

SISTEMA GNU-ESTUDIANTE. DISEÑO DE LA DISTRIBUCIÓN DE SU BASE DE DATOS

Resumen / Abstract

En una sociedad, la responsabilidad de formar a los profesionales del mañana recae en las universidades, es por ello que debe prestarse especial atención a los diferentes procesos que allí ocurren. El Ministerio de Educación Superior (MES) se ha planteado una estrategia de informatización que incluye la automatización de los procesos que tienen lugar en las universidades y sedes municipales del país. Es así que surge el Sistema GNU-Estudiente (Gestión de la Nueva Universidad), para dar respuesta a dicha problemática; este sistema, que actualmente se encuentra en desarrollo, está diseñado para trabajar con una base de datos distribuida, donde se hallará la información detallada de cada estudiante que hoy se encuentra cursando estudios universitarios, de forma tal, que pueda ser accedida en su centro y desde el propio MES.

Universities have the responsibility in the education of tomorrow's professionals therefore must care about different process taken place there. The Higher Education Ministry establish a strategy for computerize the automation of that process. Thereby emerge the "GNU-Estudiente" System (New University Management) to solve this problematical (academic management); the system, which is actually in development process, is design to work whit a distributed database which contain detail information about every university's student that can be access from the university and the ministry.

Palabras clave / Key words

Base de datos distribuida, gestión académica, SBDD, BDD, fragmentación, asignación, propagación de actualizaciones, réplica

Distributed database, academic management, SBDD, BDD, fragmentation, assignment, updates propagation, replica

INTRODUCCIÓN

En la sociedad cubana de principios de siglo, cada vez se hace más evidente el papel rector de la Educación Superior (ES) en la formación de los profesionales que tendrán sobre sí el futuro desarrollo del país.

Cada vez es mayor el número de personas que tienen acceso de una u otra forma a las universidades, esta es una tendencia que se incrementará con los años; luego, se impone un estricto control sobre el proceso (docente- educativo, metodológico e investigativo) que en los diferentes Centros de Educación Superior (CES) o Sedes Universitarias Municipales (SUM) ocurre. Este control puede llevarse manual (como actualmente ocurre en casi todo el país) pero está claro que si existiera un sistema que lo automatizara facilitaría el trabajo del personal involucrado en el mismo, contribuiría a mejorar el flujo de información entre las sedes y los CES a los que estas tributan, entre el MES y los diferentes CES y SUM del país, y dentro de cada CES y SUM; además se podría dar respuesta

Mariela Amaro López, Ingeniera Informática, Instructora, Centro de Estudios de Ingeniería de Sistemas (CEIS), Instituto Superior Politécnico José Antonio Echeverría Cujaa Ciudad de La Habana, Cuba
e-mail: mamaro@ceis.cujae.edu.cu

Recibido: octubre del 2005

Aprobado: diciembre del 2005

inmediata a peticiones informativas que se realizan al MES por la dirección del país.

Es por ello que el Ministerio de Educación Superior (MES) planteó la necesidad de automatizar la gestión universitaria teniendo como prioridad cero los procesos asociados a las secretarías docentes, que abarcan desde la matrícula de los estudiantes hasta su graduación, llevando el expediente académico del alumno, donde se registran los sucesos de su vida universitaria, en una base de datos distribuida (donde exista un sitio por cada secretaría subordinado a uno a nivel de CES) con sitio central en el MES que permita dar respuesta a las necesidades informativas que presenta ese ministerio.

El presente trabajo está orientado a describir el proceso de obtención de la distribución de la base de datos antes mencionada, es por ello que sus objetivos son:

1. Mencionar las características del sistema.
2. Obtener el diseño de la distribución de la base de datos.

SISTEMA GNU- ESTUDIANTE

El sistema surge como proyecto desde el año 2004, debido a una necesidad del MES, pero no es hasta el 2005 que empieza a desarrollarse, para lo cual se realizaron durante los meses de marzo y abril reuniones técnicas del equipo de desarrollo (compuesto por profesores y estudiantes de la Cujae, UCLV y UMCC), donde se discutió y acordó la plataforma a utilizar (J2EE, con el uso de los *enterprise java beans*) y el gestor de base de datos (PostgreSQL en su versión 8.0), todas herramientas Open Source que se ajustan a la política trazada por el gobierno cubano y el Ministerio de la Informática y las Comunicaciones (MIC); además, se definió que en el desarrollo del mismo se utilizaría el paradigma MDA (arquitectura guiada por modelos, en sus siglas en inglés, Model Driven Architecture).¹⁻¹¹

Principales características

Entre las características del Sistema GNU- Estudiante se encuentran:

- **División por módulos:** Se definieron una serie de módulos del sistema los cuales están representados en la tabla 1.
- **Arquitectura de *plugins*:** presenta este tipo de arquitectura donde la aplicación servidora queda reducida a un motor de ejecución de *plugins* sin ninguna funcionalidad inerte, donde cada *plugins* se ejecuta bajo las reglas definidas por el motor y por él mismo
- Manejo de la persistencia por el contenedor (en esta aplicación se utilizó el servidor de aplicaciones JBOSS): el contenedor conoce cómo comunicarse con el almacenamiento y genera el código necesario para ello
- Uso de *entity beans* (utilizados para representar las entidades del negocio) y *session beans* con estados (representan a un cliente dentro del servidor EJB)
- La seguridad es de dos tipos: programativa (existen creados roles, permisos y usuarios en la base de datos) y declarativa (a

través del API Java Authentication and Authorization Service (JAAS) en el servidor de aplicaciones).

- Base de datos distribuida (el diseño de esa distribución será explicado con detalle a continuación).

TABLA 1 Módulos que deben integrar el sistema	
No.	Módulo
1	Matrícula (incluye módulo para SUM)
2	Estadísticas (incluye solo nuevo ingreso)
3	Secretaría Docente
4	Plan de Estudio
5	Conexión y Réplicas
6	Profesor
7	Recuperación de Información Web
8	Estadísticas versión 2.0
9	Alumno ayudante
10	Estipendio
11	Estudiantes extranjeros
12	Planificación docente
13	Archivo histórico
14	Becas
15	Extracurricular
16	Guardia

DISTRIBUCIÓN DE LA BASE DE DATOS

Existen diferentes formas de hacer esa distribución, el estándar que se ha proyectado es: (Ver figura 1.)

- Un sitio en cada secretaría (de CES y SUM).
 - Un sitio a nivel de CES que almacenaría la información generada en todas las secretarías del CES y la información de aquellas carreras de las SUM subordinadas a esa universidad
 - Un sitio central a nivel de MES que almacenaría la información de cada CES.
- Otras de las variantes podrían ser:
- Prescindir del sitio a nivel de CES y que cada secretaría se subordine directamente al MES
 - Prescindir (en aquel CES donde exista un nivel elevado de conectividad) de los sitios en las secretarías y tener un sitio a nivel de CES contra el que trabajen las mismas
 - Tener un sitio por facultad (subordinado al del CES) para el que trabajen varias máquinas de una secretaría

- Tener un sitio a nivel de provincia (en la sede provincial del MES) subordinado al del MES que recibirá la información de cada CES y SUM de su provincia (esta variante es la menos probable en corto plazo pero se menciona pues pudiera ser interés del MES concentrar las informaciones a nivel provincial).

Cada una de las variantes descritas anteriormente estará contemplada en el diseño del sistema y la decisión de cuál escoger se dejará a elección de cada CES en el momento de la implantación. En el caso de las SUM, como el nivel de conectividad que presentan (hasta el momento) es muy escaso, solo tendrán la variante estándar, es decir, tendrán un sitio en cada secretaría de SUM subordinado al CES (puede ser más de uno y está en dependencia de la carrera a la que se refiera) y pudieran enviar sus réplicas directamente al MES.

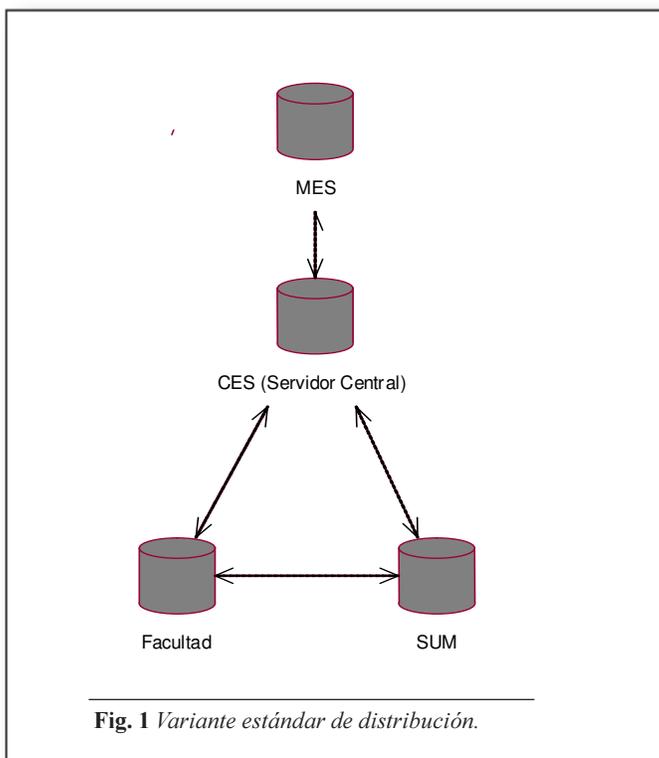


Fig. 1 Variante estándar de distribución.

Al estar la base de datos distribuida debe garantizarse la fiabilidad de la información almacenada en cada sitio, además de mantener la consistencia e integridad de los datos y plantear mecanismos de seguridad que restrinja el acceso a los mismos solo al personal autorizado (muchos de estos mecanismos quedan de parte del administrador del sistema en cada CES y SUM).

Con anterioridad se analizaron distintas variantes de donde ubicar los sitios del sistema distribuido, ahora se analizará qué información debe existir en cada sitio.

Fragmentación y asignación

Para el Sistema GNU- Estudiante la fragmentación que se utilizó fue la horizontal y los criterios para la división fueron los siguientes:

1. Dividir la información de estudiantes según la carrera

- a) En el caso de los CES en cada secretaría de facultad se ubicarán solo aquellos estudiantes que estudien carreras de esa facultad y aquellos estudiantes de una SUM cuya carrera se subordine metodológicamente a una carrera de esa facultad.

- b) En una SUM solo estarán aquellos estudiantes que estudien carreras pertenecientes a la misma.

2. La información de tipos de curso se dividirá según el centro

- a) En cada secretaría de un CES estarán todos los tipos de curso existentes (Regular Diurno, Para Trabajadores, Continuidad de estudios).

- b) En una SUM solo existirá un tipo de curso (Continuidad de estudios).

3. La información de facultades se ha dividido por centro, de forma tal que en cada uno solo esté la información relativa a las facultades que lo componen.

4. Las carreras se dividen por facultades o sedes, en cada facultad o sede solo estarán las carreras que en ella se estudien.

5. Los planes de estudio con las asignaturas y disciplinas que los conforman se han dividido por carrera y tipo de curso.

En cada sitio del sistema se encontrará el mismo esquema de base de datos, pero la información que se ubica depende de las divisiones realizadas anteriormente. En cada caso, el sitio central del CES (si existe) contendrá las informaciones relativas a todas las facultades que lo compongan (toda la información que se genere en la secretaría de cada facultad) y la de las SUM que estén subordinada metodológicamente a alguna facultad del CES. En el sitio del MES estará una copia de la información de cada CES y SUM del país.

Réplicas y propagación de actualizaciones

Para el Sistema GNU- Estudiante se han definido varias políticas:

1. Perfiles de réplicas estándares y posibilidad de crear nuevos perfiles (el módulo 5 brindará una herramienta para estos efectos a los administradores de cada sitio).

- Los perfiles de réplicas definen cuáles son las tablas (brindando la posibilidad de crear filtros, es decir, fragmentar las tablas) que van a replicarse de un sitio al otro. Los estándares serán los utilizados siempre que se replique a no ser que se especifique algún otro perfil. Algunos de los que están definidos son:

- La Dirección de Estadística del MES es la responsable de los codificadores del sistema, una vez definidos estos deben replicarse en todos los sitios centrales de los CES (con la división que se analizó anteriormente, en estos sitios estarán las copias primarias), estos replican a las facultades y a las sedes que se le subordinen administrativamente. Si ocurre algún cambio este debe propagarse a cada sitio implicado. Si no existiera un sitio central en el CES entonces se cambiaría el perfil de réplica por uno nuevo (esto se definiría en la instalación).

- Las sedes replicarán sus datos, según la carrera, a los CES a los que la carrera se subordine metodológicamente.

- Las facultades replicarán sus datos al sitio central del CES
- El sitio central de cada CES replicará al MES todos los datos recibidos de las secretarías de las facultades que lo componen y los que recibe de las sedes.

- La creación de nuevos perfiles es una necesidad y constituye una política de contingencia ante la pérdida de información sensible, por ejemplo: una facultad o sede necesita que el nivel superior replique hacia ella todos sus datos o fragmentos de estos (al existir pérdida de información por roturas en los soportes en los que se almacenó o debido a algún fenómeno natural -lugares propensos a inundaciones, temblores de tierra, descargas eléctricas, etc.-, entre otras).

2. Momento y frecuencia (está en dependencia del nivel de conectividad presente en cada sitio) con que se realizarán las réplicas (esto debe ser definido en la instalación del sistema y podrá cambiarse posteriormente si así se desea, solo por los administradores del sistema).

3. Posibilidad de replicar la base de datos en diferentes dispositivos de almacenamiento y luego recuperar la misma hacia otra base de datos para hacer la consolidación.

- Esta es una necesidad pues al ser la base de datos distribuida nacionalmente las réplicas estarían dependiendo del nivel de conectividad existente en cada lugar y definiendo esta posibilidad se disminuye esa dependencia. Con los problemas de conectividad podrían darse las siguientes situaciones:

- **Baja conectividad:** Muchos centros (como las SUM, algunas facultades de la Universidad de La Habana (UH), de la Universidad Central de Las Villas (UCLV) que se encuentran distantes de la universidad) realizan su conexión a través de *modems*, esto puede ocasionar que la conexión falle y no termine el proceso de replicación.

- **Cero conectividad:** Pueden existir centros donde no halla conexión o puede darse el caso de una secretaría que esté conectada y falle la red en el momento de realizar la réplica.

- Se han previsto además soluciones en las cuales, si la réplica no se realiza completa, se brinde posibilidad de terminarla en un momento posterior.

4. Para cada tupla de una relación existirá un **identificador único global** (*Global Unique Identifier, GUID*) que será una llave física, de forma tal que no existan llaves duplicadas lo que pudiera traer problemas a la hora de consolidar las réplicas en las bases de datos de los CES y el MES.

5. Se dará la posibilidad de realizar respaldos de la base de datos en el momento en que el usuario quiera y hacia el soporte magnético que desee; junto con esto podrán realizarse recuperaciones de esos datos.

CONCLUSIONES

El presente trabajo abordó algunas características del Sistema GNU-Estudiente, así como el diseño de la distribución de su base de datos, de este puede concluirse que:

- El Sistema GNU-Estudiente permitirá la automatización de los procesos que ocurren actualmente en las universidades y

sedes municipales del país. Brinda prioridad cero a los vinculados con la gestión docente (Matrícula, Plan de Estudio, Control Docente y Estadísticas) y distribuye la información detallada de cada estudiante de forma tal que pueda ser consultada en su centro o en el MES.

- El diseño de la distribución de la base de datos garantizará contar en cada sitio exactamente con la información que en él puede ser solicitada y en el sitio central concentrar la información de todo el país para realizar estudios estadísticos, históricos y dar respuesta a las necesidades informativas del MES.

- El proceso de réplica garantizará la consistencia de las bases de datos de cada sitio propagando las actualizaciones en cualquier sentido (desde el MES hacia los CES y SUM y al contrario), y prevé soluciones ante fallos o diferentes niveles de conectividad.

- La definición de perfiles estándares de réplica garantizará la transparencia de este proceso para los usuarios finales de la aplicación y cuando se desee modificar los datos a replicar podrán crearse nuevos perfiles. ☐

REFERENCIAS

1. **BIRSAN, DORIAN:** "On Plugins and Extensible Architectures", *ACM Queue Architectures Tomorrow's Computing*, Vol. 3, No. 2, March, 2005.
2. **CUESTA QUINTER, CARLOS ENRIQUE:** "Arquitectura de software dinámica basada en reflexión", Tesis de Doctorado. Departamento de Informática (ATC, CCIA, LSI), Universidad de Valladolid, Julio de 2002.
<http://descargas.cervantesvirtual.com/servlet/SirveObras/01361653144571409089802/010349.pdf> (enero, 2005)
3. **DATE, C. J. :** *Introducción a los sistemas de bases de datos*. 3ra. parte, Ed. Félix Varela, Ciudad de La Habana, 2003.
4. **ECKEL, BRUCE:** *Thinking in Java*, third edition, Prentice Hall, New Jersey, 2003.
5. **HERNÁNDEZ, ANAISA:** *Base de datos distribuidas*, (Bibliografía de la Asignatura Bases de Datos de la Maestría en Informática Aplicada) \\ceis\clases\postgrado\Maestría en Informática Aplicada\11na edición (Noviembre 2004)\BD, abril, 2005.
6. **IGLESIAS, EVA L.:** *Bases de datos distribuidas*. Ampliación de BD. ESEI. Universidade de Vigo, <http://trevinca.ei.uvigo.es/~evali/abd/BDDWeb.pdf>, mayo, 2005.
7. **MARTÍNEZ PRIETO, JULIO:** "Informatización de la Gestión Universitaria", Propuesta Preliminar, Informe resultante del Taller Nacional Informatización de la Gestión Universitaria. Ciudad de La Habana, 12 de marzo del 2004.
8. **FOWLER, MARTIN:** Who Needs an Architect? Martin Fowler, Editor: *IEEE Software*, July/August, 2003.
9. *Inc. Model Driven Architecture (MDA)*. Object Management Group, Joaquín Miller and Jishnu Mukerji, ed., 9 de julio de 2001.
10. *Inc. MDA Guide Version 1.0.1*, Object Management Group, ed., 12 de junio, de 2003.
11. *Inc. MDA Specifications*, Object Management Group, May 2005, <http://www.omg.org/mda/specs.htm>