por ISP. La segunda, sin embargo, es más novedosa. Se trata de establecer una relación entre las partes, en las que éstas se implican bilateralmente a través de algo más que una transacción económica: hay una colaboración en aras de un objetivo común.

Los escenarios posibles que se han detectado son varios pero tratan de ofertar almacenamiento de alta disponibilidad o redundante tanto como backup (sin acceso aleatorio) como de disco virtual.

- El primer escenario consiste en que el ISP puede ofrecer servicios basados en tecnologías GRID utilizando recursos propios (de disco en sus CPD). Los parámetros que se pueden modular y por tanto incluir en un acuerdo de nivel de servicio en este caso son además de la capacidad, el nivel de redundancia (número de copias y proveedores de conectividad) y la rapidez de replicación y acceso. En este caso, como en los siguientes, la confidencialidad de los datos es siempre un problema que, aunque resuelto técnicamente, puede frenar la utilización de estos servicios.
- El segundo escenario propuesto consiste en la realización de sistemas basados en P2P, en los que el ISP puede ser configurado como servidor de directorios y que son adecuados cuando una misma empresa tiene diversas localizaciones físicas. En este supuesto, la gestión del sistema se



El servicio de almacenamiento perdurable es un producto que ofrece a los clientes un almacenamiento de datos externo a su empresa y con la particularidad de estar disponible "siempre"

externaliza al ISP y los datos (excepto los contenidos en el directorio) residen en la infraestructura del cliente.

- En el tercer caso, el ISP utiliza los discos virtuales remotos de clientes o propios instalados en las dependencias de clientes y ofrece el servicio de almacenamiento estable y de alta disponibilidad a terceros. Como en el primer caso, existen parámetros para fijar el acuerdo de nivel de servicio en la oferta a los clientes, pero también hay otros parámetros de calidad respecto a los "clientes-repositorios", en lo que se refiere a la degradación de las comunicaciones. El aspecto colaborativo puede verse reflejado en la facturación, ya que los clientes que además de usar el servicio ofrecen sus recursos pueden ver la misma reducida. El producto obtenido se conoce como "almacenamiento perdurable".

El primer escenario en nuestra opinión es difícil de plantear hoy comercialmente, salvo a través de ISP de características muy específicas tanto por el plan de negocio como por los clientes y sus necesidades, es el caso de los NRENS. El segundo caso tiene su razón de ser, pero se ve sobrepasado en interés por el tercero. Este último es el que más interés presenta y el que se tratará de desarrollar en el proyecto de investigación que nos ocupa.

#### 4.- El servicio de almacenamiento perdurable

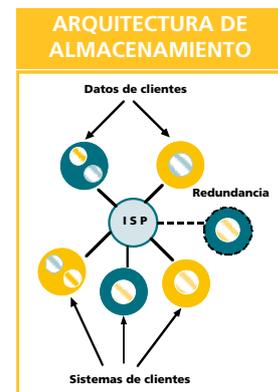
El servicio de almacenamiento perdurable es un producto que ofrece a los clientes un almacenamiento de datos externo a su empresa (no sólo a su propia delegación) y con la particularidad de estar disponible "siempre". Este servicio tiene ya un referente histórico en el artículo seminal de Ross Anderson [2], aunque en este caso el objetivo era el de proveer de anonimato de lectores y productores de contenidos y garantizar la libertad de difusión. Otro ejemplo de conceptos similares se encuentra en [3].

Dado que los términos absolutos no son prácticamente reales nunca, hay que matizar la definición de "siempre". El objetivo de este servicio es replicar datos en diversas localizaciones geográficas, empleando para ello no sólo infraestructuras (de almacenamiento y comunicaciones) del propio cliente o del ISP, sino de otros clientes. Esto permite suponer que en función del número de replicas y la localización de las mismas, la información puede superar incluso catástrofes de área geográfica extensa como inundaciones.

La arquitectura que responde a esta arquitectura se muestra en la figura donde se representa al ISP como elemento central de la lógica del sistema y a dos clientes (uno de cada color) con diversas localizaciones físicas y que tienen datos de los otros clientes.

Esto se asemeja mucho a una arquitectura de tipo P2P. Dado que estas arquitecturas son por definición no gestionables de modo centralizado es necesario dotar al sistema de mecanismos que permitan asegurar una "calidad de servicio": número de replicas, tiempo de replicación por Gb, ocupación del ancho de banda de clientes, etc. Esta calidad de servicio es imprescindible si se piensa ofertar dentro del portafolio de un proveedor y en consecuencia percibir pagos.

En el caso más general, los clientes pueden almacenar no sólo datos propios, sino también datos ajenos y pueden "comprar" o



intercambiar espacio. La característica de colaboracionismo se concreta aquí: los clientes pueden suscribir el servicio como un servicio tradicional y serle facturado de manera clásica, pero también pueden ofrecer espacio de disco (o un sitio para un equipo) y sus recursos de comunicaciones.

El ISP tiene diversas tareas. Por un lado realiza funciones de directorio avanzado: va a seleccionar lugares de réplica de acuerdo con la política de uso y al contrato específico del cliente, detallando el número de replicaciones, la localización geográfica de las mismas (por ejemplo: diversos continentes) y el ancho de banda consumido. Por otro lado, va a permitir la localización de ficheros para recuperación, la regeneración de sitios cuando éstos sufran un percance y a través de sus propios recursos va a asegurar la calidad del servicio en los momentos puntuales en los que el sistema basado en los clientes exclusivamente no sería capaz de entregarla. Finalmente, actúa como interlocutor único para los clientes y centro de compensación entre los mismos.

Lógicamente, el proveedor de Internet pasa a ser un punto crítico dentro del sistema y por tanto debería estar replicado para asegurar la supervivencia del sistema. Sin embargo, es de destacar que el ámbito de trabajo no es extraño para él y no le supone un negocio radicalmente distinto al tradicional.

El sistema tiene unos requerimientos de seguridad altos, básicamente en el campo de la autenticación, cifrado de datos y de alguna manera anonimato (los clientes no deberían saber tampoco de quién son los datos que tienen).

Es de destacar que, aun cuando parte de los objetivos pueden considerarse similares a los propuestos en arquitecturas como [4], la presencia de un ISP como repositorio centralizado, los conceptos de calidad de servicio y el tamaño del sistema (sensiblemente menor) diferencian esta propuesta de aquella.

### 5.- Trabajo futuro

Después de este primer año de trabajo, en el que se ha identificado un área de trabajo, estamos preparando el trabajo para el segundo. A lo largo de este segundo año, se desarrollará un protocolo y un demostrador del mismo. Para ello, se refinará la arquitectura y se estudiarán las opciones de desarrollo (reciclado de opciones ya existentes o utilización de toolkits específicos), la elección de las API (Adhoc o estandarizadas como Web Services) y diversos aspectos de la arquitectura de seguridad.

### 6.- Conclusiones

Aun cuando es todavía demasiado pronto para presentar las conclusiones definitivas del proyecto, en este momento sí es posible adelantar las siguientes:

- Es posible una provisión comercial de servicios basados en Grid, en los que los ISP ocupan un lugar natural.
- Es posible ofrecer un servicio atrayente a los usuarios: el almacenamiento perdurable. Además tiene diversas características que posibilitan una oferta comercial variada.
- No requiere la adquisición de habilidades adicionales para los ISP.
- De manera colaborativa los clientes implicados obtienen beneficios operativos (siempre) y económicos (si participan activamente).

  
El ISP pasa a ser un punto crítico dentro del sistema y por tanto debería estar replicado para asegurar la supervivencia del sistema



## Bibliografía

- [1] Foster, Ian; Kesselman, Carl; Nick, Tuecke, Steven; "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", 2001, "International J. Supercomputer Applications", 15(3)
- [2] Anderson, Ross J. "The Eternity Service", presentado en "Pragocrypt 96", 1996, Praga, pp. 242-252.
- [3] Clarke, Ian; Sandberg, Oskar; Wiley, Brandon; Hong, Theodore W., "Freenet: A Distributed Anonymous Information and Retrieval System", 2001, *Lecture Notes in Computer Science*, Vol. 2009.
- [4] Kubiawicz, John et al; "OceanStore: An Architecture for Global-Scale Persistent Storage", 2000, *Proceedings of ACM Proceedings of the Ninth International Conference on Architectural Support for Programming Languages and Operating Systems*.

**Eduardo Jacob, Juan José Unzilla**  
(jtpjatae@bi.ehu.es), (jtpungaj@bi.ehu.es),  
**M<sup>a</sup> Victoria Higuero, Purificación Saiz**  
(jtphiapm@bi.ehu.es), (jtpsaagp@bi.ehu.es)  
Dpto. de Electrónica y Telecomunicaciones  
UPV/EHU