



Revista Perspectivas

Rendimiento de herramientas de replicación libres para la implementación de un sistema integrado de alta disponibilidad orientado a la gestión de tutorías académicas en Instituciones de Educación Superior

Performance of free replication tools for the implementation of an integrated high availability system oriented to the management of academic tutoring in Higher Education Institutions

Diego Palacios Campana ^{1,2}, Gonzalo Allauca Peñafiel ², Danny Velasco Silva ², Estuardo Cajilema Guamán ², Rubén Saltos Chávez ¹.

¹ Escuela Superior Politécnica de Chimborazo, Dirección de Tecnologías de la Información y Comunicación, Riobamba, Ecuador.

² Universidad Nacional de Chimborazo, Facultad de Ingeniería, Riobamba, Ecuador.

Email: dpalacios@esPOCH.edu.ec, gallauca@unach.edu.ec, dvelasco@unach.edu.ec, ecajilema.fis@unach.edu.ec, ruben.saltos@esPOCH.edu.ec

Resumen- Se realizó un análisis comparativo de las herramientas de replicación de datos Slony-I y PgPool-II, para seleccionar la mejor alternativa en la implementación de un sistema integrado de alta disponibilidad para la gestión de tutorías académicas en Instituciones de Educación Superior. Se establecieron como parámetros de estudio el tiempo de respuesta frente a la ejecución de sentencias SQL el uso de CPU, memoria, y disco duro. Se desarrolló una aplicación web N-Capas utilizando PrimeFaces y PostgreSQL, integrada a un Sistema Académico a través de servicios web SOAP. Para la realización de las pruebas se configuraron servidores de base de datos de tipo master y esclavo bajo un entorno Linux. Las pruebas de rendimiento estuvieron basadas en la herramienta Apache JMeter, simulando 282 peticiones por segundo en un instante de tiempo, observándose que PgPool-II consumió menos recursos con un promedio de 2,80%, sin embargo, Slony-I fue más eficaz al realizar las transacciones con apenas un 0,009% de error.

Palabras Clave- Replicación, Slony, PgPool, PrimeFaces, PostgreSQL

Abstract- A comparative analysis of the Slony-I and PgPool-II data replication tools was carried out, to select the best alternative in the implementation of an integrated system of high availability for the management of academic tutoring in Higher Education Institutions, established as Study parameters the response time against the execution of SQL statements, the use of CPU, memory, and hard disk. An N-Layers web application was developed using PrimeFaces and PostgreSQL, integrated into an Academic System through SOAP web services. To carry out the tests, master and slave database servers were configured under a Linux environment. The performance tests were based

on the Apache JMeter tool, simulating 282 requests per second in an instant of time, observing that PgPool-II consumed less resources with an average of 2.80%, however, Slony-I was more effective at carry out transactions with just 0.009% error.

Keywords- Replication, Slony, PgPool, PrimeFaces, PostgreSQL

I. INTRODUCCIÓN

En el ámbito de las bases de datos distribuidas, la replicación es un conjunto de tecnologías destinadas a la copia y distribución de datos y objetos de una base de datos a otra, para luego sincronizar ambas bases de datos y lograr mantener su coherencia. La réplica permite distribuir datos a diferentes ubicaciones y a usuarios remotos o móviles mediante redes de área local y de área extensa, conexiones de acceso telefónico, conexiones inalámbricas e internet [1].

La replicación es un proceso clave para lograr el comportamiento exitoso de las bases de datos, ya que su funcionalidad ayuda a garantizar la consistencia de los datos y permite que el motor de base de datos siga funcionando en caso de fallas de la red. Esto implica copiar y mantener cada manipulación de datos desde una ubicación a los otros nodos que se actualizarán. Por lo tanto, la configuración de la red y la comunicación de nodos tienen un gran impacto en el rendimiento de la replicación. Sin embargo, los modelos internos para la replicación provistos en la mayoría de los motores carecen de flexibilidad y sufren otros problemas como las comunicaciones de cuello de botella. Actualmente existe una diversidad de herramientas que facilitan el proceso

de replicación de datos logrando mejorar el rendimiento y proteger la disponibilidad de las aplicaciones, ayudando al auge, cada día más creciente, de sistemas de base de datos distribuidos. Una correcta identificación de las necesidades de replicación asegura la consistencia de los datos y disminuye los errores de replicación [2] [3].

Slony-I es un sistema de replicación maestro-esclavo que incluye todas las características y capacidades necesarias para replicar grandes bases de datos a un número razonablemente limitado de sistemas esclavos. Es un sistema diseñado para su uso en centros de datos y sitios de respaldo, donde el modo normal de operación es que todos los nodos están disponibles [4].

PgPool-II es un middleware que se ubica entre los servidores y un cliente de base de datos PostgreSQL, proporcionando las siguientes características: agrupación de conexiones, equilibrio de carga, limitación del exceso de conexiones y replicación; esta última característica se consideró en esta investigación [5].

Con el avance de la tecnología, tanto en el ámbito del desarrollo web y de las bases de datos, se han incorporado nuevas tecnologías, que permiten el desarrollo de soluciones informáticas mucho más robustas, escalables, productivas, seguras, flexibles, con un mayor rendimiento y de alta disponibilidad. Java 2 Enterprise Edition (J2EE) es una plataforma y filosofía de diseño para grandes sistemas empresariales [6].

JavaServer Faces (JSF) es un marco de componentes de interfaz de usuario (UI) para aplicaciones web J2EE que, una vez adoptadas, permite a las organizaciones migrar desde tecnologías antiguas, como plataformas basadas en caracteres para terminales virtuales, a plataformas y tecnologías más actualizadas basadas en estándares, como JSF y Java [7]. JSF es considerado como un framework, el cual tiene como propósito hacer más con menos código, para agilizar el desarrollo [8].

PrimeFaces es una suite de componentes open source de JSF, que provee un conjunto de componentes enriquecidos, con Ajax incorporado, ligero, con soporte a través de Atmosphere Framework, con un kit de UI Mobile para la creación de aplicaciones web móviles, y con mucha documentación de apoyo para el desarrollador [9].

Por otro lado, las tutorías académicas en las instituciones de educación superior significan un proceso educativo dirigido a los estudiantes, con el objetivo de orientar, atender y acompañar de forma individual o grupal, en el desarrollo académico conforme a dimensiones basadas en criterios y mecanismos de monitoreo y control, a fin de fortalecer su formación integral.

En el Art. 15 componente 1 literal. b) 2do. Párrafo del Reglamento de Régimen Académico del Consejo de Educación Superior (CES), se menciona: Son actividades de aprendizaje colaborativo, entre otras: la sistematización de prácticas de investigación-intervención, proyectos de integración de saberes, construcción de modelos y prototipos, proyectos de problematización y resolución de problemas o casos. Estas actividades deberán incluir procesos colectivos de organización del aprendizaje con el uso de diversas tecnologías de la información y la comunicación, así como metodologías en red, tutorías in situ o entornos virtuales [10].

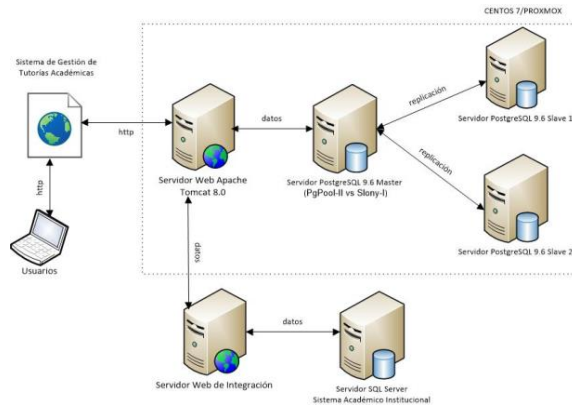


Fig. 1. Arquitectura de la Solución del Sistema integrado de alta disponibilidad para la gestión de tutorías académicas en instituciones de Educación Superior

En base a éste contexto, y, con el propósito de contribuir a la automatización de los procesos académicos y a los indicadores del modelo para la evaluación de las carreras, presentado por el Consejo de Aseguramiento de la Calidad de la Educación Superior (CACES), en la Universidad Nacional de Chimborazo (UNACH) se implementó el Sistema integrado de alta disponibilidad para la gestión de tutorías académicas orientado a las Instituciones de Educación superior, para lo cual en el desarrollo de la solución se aplicó la arquitectura N Capas, utilizando tecnologías con versiones estables de acuerdo a la tendencia de uso establecida en google trends, como es el caso de Netbeans 8.0, JSF2.2, PrimeFaces 6.2.27 y Apache Tomcat 8.0. Posteriormente se desarrollaron servicios web para su integración con un sistema académico heterogéneo implementado bajo software propietario, y por último se realizó un análisis comparativo de las herramientas de replicación de datos Slony-I y PgPool-II con respecto al rendimiento a través de pruebas de estrés utilizando Apache JMeter.

La solución permite llevar un control y seguimiento de las tutorías académicas in situ o virtuales, de tipo individual o grupal, dirigidas a los estudiantes con inconvenientes en tres tipos de dimensiones: académica, profesional y personal. Cuenta con los siguientes módulos: integración, administración, planificación de actividades, gestión de actividades y reportes.

II. METODOLOGÍA

A. Arquitectura de la solución

El estudio de rendimiento de las herramientas de replicación se la realizó en función al tiempo de respuesta frente a la ejecución de sentencias SQL, uso de CPU, uso de memoria, y uso de disco duro, con la finalidad de obtener una arquitectura de la solución como se describe en la Figura 1.

B. Materiales

La infraestructura sobre la cual se implementó la solución tecnológica es una red LAN de tres capas como se muestra en la Figura 2. En donde se puede observar como parte de la solución un servidor web que alojará el sistema de gestión de tutorías, un servidor web que aloja los servicios web para la

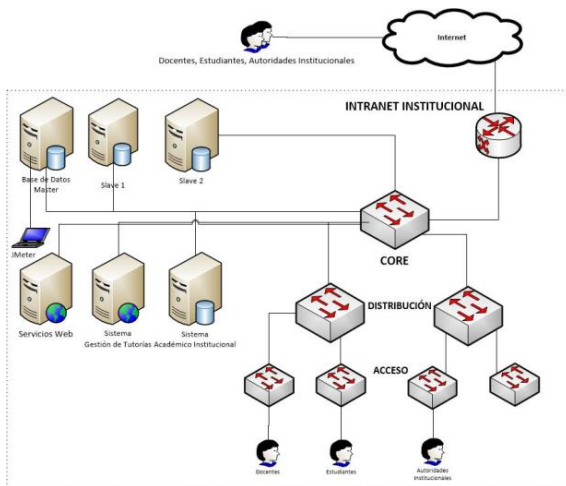


Fig. 2. Infraestructura base para el despliegue de la Solución Tecnológica.

Tabla I
INSUMOS PARA LA IMPLEMENTACIÓN DE LA APLICACIÓN WEB

Insumo	Descripción	Versión
Sistema Operativo	CentOS	7
Arquitectura de desarrollo de Software	N Capas	N/A
IDE	Netbeans	8
Framework	JSF	2.2
	PrimeFaces	5.2.1
Servidor Web	Apache Tomcat	8
Motor de bases de datos aplicación web	PostgreSQL	9.6
Servicios Web	Visual Studio .NET	2017
Motor de bases de datos del sistema académico institucional	SQL Server	2008

integración con el sistema académico institucional, un servidor de base de datos que resguarda la información del sistema académico institucional, y tres servidores de bases de datos para la replicación de datos. Los estudiantes, docentes y autoridades institucionales, considerados como usuarios finales pueden acceder a la aplicación desde cualquier dispositivo tanto desde la intranet como desde el internet. En la Tabla I y II, se detallan los insumos para la solución, y en la Tabla III las características de los servidores utilizados.

C. Procesamiento y análisis

El procesamiento y análisis se basa en la siguiente propuesta metodológica de comparación de herramientas de replicación de datos, la que está compuesta de los siguientes pasos:

- Análisis de herramientas de replicación de datos
- Desarrollo del Sistema integrado de alta disponibilidad para la gestión de tutorías académicas en instituciones de Educación Superior
- Implementación de ambientes de prueba
- Ejecución de pruebas de rendimiento
- Resultados

Paso 1: Análisis de herramientas de replicación de datos

Tabla II
INSUMOS PARA LA IMPLEMENTACIÓN DE REPLICACIÓN DE DATOS Y PRUEBAS DE RENDIMIENTO

Insumo	Descripción	Versión
Sistema Operativo	CentOS	7
Motor de bases de datos	PostgreSQL (base de datos del sistema de gestión de tutorías)	9.6
Herramienta replicación 1	de Slony -I	2.2.3
Herramienta replicación 2	de PgPool-II	3.7.0
Herramienta de estrés	Apache JMeter	5.2.1
Complemento	Apache JMeter PerfMon (Servers Performance Monitoring)	2.1
Herramienta de estrés	Apache JMeter PerfMon Server Agent	2.23

Tabla III
CARACTERÍSTICAS TÉCNICAS DE LOS SERVIDORES DE LA SOLUCIÓN BAJO LA PLATAFORMA DE VIRTUALIZACIÓN PROXMOX VE 4.2-2

Servidor	RAM (GB)	CPU	Disco Duro(GB)
Web Tomcat	16	12 x Intel(R) Xeon(E) CPU E5-2620 2GHz (2 sockets, 4 cores)	150
Base de Datos PostgreSQL (Master)	16	12 x Intel(R) Xeon(E) CPU E5-2620 2GHz (2 sockets, 4 cores)	500
Base de Datos PostgreSQL (Slave 1)	16	12 x Intel(R) Xeon(E) CPU E5-2620 2GHz (2 sockets, 4 cores)	500
Base de Datos PostgreSQL (Slave 2)	16	12 x Intel(R) Xeon(E) CPU E5-2620 2GHz (2 sockets, 4 cores)	500

Tabla IV
HISTORIAS DE USUARIO

Nº	Nombre	Iteración
1	Gestión de Acceso de Usuarios	1
2	Gestión de Seguridad	1
3	Gestión de Facultades	1
4	Gestión de Carreras	1
5	Gestión de Niveles	1
6	Gestión de Periodos	1
7	Gestión de Tutores	1
8	Gestión de Actividades Planificadas	2
	Gestión de Actividades Realizadas	2
	Emisión de Reportes	3

Tabla V
PLAN DE ENTREGA ITERACIONES

Iteración	Historia de Usuario	Duración en semanas
1	Gestión de Acceso de Usuarios	2
1	Gestión de la Seguridad	3
1	Gestión de Facultades	1
1	Gestión de Carreras	1
1	Gestión de Niveles	1
1	Gestión de Periodos	1
1	Gestión de Tutores	3
2	Gestión de Actividades Planificadas	2
2	Gestión de Actividades Realizadas	3
3	Emisión de Reportes	3

El análisis corresponde a la revisión de elementos técnicos y de versiones de las herramientas de replicación de datos Slony-I y PgPool-II, para seleccionar la mejor alternativa en la implementación de la solución.

Paso 2: Desarrollo del Sistema integrado de alta disponibilidad para la gestión de tutorías académicas en instituciones de Educación Superior

Para la implementación del software, se aplicó la metodología Extreme Programming (XP), por considerarse una metodología ágil, centrada en potenciar las relaciones interpersonales como clave para el éxito en desarrollo de software, promoviendo el trabajo en equipo, preocupándose por el aprendizaje de los desarrolladores, y propiciando un buen clima de trabajo [11].

Se desarrollaron los módulos de: integración, administración, planificación de actividades, gestión de actividades y reportes. Para lo cual se utilizaron los insumos detallados en la Tabla I.

Como roles del equipo de trabajo se definieron un rastreador, testeador, un programador y un consultor, en la Tabla IV se detallan las historias de usuario con un riesgo, prioridad y esfuerzo alto y en la Tabla V se muestra el plan de entrega de las historias de usuario de acuerdo a iteraciones.

La base de datos de la solución se implementó utilizando el Front-End de PostgreSQL, Pg-Admin IV. La arquitectura utilizada para la solución es la N Capas, conformada por las capas de: Base de Datos, Acceso a Datos, Reglas de Negocio, Integración y Presentación. En base a los requerimientos recolectados, la descripción detallada en las historias de usuario y el diagrama de procesos se definió las interfaces de usuario definitivas, las cuales fueron implantadas en el sistema, como se muestra en la Figura 4.

Paso 3: Implementación del ambiente de pruebas

Una vez implementado el sistema, se definió el ambiente de pruebas que permitió la comparación de ambas herramientas de replicación de datos, como se muestra en la Figura 3., utilizando los insumos detallados en la Tabla I y II.

Paso 4: Ejecución de pruebas de rendimiento

Para la toma y recolección de datos, se generaron las pruebas de rendimiento para las dos herramientas, ejecutando Apache JMeter, realizando por varias ocasiones una simulación de 282 peticiones concurrentes por segundo durante un lapso de 30 minutos. Este número corresponde al número aproximado de usuarios que acceden normalmente al aplicativo.

Paso 5. Resultados

Dado que los datos de los parámetros de la investigación no cumplen los supuestos de normalidad y homogeneidad de

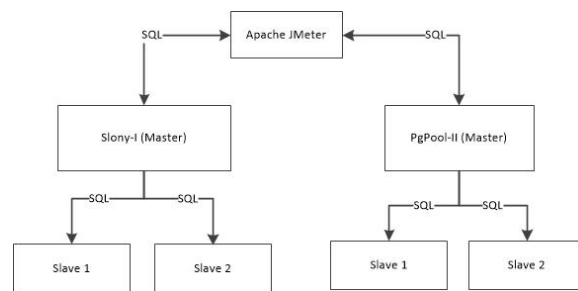


Fig. 3. Arquitectura de pruebas de rendimiento con Apache JMeter



Fig. 4. Interfaz de usuario del Sistema integrado de alta disponibilidad para la gestión de tutorías académicas, disponible en <http://siget.unach.edu.ec>

varianzas, se aplicó la prueba U de Mann-Whitney, se compararon las medias, posterior a ello se interpretaron los resultados y se obtuvieron las conclusiones respectivas.

III. RESULTADOS

A. Resultados de las mediciones de rendimiento de las herramientas SLONY-I y PGPOOL-II

Como se mencionó anteriormente, para la toma y recolección de datos, se generaron las pruebas de rendimiento, ejecutando Apache JMeter, tanto para el tiempo de respuesta frente a la ejecución de sentencias SQL, uso de CPU, uso de memoria, y uso de disco duro de manera independiente. Se Aplicó la prueba estadística, estableciendo un nivel de significancia de 0,05% con una muestra de 384 por grupo, obteniéndose los resultados presentados en la Tabla VI, Tabla VII y Tabla VIII.

Tabla VI
ANÁLISIS COMPARATIVO SLONY-I FRENTE PGPOOL-II

Herramientas	Método de replicación	Sincronización	Pool de Conexión	Balaceo de la Carga	Posicionamiento de Consultas
SLONY-I	Maestro/Esclavo	Asíncrono	NO	NO	NO
PGPOOL-II	Maestro/Esclavo	Síncrono	SI	SI	SI

Tabla VII
TIEMPO DE RESPUESTA Y PORCENTAJE DE USO DE RECURSOS DE SLONY-I Y PGPOOL-II

Indicadores	Slony-I	PgPool-II	Diferencia
Tiempo de respuesta	2,22s	0,87s	1,35s
Uso de CPU	10,37 %	3,27%	7,1%
Uso de RAM	14,49%	5,05%	9,44%
Uso de DISCO	0,25%	0,09%	0,16%
Promedio uso de recursos	8,37%	2,80%	5,57%

Tabla VIII
PORCENTAJE DE ERROR EN EL PROCESO DE REPLICACIÓN

Indicadores	Slony-I	PgPool-II	Diferencia
Error de procesamiento	0,009%	1,56%	1,551%

IV. CONCLUSIONES

Aplicando la prueba estadística U de Mann-Whitney se comprobó que existe diferencias de rendimientos en las herramientas de replicación. Además, en la Tabla VII se demuestra que la herramienta de replicación PgPool-II presentó ventajas sobre Slony-I con respecto al rendimiento, y ejecución de sentencias SQL en ambientes Linux CentOS 7 de 64 bits, luego de ejecutar Apache JMeter sobre la infraestructura base para la solución. Sin embargo, es importante recalcar que al realizar las pruebas de rendimiento por medio de la ejecución de transacciones de replicación se comprobó que Slony-I es más eficaz que PgPool-II.

Los resultados obtenidos tanto en el ambiente de pruebas y de producción arrojaron los mismos resultados. El método de replicación asincrónica usado por la herramienta Slony-I, es un método de replicación que provee menor carga a la red, por tal razón los datos se pueden replicar en un momento determinado; es una herramienta útil a la hora de realizar replications de datos de un servidor maestro a varios esclavos, puede replicar parte de las tablas a otros servidores esclavos diferentes, su configuración es menos compleja que PgPool-II.

El método de replicación síncrona usada por la herramienta PgPool-II, es propenso a sobrecargar la red por razón que el servidor esclavo debe estar en escucha todo el tiempo, a posibles cambios que se realice en servidor maestro, permite replicar toda una base de datos, así como esquemas y tablas desde el servidor PgPool-II a clientes, presenta su nivel de complejidad al realizar las configuraciones ya que no existe mucha documentación disponible. PgPool-II está

expuesto a tener errores con PostgreSQL versión 9.6 o inferior por temas de compatibilidad y comunicación.

Se recomienda trabajar con cualquiera de los dos métodos de replicación síncrona o asincrónica, dependiendo del contexto y de la infraestructura base que soportará el sistema integrado, si se trabaja con el método de replicación síncrona se debe tomar en cuenta que esta propenso a sobre cargar la red, mientras con el método de replicación asincrónica se debe tomar en cuenta los datos se puede actualizar más tarde desde el servidor maestro a servidores esclavos.

PrimeFaces cuenta con 117 componentes, no requiere dependencias para utilizar la librería, es compatible con otras librerías de componentes, posee soporte Ajax, su documentación es actualizada, además los componentes son amigables, atractivos e innovadores tanto para el desarrollador como para el usuario.

AGRADECIMIENTOS

A las autoridades de la Universidad Nacional de Chimborazo, a la Facultad de Ingeniería, al Centro de Tecnologías Educativas, a la Unidad Técnica de Control Académico, y a la Dirección General Académica, por brindar las facilidades técnicas y administrativas para la implementación del presente proyecto de investigación.

REFERENCIAS

- [1] M. Marqués. SQL Server 2008 R2 Motor de base de datos y administración. Grupo RC, España, 2011.
- [2] F. Asteasuain y A. Javed, «Tecnología aplicada a aspectos en base de datos relacionales: El caso de replicación.» *Ingeniería & Desarrollo*, n° 25, p. 232, 2009.
- [3] Y. Guerra, E. Altuna, A. García y. Pérez, «Experiencias en el uso del symmetric para la replicación de datos en la plataforma educativa ZERA.» *Revista Cubana de Ciencias Informáticas*, vol. 9, n° 4, p. 197, 2015.
- [4] S.D. Group. Slony-I, 2007. [En línea]. Disponible: <http://www.slony.info/>.
- [5] MediaWiki. PgPool, 2011. [En línea]. Disponible: <http://www.pgpool.net>
- [6] G. Brill. Codenotes for J2EE: EJB, JDBC, JSP, and Servlets. Random House Trade Paperbacks, New York, 2002.
- [7] J. Jacobi, J. Fallows. Pro JSF and AJAX. Building Rich Internet Components (1 ed.). Apress, USA, 2006.
- [8] J. Arrambide. Desarrollo Web en Java: Configura paso a paso todas las herramientas para crear proyectos web. Edición Kindle, Estados Unidos, 2017.
- [9] PrimeTek. Leading Provider of Open Source UI Component Libraries, 2009. [En línea]. Disponible: <https://www.primefaces.org/>
- [10] Consejo de Educación Superior. Reglamento de Régimen Académico. CES, Ecuador, 2019.
- [11] K. Beck y C. Andres, Extreme Programming Explained: Embrace Change. Addison-Wesley Educational Publishers Inc, 2004.