

DEMO_REP: HERRAMIENTA PARA LABORATORIOS EN LA ENSEÑANZA DE RÉPLICA DE DATOS

AUTORES: Anthony Rafael Sotolongo León¹

Ileana Martí Pérez²

DIRECCIÓN PARA CORRESPONDENCIA: Universidad de la Ciencias Informática (UCI), La Habana, Cuba, Email: asotolongo@uci.cu

Fecha de recepción: 10 - 06 - 2012

Fecha de aceptación: 11 - 01 - 2013

RESUMEN

El proceso de formación de los profesionales en tecnologías de la informática es un aspecto clave en el desarrollo del país. La Universidad de las Ciencias Informáticas es uno de los principales centros que se dedica a este proceso, la misma consta de programas de formación postgraduada para profundizar en temas que no aparecen incluidos en los programas analíticos de las asignaturas de pregrado. Uno de estos programas es el "DIPLOMADO EN TECNOLOGÍAS DE BASES DE DATOS POSTGRESQL", el cual cuenta con una serie de cursos para lograr experticia en la temática de las bases de datos en el gestor PostgreSQL. En el presente trabajo se muestra cómo incide en la comprensión del proceso de la replicación de datos, la utilización de una herramienta informática "Demo_rep" para realizar los ejercicios de laboratorio de réplica. Se describe además las principales funcionalidades de la misma.

PALABRAS CLAVE: Herramienta informática, PostgreSQL, Laboratorio de réplica de datos

DEMO_REP: A LABS TOOL FOR TEACHING DATA REPLICATION

ABSTRACT

The process of training professionals in information technologies is a key aspect in the development of the country. University Informatics Science is one of the main centers dedicated to this process, also has postgraduate training programs to further advanced topics not mentioned in the pregraduate. One such program is the "DIPLOMADO EN TECNOLOGÍAS DE BASES DE DATOS POSTGRESQL," which has a number of courses to obtain expertise in the subject of databases in PostgreSQL Manager. In this paper show how affects in the understanding of the process of data replication, the use of the software

¹ Graduado de ingeniero informático en la Universidad de Cienfuegos(UCF) en el 2006, labora en la Universidad de Ciencias Informáticas(UCI) como profesor del Dpto de PostgreSQL del centro DATEC de la facultad 6, obtuvo el grado de Máster en ciencias en el 2010 , además es profesor asistente.

² Graduado de Ingeniería en Ciencias Informáticas en la Universidad de de las Ciencias Informáticas en el 2007, labora en la Universidad de Ciencias Informáticas(UCI) como profesor del Dpto de Sistemas Digitales de la facultad 6, y posee categoría docente de Instructor.

tool "Demo_rep" for laboratory exercises replication. It also describes the main features of software tool.

KEYWORDS: Informatics Tool, PostgreSQL, Data Replication labs

INTRODUCCIÓN

En la actualidad las bases de datos constituyen una parte importante del desarrollo de múltiples aplicaciones. Sin embargo, la concepción e implementación de las bases de datos continúa siendo una tarea costosa y difícil en la cual no debe improvisarse, sobre todo en entornos críticos. Para enfrentar estas situaciones se ha incluido en la formación de los profesionales planes de estudio relacionado con el diseño e implementación de bases de datos, fundamentalmente en las carreras afines a la informática.

En investigaciones como (IÑIGO, 2003), (GONZÁLEZ, 2007, 2011) se manifiesta que el desarrollo de las capacidades profesionales básicas en los estudiantes graduados cubanos tiene aún que avanzar para lograr la graduación de un profesional competente. Lo que constituye un reto educacional, en especial en la educación superior. En este sentido la educación de postgrado es una de las direcciones principales de trabajo de la Educación Superior en Cuba, dirigida a promover la educación permanente de los graduados universitarios. Una de las instituciones que tiene en cuenta estos aspectos es la Universidad de las Ciencias Informáticas (UCI), en la cual se ha desarrollado un diplomado para profundizar en temas avanzados relacionados con las bases de datos, el mismo se titula: "DIPLOMADO EN TECNOLOGÍAS DE BASES DE DATOS POSTGRESQL". Uno de los temas que se imparten es la replicación de datos (SOTOLONGO, VAZQUEZ, 2011), aspecto eficiente y estratégico para mantener disponibles los datos en los sistemas en producción (BRAYNER, 2009).

El proceso de enseñanza de esta temática de las bases de datos *puede resultar complicado a la hora de realizar prácticas de laboratorio* pues las soluciones de réplica que existen son robustas y tienden a ser costosas a la hora de configurar, pues están desarrolladas en diferentes lenguajes de programación y sobre todo para entornos de producción tales como SLONY-I (SLONY-I, 2011), pgpool-II (PG-POOL-II, 2011) y la replicación nativa de PostgreSQL: "Hot Standby-Streaming Replication" (PGDG, 2010). *Esto imposibilita a los estudiantes comprender mejor el proceso de replicación.* Teniendo en cuenta que el término "comprender" en la educación es: "pensar y actuar flexiblemente en cualquier circunstancia a partir de lo que se sabe acerca de algo" (Pogré, Lombardi, 2004), significaría en este contexto que los estudiantes sean capaces de analizar cada parte del proceso de replicación, puedan tomar decisiones de acorde a las necesidades de la situación y ejecutar dicha decisión.

El presente trabajo tiene como objetivo presentar las experiencias y resultados obtenidos en el curso de réplica de datos del "DIPLOMADO EN TECNOLOGÍAS DE BASES DE DATOS POSTGRESQL" en su segunda versión, al aplicar en las

clases de laboratorio la herramienta “Demo_rep” y determinar si tiene incidencia en la mejora de la comprensión del proceso de replicación.

DESARROLLO

El laboratorio es un tipo de clase, lo cual queda expresado en la Resolución No. 210/07 del Reglamento del Trabajo Docente y Metodológico en la Educación Superior (RESOLUCIONES, 2007):

ARTÍCULO 112: La práctica de laboratorio es el tipo de clase que tiene como objetivos que los estudiantes adquieran las habilidades propias de los métodos y técnicas de trabajo y de la investigación científica; amplíen, profundicen, consoliden, generalicen y comprueben los fundamentos teóricos de la disciplina mediante la experimentación, empleando para ello los medios necesarios.

En las carreras técnicas, particularmente las relacionadas con la informática los tipos de clase utilizadas en el postgrado básicamente son: la conferencia, la clase práctica, el seminario, la práctica de laboratorio y el taller. La replicación de datos como curso dentro del diplomado tiene en cuenta en el diseño de su programa analítico estos aspectos. En los temas teóricos se tratan materias como (SOTOLONGO, VAZQUEZ, 2011):

- Heterogeneidad (homogénea y heterogenia).
- La captación de los cambios (basaba en disparadores y minería de registros).
- Dirección de transmisión de los datos (maestro-esclavo y maestro-maestro).
- Forma de transmisión de los datos (síncrona y asíncrona)
- Fragmentación de los datos.

En los aspectos prácticos se orientan ejercicios relacionados con (SOTOLONGO, VAZQUEZ, 2011):

- Implementación de técnicas de detección de cambios basados en disparadores.
- Implementación de mecanismos de fragmentación de datos horizontal y vertical.
- Implementación de mecanismos que capaces de lograr heterogeneidad.

Para realizar las prácticas de laboratorio se desarrolló por el grupo de profesores del diplomado una herramienta que permite realizar sobre la misma los aspectos prácticos mencionados anteriormente. Esta herramienta con nombre “Demo_rep” está desarrollada en el lenguaje java y utiliza un script SQL con tablas, funciones y disparadores en plpgsql, lenguaje que brinda PostgreSQL para el desarrollo de rutinas del lado del servidor.

“Demo_rep” es una aplicación que realiza replicación asíncrona, maestro-esclavo, utilizando disparadores para la captación de los cambios, sobre servidores PostgreSQL, está orientada totalmente a entornos docentes, para ejemplificar aspectos teóricos de la replicación y permitir su modificación,

además realiza la replicación bajo demanda (a petición del usuario y/o automática cada 5 segundos).

En el script SQL se encuentra la creación de un esquema llamado “replica” el cual tiene una tabla con nombre “replica.registro_replica” que es la que almacena la información que se va a replicar.

```
CREATE TABLE "replica".registro_replica
(id_registrito serial NOT NULL, -- id
tabla character varying(50), -- nombre la tabla modificada
sql text, -- sentencia SQL
replicado boolean DEFAULT false, -- estado del registro (por defecto "false")
usuario character varying(20), -- usuario que provocó el cambio en la tabla
fecha date, -- fecha que se capturó el cambio
operacion character varying(10) -- operación que se capturó (INSERT, UPDATE, INSERT));
```

También hay una función “replica.demo_rep_generar_trigger(anombretabla text)” con la cual se generan los disparadores(triggers) necesarios para las tablas que desee, a esta función se le debe pasar los nombres de las tablas que desea replicar separado por comas: 'tabla1,tabla2' , y con ella se realiza la configuración de las tablas a replicar.

Los disparadores generados dinámicamente hacen un llamado a la función disparadora "replica.tr_demo_rep_registro()" que es la encargada de almacenar los cambios en la tabla “replica.registro_replica”(tabla de control). Esta tabla es la que consulta la aplicación “Demo_rep” para ver qué cambios se han realizado y los mismos los envía para el servidor esclavo o permite exportar a un archivo *.sql para replicar en entornos desconectados.

Por último cuenta con una función "replica.demo_rep_limpiar_registro()" que se encarga de eliminar los registros replicados de la tabla “replica.registro_replica”.

La interfaz de la aplicación se muestra en la Figura 1. La misma cuenta con un área para configurar la conexión a las bases de datos PostgreSQL de los servidores maestros y esclavo de respectivamente. También en la parte inferior se encuentran los botones que le permiten al usuario realizar la replicación manual o por exportación de archivo *.sql, además brinda la opción de ejecutar réplica automáticas cada 5 segundos.

A los estudiantes se les muestra dicha herramienta y se explica su funcionamiento comenzado por detallar la estructura del script SQL que está escrito un lenguaje ya conocido y común(sql y plpgsql), además se les exigen que realicen cambios pertinentes en las estructuras de tablas y funciones dentro del mismo script SQL para lograr los tres los aspectos prácticos mencionados anteriormente, con ideas propias, demostrando así sus habilidades de programadores y si han sido capaces de comprender el proceso de replicación, ajustándolo a sus necesidades.

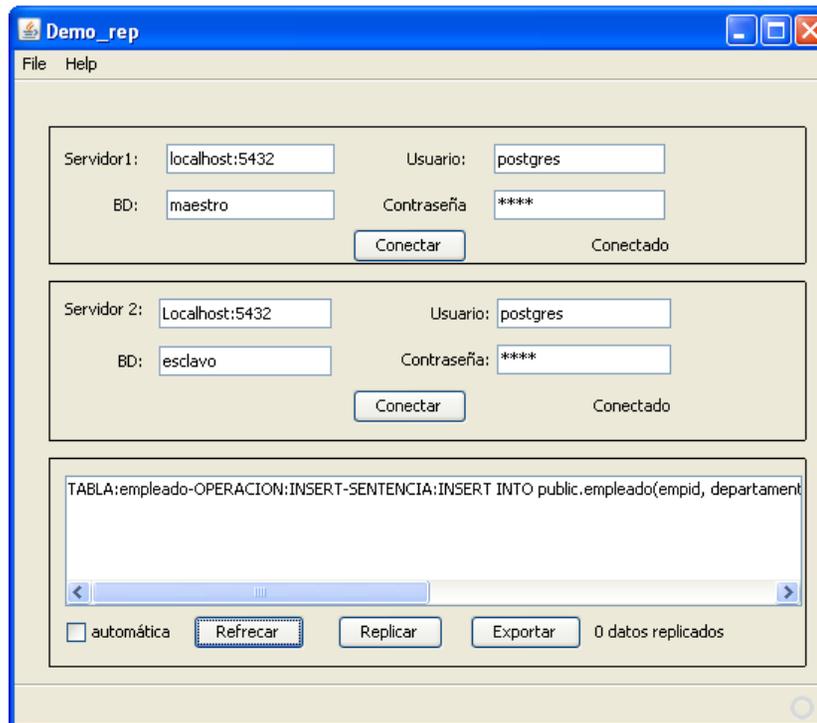


Figura 1: Interfaz de la aplicación Demo_Rep

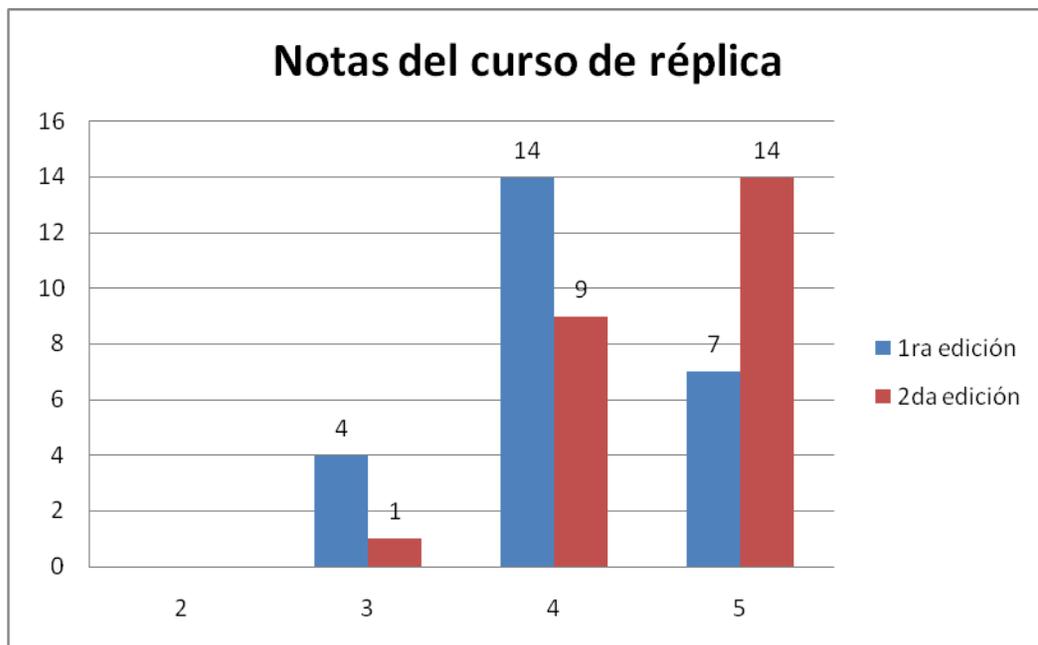
Algunos de los ejercicios que se les orientan a los estudiantes son:

- Permitir detectar cuando se realiza una operación “truncate” sobre las tablas a replicar y seleccionar la operación a replicar.
- Lograr realizar fragmentación horizontal sobre los registros que cumplan determinada condición.
- Realización de transformaciones de los datos a replicar.

El diplomado se ha impartido en dos ediciones, la primera para 25 estudiantes en la escuela de verano de la UCI 2011 donde se utilizó la replicación nativa de PostgreSQL para los laboratorios, la composición de este grupo es de 19 graduados de la carrera de Ingeniería en Ciencia Informáticas [graduados UCI], 4 graduados de Ingeniería Informática, 2 de Matemática Computación [cibernética], donde todos eran de graduaciones posteriores al año 2006 y la segunda fue para 24 estudiantes en una edición especial en Santiago de Cuba en febrero 2012 donde si se utilizó la herramienta “Demo_rep” para las clases de laboratorios, la composición de este grupo es de 20 graduados de la carrera de Ingeniería en Ciencia Informáticas [graduados UCI], 3 graduados de Ingeniería Informática, 1 de Matemática Computación [cibernética], donde todos eran de graduaciones posteriores al año 2006 .

Para determinar si la aplicación de la herramienta “Demo_rep” se midieron dos variables, *calidad de las notas y cantidad de tesinas relacionadas con la temática de réplica*. A continuación se muestran los resultados:

La Gráfica 1 muestra la *calidad de las notas* de las ambas ediciones del curso de réplica.



Gráfica 1: notas del curso de réplica.

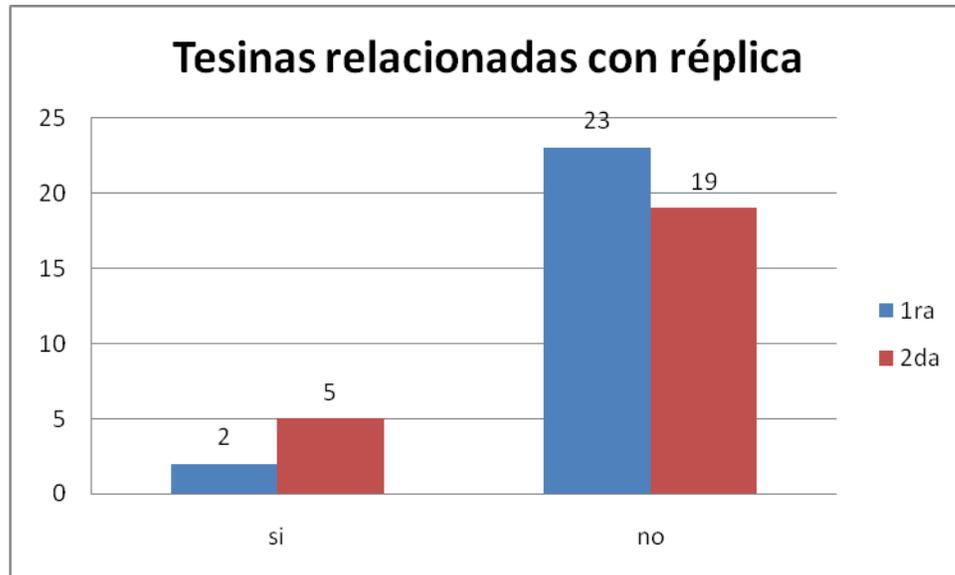
Los exámenes aplicados fueron similares en cuanto a su diseño donde se les planteaba una situación y debían describir las características de la herramienta que supiera las necesidades del problema y decir de las estudiadas en clase cual cumplía con dichas características (anexo 1 y 2).

Se puede observar en la gráfica 1 que el número de la menor nota obtenida, 3 en este caso, se redujo en un 75 % de la primera a la segunda edición, y la mayor nota, 5 en este caso, mejoró en un 100 % de la primera a la segunda edición, es decir que la calidad de las notas mejoró.

La Gráfica 2 muestra la cantidad de tesinas relacionadas con la temática de réplica.

Para la graduación del diplomado se debía presentar una tesina donde presentaran resultados de investigaciones propias donde se estuvieran contenidos de al menos dos cursos del diplomado. Se puede observar en la gráfica 2 que el número de tesinas que incluyeron temas de réplica aumentó en un 150 % de la primera a la segunda edición, es decir que el interés por desarrollar investigaciones con temas de réplica aumentó.

Los resultados anteriores pueden argumentar que los estudiantes de la segunda edición, donde se aplicó la herramienta “Demo_rep” tuvieron una mejor comprensión del proceso de replicación de datos.



Gráfica 2: Tesinas relacionadas con réplica.

CONCLUSIONES

Habiendo identificado que las herramientas existentes para realizar el proceso de replicación en PostgreSQL, no responden a necesidades docentes, sino más bien a entornos productivos, lo cual imposibilita en mejor entendimiento del proceso, propició desarrollar alternativas que pudieran utilizarse para resolver dicha situación. La utilización de la herramienta de réplica “Demo_rep” para PostgreSQL, permitió a los estudiantes de la segunda edición del “DIPLOMADO EN TECNOLOGÍAS DE BASES DE DATOS PostgreSQL”, tener un mejor comprensión del proceso de replicación.

BIBLIOGRAFÍA

IÑIGO, E. (2003). "El proceso de reproducción social de los jóvenes profesionales en cuba. Análisis de su impacto en los graduados de 1996 al 2000". Informe al Ministerio de Educación Superior de Cuba. Centro de Estudios para el Perfeccionamiento de la Educación Superior. Universidad de La Habana. Ciudad de La Habana. 20 pp.

GONZÁLEZ, G. (2007)"Experiencias en la formación de competencias profesionales en estudiantes de ingeniería mecánica mediante el desarrollo de proyectos de cursos". En: Memorias de las Primeras Jornadas de Innovación Educativa, España 2007. 8 pp. Referido en línea en: <http://dialnet.unirioja.es/servlet/articulo?codigo=2901942> [consulta: 22 febrero 2011].

GONZÁLEZ, G. (2011). "La actividad de normalización y el desarrollo de competencias profesionales en estudiantes de ingeniería". Revista Cubana de Ingeniería, vol. 2, núm.1, pp. 13-20. Ciudad de La Habana.

POGRÉ, P. y Lombardi, G (2004) , “Escuelas Que Enseñan a Pensar”, Buenos Aires, Argentina pp 40,

SOTOLONGO, A y VAZQUEZ, Y (2011). Diplomado en tecnologías de bases de datos PostgreSQL. 2011 (CD-ROM) En: MEMORIAS DE Universidad 2012 a nivel provincial en la CUJAE). La Habana.

BRAYNER, J. M. Monteiro, S. Lifschitz (2009). Using transaction isolation levels for ensuring replicated database consistency in mobile computing environments, Proceedings of the Eighth ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE '09) pp. 85-91

SLONY-I (2011). Slony-I , Enterprise-level replication system. En línea disponible en: <http://slony.info/> [consultado el 1 mayo 2012]

PG-POOL-II (2011). Main Page. En línea disponible en: http://www.pgpool.net/mediawiki/index.php/Main_Page [consultado el 1 mayo 2012]

PGDG (2010). PostgreSQL 9.0.0 Documentation, High Availability, Load Balancing, and Replication (pp. 548-568)

RESOLUCIONES (2007). Reglamento sobre el Trabajo Docente y Metodológico donde se norma el Trabajo Metodológico, el Trabajo Docente, la Evaluación del Aprendizaje y el Trabajo con los Alumnos Ayudantes. Cuba (en línea). Disponible en: <http://www.matcom.uh.cu/Portals/0/recursos/Resolucon%20210-2007.doc> [Consultado el 11 de mayo de 2012]

ANEXOS

Anexo 1. Examen de la primera edición

Se tiene una solución de software que necesita implementar un sistema de replicación de datos entre sus servidores, debido a que requiere que sus bases de datos tengan la misma información, el sistema debe ser desplegado en la capital de la provincia de Holguín, se tendrá un nodo de la base de datos en el reparto “Vista alegre” y el mismo está distante a 1 Km de el centro de la ciudad donde está otro nodo y por último otro servidor en el aeropuerto a unos 10 Km del centro.

Se necesita que los datos contenidos en el servidor de vista alegre y del aeropuerto se repliquen hacia el servidor del centro de la ciudad. Fungiendo el servidor del centro de la ciudad como receptor de la información de los demás. Y este servidor central no envía ninguna información a los otros nodos es solo para que los directivos consulten y vean el comportamiento de las entidades relacionadas.

Es necesario tener en cuenta que la comunicación entre el aeropuerto y el centro de la ciudad es afectada por interferencias generadas por radares de una unidad militar que se encuentra cerca del aeropuerto, lo que hace que corte la comunicación por pequeños espacios de tiempo.

- A) Diga las características que debe tener una herramienta de réplica que supla estas necesidades.

Anexo 2. Examen de la primera edición

Se tiene una solución de software que necesita implementar un sistema de replicación de datos entre sus servidores, debido a que necesita que sus bases de datos tengan la información requerida, el sistema debe ser desplegado en la capital de la provincia de Santiago de Cuba, se tendrá un nodo de la base de datos en el reparto “Boniato” y el mismo está distante a unos 2 Km de el centro de la ciudad donde está el nodo principal y por último otro servidor en el aeropuerto a unos 6 Km del centro.

Se necesita que los datos modificados en el servidor principal se repliquen a los servidores de vista alegre y del aeropuerto pero solo los que le incumbe a cada uno de ellos. Además debe permitir enviar información modificada en los servidores de “Boniato” y del aeropuerto hacia en central.

Es necesario tener en cuenta que la comunicación entre los nodos y el centro de la ciudad es perfecta y no existe interrupciones.

- A) Diga las características que debe tener una herramienta de réplica que supla estas necesidades.

