

Instituto Tecnológico de Costa Rica  
Centro Académico de Alajuela  
Bases de Datos II

Grupo 8:

Marco Rodríguez Vargas – Carné: 2022149445

Kevin Carranza Jiménez – Carné: 2015100260

Jorge Esteban Benavides Castro – Carné: 2022230697

Profesor:

Alberto Shum Chan

Fecha de entrega:

20/9/2023 a las 16:59 pm

II Semestre, 2023

**Tema:** Práctica Replicación en PostgreSQL

**Objetivo:** Evaluar al menos tres opciones de replicación disponibles en PostgreSQL o cualquier motor de base de datos y seleccionar una para utilizar en la implementación del segundo proyecto programado.

**Entrega en el TecDigital:**

- Deben presentar un archivo pdf.
- Forma de trabajo en grupos de 3 personas.

**Actividades:**

1. Lea el capítulo 26 del manual de PostgreSQL High Availability, Load Balancing, and Replication disponible en [1] o investigue opciones de replicación desarrolladas por terceros.
2. Seleccione tres de las opciones que considere que podrían servir para implementar el segundo proyecto del curso que tiene como objetivos:

**Objetivo proyecto:** Desarrollar un proyecto básico de inteligencia de negocio que incluya desde la configuración de algunos elementos de la infraestructura tecnológica requerida hasta los elementos de visualización de datos al servicio de los usuarios finales. Para lograr el objetivo se debe implementar un ambiente de replicación y análisis de datos.

**Objetivos específicos:**

- Diseñar e implementar un *datamart* basado en un modelo multidimensional.
  - El *datamart* se debe alimentar de una base de datos transaccional por medio de un proceso de replicación.
  - Diseñar e implementar un *dashboard* que resuma información importante sobre los datos utilizando el software *Tableau*.
3. Evalúe las tres herramientas teniendo en cuenta los requerimientos del proyecto.

Para la evaluación llene la siguiente tabla:

Nombre de la herramienta	Sincrónica	Continua	Cascada
Desarrollada por:	La replicación sincrónica no fue desarrollada por un grupo en concreto sino que evolucionó a lo largo de los años y con aportes de varias entidades, como por ejemplo: Oracle, IBM, Microsoft, PostgreSQL y MySQL.	La replicación continua no es una herramienta desarrollada por una sola entidad, sino que existen diferentes herramientas de replicación continua desarrolladas por diferentes entidades. Algunas de ellas son MySQL Replication, desarrollada por MySQL que ahora pertenece a Oracle Corporation o PostgreSQL Streaming Replication, desarrollada en Postgre, una base de datos de código abierto.	PostgreSQL (agregada en la versión 9.5), por miembros de la comunidad (Andrew Gierth, Peter Eisentraut, Tom Lane, Magnus Hagander, David Fetter, Josh Berkus, Jonathan Perkin, Stephen Frost, Bruce Momjian).
Documentación disponible en:	<a href="#">PostgreSQL</a> <a href="#">Oracle</a> <a href="#">MySQL</a>	<a href="#">PostgreSQL: Documentation: 16: 20.6. Replication</a>	<a href="https://www.postgresql.org/docs/current/warm-standby.html#SYNCHRONOUS-REPLICATION">https://www.postgresql.org/docs/current/warm-standby.html#SYNCHRONOUS-REPLICATION</a> (sección 27.2.7)
¿Es software libre? / Licencia de uso	Existen diferentes versiones de la herramienta, donde algunas son de software libre y otras licenciadas.  Código abierto: PostgreSQL. MySQL.	Dependiendo de la herramienta utilizada puede ser de código abierto. En el caso de PostgreSQL es de software libre.	La replicación en cascada de PostgreSQL es un método de replicación de datos desarrollado por la comunidad de PostgreSQL y es software libre.

	Licenciada: Oracle. ( <a href="#">Licencia</a> )		El código fuente está disponible gratuitamente bajo la Licencia PostgreSQL.
Tipo de replicación que implementa	La replicación síncrona implementa el que la información se escribe simultáneamente en varios servidores y se confirma el cambio si se escribió en más de un servidor	La replicación continua se implementa mediante la configuración de un sistema maestro y uno o más sistemas secundarios en una red de bases de datos.	La replicación en cascada de PostgreSQL es un método de replicación de datos de tipo maestro-esclavo que propaga los cambios de la base de datos maestra a las bases de datos secundarias de forma asíncrona.
Forma de transmisión de los datos	<p>La conexión de los servidores se da a través de archivos WAL donde después de que un commit se escribe en el servidor principal los servidores de apoyo responde cada vez que un archivo WAL es escrito.</p> <p>Estos mensajes van a ser comparados con los enviados por otros servidores de apoyo para decidir cuándo liberar transacciones que están pendientes de confirmar que el commit se hizo.</p> <p>En resumen, transmite los datos mediante archivos</p>	<p>El maestro registra todas las operaciones de escritura en un archivo de registro binario llamado Write-Ahead Logging (WAL). Estos registros se envían de manera continua a los sistemas secundarios a través de una conexión de flujo, donde se aplican en el mismo orden en que se reciben.</p>	<p>La replicación en cascada de PostgreSQL utiliza una forma de transmisión de datos basada en archivos de registro de transacciones (WAL). Los archivos WAL registran todos los cambios realizados en la base de datos maestra. Los suscriptores utilizan estos archivos para replicar los cambios en sus propias bases de datos.</p> <p>En resumen, la replicación en cascada de PostgreSQL</p>

	WAL donde se espera a la respuesta de todos los servidores para ejecutar el cambio.		transmite los datos mediante archivos WAL.
¿Permite que el esclavo realice consultas a la base de datos?:	Sí, ya que una vez configurada la herramienta el servidor esclavo puede hacer consultas de lectura al servidor principal, como por ejemplo consultas: SELECT, SELECT WHERE, JOIN entre otras que no modifiquen el servidor principal.	Sí, los servidores secundarios en un sistema de replicación continua pueden realizar consultas a la base de datos. Estos servidores secundarios mantienen una copia actualizada de los datos del servidor principal, lo que les permite acceder y consultar la base de datos de manera independiente.	La replicación en cascada de PostgreSQL permite que el esclavo realice consultas a la base de datos, ya que los cambios realizados en la base de datos maestra se propagan a los suscriptores, incluyendo las consultas.
Complejidad de instalación	<p>El proceso completo de instalación de este sistema se puede concentrar en:</p> <ul style="list-style-type: none"> <li>-Configurar el repmanager* en tres equipos.</li> <li>-Configurar el servidor primario.</li> <li>-Copia de los datos a servidores secundarios.</li> <li>-Añadir el repmanager los tres servidores.</li> <li>-Configurar la réplica asíncrona.</li> </ul> <p>Si bien este proceso puede ser bastante técnico su instalación se puede llevar a cabo con relativa facilidad con</p>	<p>La instalación de la replicación continua implica configurar un servidor como el "maestro" y uno o más servidores como "réplicas". El maestro registra cambios en un archivo de registro (WAL) y los envía constantemente a las réplicas. Esto asegura que las réplicas mantengan copias actualizadas. Aunque las réplicas pueden realizar consultas a la base de datos, suelen estar configuradas en modo de solo lectura por</p>	<p>La configuración del publicador es relativamente sencilla. Se trata de habilitar la replicación en cascada y especificar la ruta a los archivos WAL.</p> <p>La configuración de los suscriptores es más compleja. Se trata de establecer la conexión con el publicador, configurar la sincronización de los datos y habilitar las consultas.</p>

	ayuda de manuales y tutoriales.  *repmanager: administrador de réplica.	defecto. Para habilitar operaciones de escritura en las réplicas, puede ser necesario realizar configuraciones adicionales y manejar conflictos de escritura.	En resumen, la instalación de la replicación en cascada de PostgreSQL es moderadamente compleja.
Complejidad de administración	Medio alta, si no se hace bien se vuelve muy lento todo y hasta puede tirar y aumentar la probabilidad de errores	La administración de un sistema de replicación continua puede implicar cierta complejidad. Es necesario supervisar regularmente el estado de los servidores maestro y secundarios para asegurarse de que estén funcionando correctamente y que la replicación esté en curso.	En general, la replicación en cascada de PostgreSQL es relativamente fácil de administrar. Sin embargo, es importante comprender los conceptos básicos de la replicación para poder realizar las tareas de administración de forma eficaz.
Ventajas	Se asegura que no haya pérdida de información, pues toda la información está en todos los servidores.	<ul style="list-style-type: none"> <li>• Alta disponibilidad.</li> <li>• Rendimiento mejorado.</li> <li>• Tolerancia a fallos.</li> <li>• Escalabilidad.</li> <li>• Copias de seguridad sin interrupciones.</li> <li>• Georredundancia.</li> <li>• Balanceo de carga.</li> </ul>	La capacidad de mejorar el rendimiento y la disponibilidad de la base de datos mediante la escalabilidad, garantizar una alta disponibilidad al preservar la funcionalidad en caso de fallos del publicador, y la capacidad de respaldar los datos replicados para protegerlos ante posibles obstáculos, brindando así una solución

			segura para la gestión de bases de datos.
Limitaciones	Si no se implementa cuidadosamente puede reducir rendimiento al aumentar el tiempo de respuesta y la contención de datos.	<ul style="list-style-type: none"> <li>• Consistencia eventual.</li> <li>• Posibles conflictos.</li> <li>• Requisitos de ancho de banda.</li> <li>• Latencia de replicación.</li> <li>• Complejidad de configuración y mantenimiento.</li> <li>• Costos de almacenamiento adicional.</li> <li>• Escasa tolerancia a fallos de red.</li> </ul>	<p>En primer lugar, es crucial que la base de datos maestra esté disponible en todo momento para que la replicación funcione adecuadamente. Además, los cambios efectuados en la base de datos maestra se propagan de manera asincrónica a las bases de datos secundarias, lo que puede ocasionar que los datos no estén completamente sincronizados en todo momento. Además, los suscriptores no tienen la capacidad de realizar modificaciones en los datos replicados que no sean coherentes con la estructura de la base de datos maestra.</p>
¿Existe un grupo de desarrolladores dando soporte a la herramienta?	Esta tecnología es común de encontrar en DBMSs, por lo que cuenta con el	Para PostgreSQL, el soporte puede ser proporcionado por varias fuentes, incluyendo:	La herramienta cuenta con respaldo de un equipo de desarrolladores,

	<p>soporte de grupos como: PostgreSQL, MySQL Cluster, Oracle Database.</p> <p>Además de foros de software libre y similares.</p>	<p>La comunidad de PostgreSQL a través de foros y listas de correo. Empresas que ofrecen servicios de soporte y consultoría para PostgreSQL, como EnterpriseDB, 2ndQuadrant, entre otras.</p>	<p>encabezado por el PostgreSQL Global Development Group (PGDG), el cual tiene la responsabilidad de supervisar tanto el desarrollo como el mantenimiento continuo de PostgreSQL.</p>
Existe documentación	<p>Sí, cada grupo tiene su propia documentación sobre su herramienta.</p>	<p>Documentación oficial de PostgreSQL:  <a href="https://www.postgresql.org/docs/">https://www.postgresql.org/docs/</a></p> <p>La sección específica sobre replicación:  <a href="https://www.postgresql.org/docs/current/high-availability.html">https://www.postgresql.org/docs/current/high-availability.html</a></p>	<p>Sí, existe documentación disponible. El enlace está arriba.</p>
Costo de la herramienta	<p>Al contar con versiones de código abierto se puede considerar como una herramienta gratuita.</p>	<p>PostgreSQL es un sistema de gestión de bases de datos de código abierto y la replicación continua (como la Streaming Replication) está incluida de forma gratuita en la distribución estándar de PostgreSQL. No hay costos de licencia asociados.</p>	<p>La replicación en cascada de PostgreSQL es una herramienta gratuita. PostgreSQL es un software de código abierto, lo que significa que es libre de usar, modificar y redistribuir.</p>

4. Seleccione la herramienta que, de acuerdo a lo investigado, solventa las necesidades del proyecto. Explique por qué recomienda la herramienta.



La herramienta que de acuerdo a lo investigado solventa las necesidades del proyecto es la replicación sincrónica, la cual garantiza una sincronización en tiempo real de los datos entre el servidor principal (maestro) y los servidores secundarios (esclavos). Esto es esencial para un proyecto de inteligencia de negocio, ya que los usuarios finales requieren acceso a información actualizada al instante. Los datos que se muestran en los informes y visualizaciones de Tableau estarán siempre al día, lo que mejora significativamente la toma de decisiones basada en datos.

Además, la replicación sincrónica proporciona alta disponibilidad al garantizar que los servidores secundarios estén siempre actualizados y listos para asumir la carga en caso de que el servidor principal tenga una interrupción. Esto minimiza el tiempo de inactividad y asegura que los usuarios puedan acceder a los datos incluso en situaciones de falla del servidor principal.

Si bien la replicación sincrónica puede introducir una pequeña latencia en el maestro debido a la espera de la confirmación de los servidores secundarios, esta latencia es generalmente insignificante en comparación con los beneficios que brinda en términos de actualización en tiempo real y alta disponibilidad. En resumen, la replicación sincrónica es una buena elección para el proyecto porque cumple con los requisitos de sincronización en tiempo real y alta disponibilidad, asegurando que los datos estén siempre actualizados y disponibles para los usuarios.

## Referencias

[1] PostgreSQL (nd.). High Availability, Load Balancing, and Replication. Recuperado de <https://www.postgresql.org/docs/13/high-availability.html>

**Rúbrica**

Rubro	Puntos
Seleccionó tres opciones de replicación	1 punto
La tabla de evaluación está completa y se responden las consultas de todas las líneas de la tabla para las tres herramientas (5 pts.), Se responden las líneas de evaluación para dos herramientas (4 pts.), Se responden para una herramienta completa y las otras dos incompletas (3 pts.), Se responden solo para una herramienta, pero se hace una buena evaluación (2 pts.), Se responde para una herramienta no completa (1 pt.).	5 puntos
La recomendación del punto 4 está bien fundamentada	2 puntos
Total	8 puntos