

Integración de interfaces web con el portal grid GENIUS

Web interface integration with the GENIUS grid portal

M. Oldenhof

Centro de Química, Instituto Venezolano de Investigaciones Científicas (IVIC), Apdo. Postal 21827, Caracas 1020A, Venezuela, y Facultad de Ciencias, Escuela de Computación, Universidad Central de Venezuela, Caracas 1020A, Venezuela.
moldenho@ivic.ve

C. Mendoza

Centro de Física, Instituto Venezolano de Investigaciones Científicas (IVIC), Apdo. Postal 21827, Caracas 1020A, Venezuela, y Centro Nacional de Cálculo Científico Universidad de Los Andes (CeCalCULA), Mérida 5101, Venezuela.
claudio@ivic.ve

Resumen

La interfaz web del programa de estructura atómica AUTOSTRUCTURE ha sido integrada al portal grid GENIUS, lo que facilita su uso en ambientes grid de computación distribuida. También establece una metodología general para adaptar programas de computación científica de alto rendimiento a la nueva e-infraestructura.

Abstract

The web-based interface of the atomic structure code AUTOSTRUCTURE has been integrated to the GENIUS grid portal, which facilitates its use in distributed grid environments. It also establishes a general methodology for adapting high-performance scientific computing codes to the new e-infrastructure.

1. Introducción

Con el uso intensivo de reservorios distribuidos de datos científicos en una cantidad cada vez mayor de proyectos en el área de salud, biomedicina, física, química, astronomía, ambiente, petróleo, entre otros, ha nacido una nueva forma de hacer ciencia: la e-

Ciencia [1]. Visualizamos la e-Ciencia como colaboraciones globales y dinámicas sobre una Internet de segunda generación. Esto implica, por una parte, la explotación de herramientas de tecnología avanzada para recolectar, integrar, compartir, minar, analizar y visualizar grandes volúmenes de datos, y por la otra, simulaciones numéricas en la tera-escala. Estos requerimientos demandan una nueva cyber-infraestructura lo que ha implicado la adopción de ambientes de computación en malla (*grid*) [2].

En el Instituto Venezolano de Investigaciones Científicas (IVIC) se emplea una diversidad de programas de cálculo intensivo en diferentes áreas de investigación. Entre los programas que se utilizan actualmente en las áreas de física atómica, astrofísica y química cuántica, podemos mencionar *AUTOSTRUCTURE* [3, 4, 5], *R-MATRIX* [6], *OPSERVER* [7], *XSTAR* [8] y *CATVIC* [9]. Con las nuevas posibilidades de la computación en malla en puerta, nos vemos en la necesidad de adaptar este portafolio de programas *FORTRAN* a un nuevo esquema operacional donde la computación ahora se concibe como un servicio. Esto implica el desarrollo de interfaces web, paralelización de los códigos, su optimización en ambientes distribuidos y su integración por medio de un portal grid [10] o, alternativamente, una máquina virtual [11].

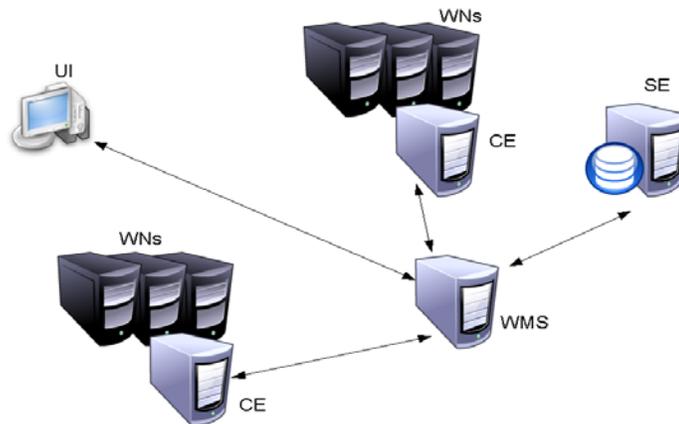


Fig. 1. Arquitectura de GLITE.

En el presente artículo reportamos nuestras experiencias utilizando el portal grid *GENIUS* [12] en la adaptación del programa *AUTOSTRUCTURE* a los requerimientos del entorno de computación grid. Comenzamos en la próxima sección con un recuento de nuestra contribución en la instalación de una red de computación distribuida (GridVenezuela) con clusters de PCs bajo el *middleware* GLITE [13] en cuatro instituciones académicas de Venezuela; específicamente, el IVIC, la Universidad Central de Venezuela (UCV), la Universidad Simón Bolívar (USB) y la Universidad de Los Andes (ULA). También incluimos su entonación y optimización para el cálculo distribuido, la “gridificación” de los códigos y el desarrollo de interfaces web para facilitar interactividad y reducir curvas de aprendizaje. Continuamos con breves descripciones de la aplicación *AUTOSTRUCTURE* y el portal *GENIUS*, para entonces centrarnos en el proceso de integración de la interfaz web de la primera. Finalizamos con las conclusiones del proyecto y algunas recomendaciones sobre su aplicabilidad como metodología general para adaptar programas científicos de computación intensiva a la nueva e-infraestructura.

2. IVIC en el GridVenezuela

La incorporación del IVIC en el GridVenezuela se lleva a cabo con la participación de los Laboratorios de Física Computacional y Química Computacional y la Gerencia de Informática y Sistemas, y con la colaboración cercana de la ULA, USB y UCV. La conexión entre estas instituciones a 34 MB/s se enmarca dentro del programa de Internet2 (REACCIUN2) que patrocina actualmente el Ministerio del Poder Popular para la Ciencia y la

Tecnología. El *middleware* que se utiliza para dicho grid es GLITE [13]. En cualquier proyecto de computación grid, la capa intermedia que proporciona el *middleware* es un componente crucial. GLITE combina una capa intermedia básica con una gama de servicios de mayor nivel.

GLITE se distribuye bajo una licencia de fuente abierta que integra componentes de las mejores herramientas actuales de capa intermedia, tales como CONDOR y GLOBUS TOOLKIT, así como componentes desarrollados para el Large Hadron Collider (LHC) Computing Project (LCG). El resultado es una solución óptima de capa intermedia compatible con el PORTABLE BATCH SYSTEM (PBS), CONDOR y el LOAD SHARE FACILITY (LSF). Tiene como objetivo la interoperabilidad y la provisión de servicios fundamentales que faciliten el desarrollo de aplicaciones grid en todos los campos. GLITE está considerado como un sistema modular que permite que los usuarios implementen diferentes servicios según sus necesidades, sin verse obligados a utilizar el sistema completo. Con esto se pretende que cada usuario adapte el sistema a su situación particular. En la Fig. 1 se puede ver un ejemplo del flujo de una tarea en el grid y la interacción de los diferentes componentes de GLITE.

Los componentes principales de GLITE son:

1. El **User Interface (UI)** es el único componente con el cual el usuario interactúa directamente;
2. El **Workload Management System (WMS)** decide a cuál elemento de cómputo se envía la tarea;
3. El **Computing Element (CE)** se encarga de recibir los trabajos del WMS para luego enviarlos a ejecutar en un nodo de trabajo;

4. El **Worker Node** (WN) es el nodo que finalmente ejecuta el trabajo. La adopción de GLITE asegura la compatibilidad con los grids europeos del Enabling Grids for E-Science (EGEE) y su integración con los de Latinoamérica bajo la E-Infrastructure shared between Europe and Latin America (EELA).

Actualmente el IVIC contribuye al GridVenezuela con un cluster de 9 procesadores, planificando aumentar a 64 procesadores en el corto plazo. En los clusters se ha instalado el sistema operativo SCIENTIFIC LINUX 3.0.7. En el nodo principal del cluster está instalado el CE de GLITE. Los otros nodos figuran como WNs. El único componente con que interactúan los usuarios es el UI. Para facilitar el uso del grid por parte de los usuarios e integrar las aplicaciones de cálculo intensivo (por ejemplo, AUTOSTRUCTURE), se ha instalado en el UI el portal grid GENIUS [12].

3. AUTOSTRUCTURE

El programa AUTOSTRUCTURE [3, 4, 5] es una versión actualizada y extendida del código de estructura atómica SUPERSTRUCTURE [3], ampliamente utilizado por la comunidad internacional de física atómica y astrofísica. Para un sistema iónico seleccionado por el usuario, AUTOSTRUCTURE calcula sus propiedades atómicas tales como los niveles energéticos, longitudes de onda y las probabilidades de transición tanto radiativas como de autoionización. Los problemas de utilización de SUPERSTRUCTURE tienen que ver con la gerencia de versiones, la estructura complicada de los datos de entrada/salida, rendimiento y documentación. Aunque AUTOSTRUCTURE ha mejorado los dos primeros puntos, especialmente el de los datos de entrada, a través de una interfaz basada en la estructura de FORTRAN NAMELIST, ésta puede ser simplificada y refinada con tecnología web. Este enfoque también se puede extender al archivo de salida que básicamente no se ha cambiado en 20 años de uso. Palacios & González [14] han desarrollado una interfaz de usuario web para AUTOSTRUCTURE basada en JAVA SERVER PAGES (JSP), la cual hemos integrado en el presente trabajo, a manera de proyecto piloto, al portal GENIUS.

4. GENIUS

El portal grid GENIUS [12] ha sido desarrollado para facilitar a los científicos el acceso, ejecución y

monitoreo de sus propios trabajos en el grid usando un simple navegador web. Como se aprecia en la Fig. 2, GENIUS está implementado sobre varios servicios del *middleware*.

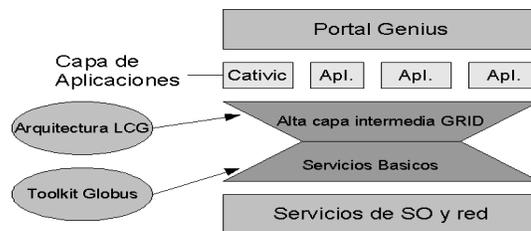


Fig. 2. Estructura general de GENIUS.

La estructura del portal se puede describir con un modelo de tres capas (Fig. 3):

1. El **cliente** es la estación de trabajo del usuario que contiene un simple navegador web;
2. El **servidor** es la máquina donde reside el UI de GLITE en concurrencia con los servicios de APACHE WEB, JAVA/XML FRAMEWORK ENGINFRAME [15] y GENIUS;
3. Los recursos remotos del **grid**.

El usuario puede acceder a sus archivos en el UI a través de los servicios de ENGINFRAME con capacidad de enviar trabajos al grid y manejar la data. La interfaz web reemplaza los complicados comandos de texto de GLITE, eliminando las limitaciones asociadas al sistema operativo, a la conexión y seguridad de red y a las del sitio y dispositivo de trabajo. El usuario puede, en efecto, interactuar con el grid desde cualquier sitio (oficina, habitación, hotel y cybercafé, entre otros) y con cualquier dispositivo (PC, PDA, teléfono WAP, etc.).

5. Integración de AUTOSTRUCTURE con GENIUS

GENIUS ofrece la posibilidad de integrar las interfaces graficas web de diferentes aplicaciones. Así el usuario sólo tiene que acceder a un portal maestro para enviar trabajos al grid, usando las mismas interfaces gráficas de las aplicaciones con las cuales está acostumbrado. La ventaja de este enfoque es que generaliza el motor de manejo del grid de GENIUS a todas las aplicaciones, en vez de tener que desarrollarlo para cada una de ellas. En el presente trabajo, a manera de proyecto piloto, se ha

investigado el grado de integración que se puede lograr para la interfaz web de AUTOSTRUCTURE.

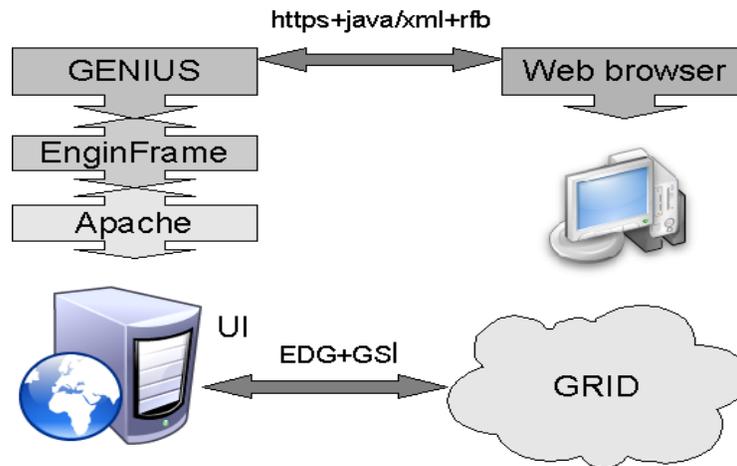


Fig. 3. Modelo de tres capas de GENIUS.

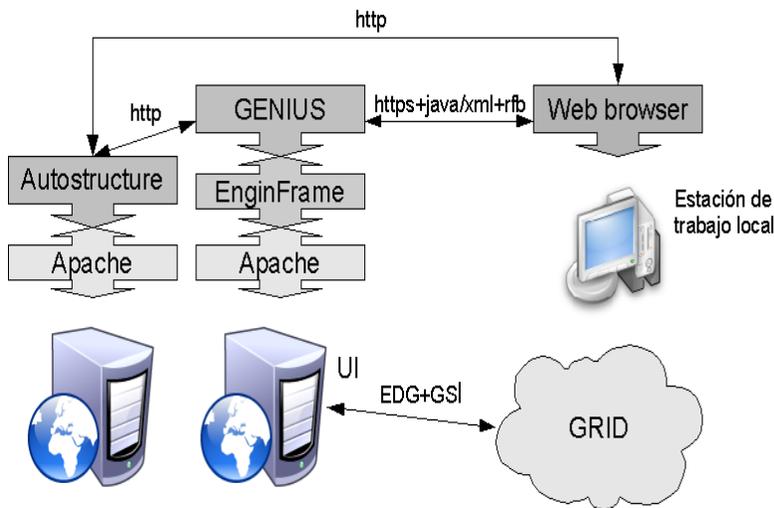


Fig. 4. Integración de AUTOSTRUCTURE con GENIUS.

El esquema de integración de AUTOSTRUCTURE con GENIUS se puede apreciar en la Fig. 4. En nuestro caso, la interfaz web de AUTOSTRUCTURE y el portal GENIUS se han instalado en el UI, pero también es posible instalar la primera en otro servidor web ya que ambos utilizan el protocolo HTTP para intercomunicación. Esta implementación es un ejemplo del esquema triangular propuesto por

Chaves et al. [16] para adaptar aplicaciones científicas que utilizan tecnología web en ambientes distribuidos.

En la Fig. 5 mostramos la portada de GENIUS desde cuya lista de aplicaciones a la izquierda se puede invocar a AUTOSTRUCTURE en una segunda pantalla. A la interfaz web original de AUTOSTRUCTURE se le ha agregado una opción para

enviar la tarea al grid (ver Fig. 5). Esta opción crea

los

archivos

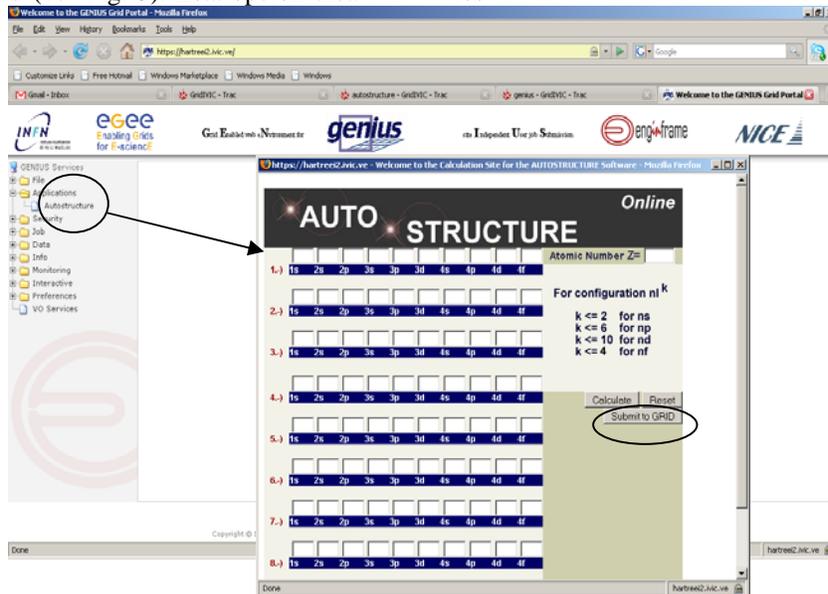


Fig. 5. Interfaz web de AUTOSTRUCTURE invocada desde el portal GENIUS.

de entrada necesarios para correr AUTOSTRUCTURE en el grid, el cual entonces llama a una función en GENIUS a través del protocolo HTTP. Finalmente, GENIUS baja los archivos necesarios al UI y envía la tarea al grid. También se encarga de monitorear el status de la tarea hasta su conclusión.

El primer paso para la integración es crear la posibilidad de abrir la aplicación en GENIUS. Como GENIUS está basado en ENGINFRAME, es fácil agregar esta opción editando un archivo de configuración XML del ENGINFRAME. También se puede indicar la función a llamar relacionando la opción agregada en el mismo archivo XML. Dicha función, en nuestro caso, abre la página web de AUTOSTRUCTURE. El segundo paso es adaptar la aplicación al ambiente grid. La aplicación web de AUTOSTRUCTURE originalmente genera los archivos de entrada y manda a ejecutar la aplicación AUTOSTRUCTURE. Se ha agregado la opción 'Submit to Grid' a la interfaz web de AUTOSTRUCTURE. La opción también genera los archivos de entrada pero los almacena en una carpeta en el servidor web de AUTOSTRUCTURE. Posteriormente, del lado del usuario se realiza un traslado automático a una función de GENIUS. Esta nueva función baja los archivos del servidor web de AUTOSTRUCTURE al servidor de GENIUS, y llama a la ya existente función 'job-submit' para enviar la tarea al grid.

Una de las dificultades que hemos encontrado es que el manejo de los archivos de entrada/salida dentro la interfaz web de AUTOSTRUCTURE se lleva a

cabo con un sistema de bases de datos (MYSQL), lo que hace necesario destacarle al usuario el momento cuando el control del proceso retorna del portal a la aplicación después de su ejecución. Una situación similar se presenta en la aplicación de química cuántica CATIVIC.

6. Conclusiones

La ventaja principal de integrar aplicaciones web con GENIUS es que generaliza su motor de manejo del grid a todas las aplicaciones, en vez de tener que desarrollarlo para cada una de ellas. El usuario se autentica en el grid sólo una vez por medio de GENIUS, y de la misma sesión del usuario se envían las tareas de otras aplicaciones al grid. Una mejora para el portal GENIUS podría ser un módulo de administración para facilitar la integración de interfaces web en vez de integrarlas manualmente.

La integración de aplicaciones con interfaces web al portal GENIUS no presenta muchas dificultades y facilita considerablemente el manejo del grid por parte del usuario. Entre las ventajas podemos citar las siguientes:

- No hay que aprender los múltiples y complicados comandos de texto para hacer uso del grid;
- No es necesario instalar software localmente sino disponer de un simple navegador web;

- Desde cualquier sitio y cualquier dispositivo (PDA, portátil, PC, teléfono móvil) se puede acceder al grid;

- GENIUS no aparenta tener limitaciones en relación al lenguaje de programación que se utilice para desarrollar la interfaz web (por ejemplo, JSP y PHP) lo que evita molestas transcripciones. La razón es porque toda comunicación entre GENIUS y las aplicaciones se lleva a cabo a través el protocolo HTTP.

Más aún, el presente enfoque se puede generalizar fácilmente a otras aplicaciones científicas (por ejemplo, CATIVIC), conformándose así en una metodología que estamos adoptando para adaptarlas a la nueva e-infraestructura.

Reconocimientos

Este trabajo ha sido parcialmente financiado por el FONACIT a través del Proyecto de Grupo G-97000667.

Referencias

- [1] Hey, T., Trefethen, A. 2003, The data deluge: an e-Science perspective, en *Grid Computing: Making the Global Infrastructure a Reality*, Berman, F., Fox, G, Hey, A.J.G. (Eds), Wiley, New York, p. 809
- [2] Foster, I., Kesselman, C. (Eds). 1999, *The grid: blueprint for a new computing infrastructure*, Morgan Kaufmann, San Francisco
- [3] Eissner, W., Jones, M., Nussbaumer, H. 1974, Techniques for the calculation of atomic structures and radiative data including relativistic corrections, *Comput. Phys. Commun.*, 8, 270
- [4] Badnell, N.R. 1986, Dielectronic recombination of Fe²²⁺ and Fe²¹⁺, *J. Phys. B*, 19, 3827
- [5] Badnell, N.R. 1997, On the effects of the two-body non-fine-structure operators of the Breit-Pauli Hamiltonian, *J. Phys. B*, 30, 1
- [6] Berrington, K.A., et al. 1978, A new version of the general program to calculate atomic continuum processes using the R-MATRIX method, *Comput. Phys. Commun.*, 14, 367
- [7] Mendoza, C., et al. 2007, OPSERVER: interactive online-computations of opacities and radiative accelerations, *Mon. Not. R. Astron. Soc.*, en prensa
- [8] Bautista, M. A., Kallman, T. R. 2001, The XSTAR atomic database, *Astrophys. J. Supl. Ser.*, 134, 139
- [9] Ruetter, F., Sanchez, M., Mendoza, C., Sierraalta, A., Martorell, G., Gonzalez, C. 2003, CATIVIC: parametric quantum chemistry package for catalytic reactions, *Int. J. Quant. Chem.*, 303, 96

- [10] Thomas, M.P., et al. 2005, Grid portal architectures for scientific applications, *J. Phys. Conf. Ser.*, 16, 596

- [11] Talia, D. 2005, Towards GRID Operating Systems: from GLinux to a GVM, en *Workshop on Network Centric Operating Systems*, http://www.cetic.be/coregrid/NCOS/papers/PP_Domenico_Talia.pdf

- [12] Barbera, R., Falzone, A., Rodolico, A. 2003, The GENIUS Grid Portal, en *Computing in High Energy and Nuclear Physics*, <http://www.slac.stanford.edu/econf/C0303241/proc/papers/TUCT001.PDF>

- [13] Laure, E., et al. 2006, Programming the Grid with GLITE, *Comp. Meth. Sci. Tech.*, 12, 33

- [14] Palacios, E.J., González, J.A. 2004, Paralelización de aplicaciones de cálculo intensivo e implementación de interfaces web para el entorno de trabajo colaborativo grid, Tesis de grado, Universidad de Carabobo

- [15] <http://www.nice-italy.com/main/index.php?id=32>

- [16] Chaves, J.L., et al. 2007, e-Science initiatives in Venezuela, en *Spanish Conference on e-Science Grid Computing*,

http://www.walc03.ula.ve/db/ssaber/Edocs/centros_investigacion/cat/publicaciones/papers/luisnunez/escience.pdf