

Titulación: Grado en Ingeniería Informática e Ingeniería en Sistemas de Información
Curso: 2022-2023. Convocatoria Ordinaria de Junio
Asignatura: Bases de Datos Avanzadas – Laboratorio

Practica 4: Replicación e Implementación de una Base de Datos Distribuida.

ALUMNO 1:

Nombre y Apellidos: _____

DNI: _____

ALUMNO 2:

Nombre y Apellidos: _____

DNI: _____

Fecha: _____

Profesor Responsable: _____

Mediante la entrega de este fichero los alumnos aseguran que cumplen con la normativa de autoría de trabajos de la Universidad de Alcalá, y declaran éste como un trabajo original y propio.

En caso de ser detectada copia, se puntuará **TODA** la asignatura como Suspenso – Cero.

Plazos

Tarea en laboratorio: Semana 1 de mayo y semana 8 de mayo.

Entrega de práctica: **Día 18 de mayo.** Aula Virtual

Documento a entregar: Este mismo fichero con la implementación de la replicación y la base de datos distribuida, las pruebas realizadas de su funcionamiento; y los ficheros de configuración del maestro y del esclavo utilizados en replicación; y de la configuración de los servidores de la base de datos distribuida (obligatorio). También los ficheros de log generados en la realización de la práctica de los postgres involucrados (obligatorio). Fichero para subir **DNI'sdelosAlumnos_PECL4.zip**

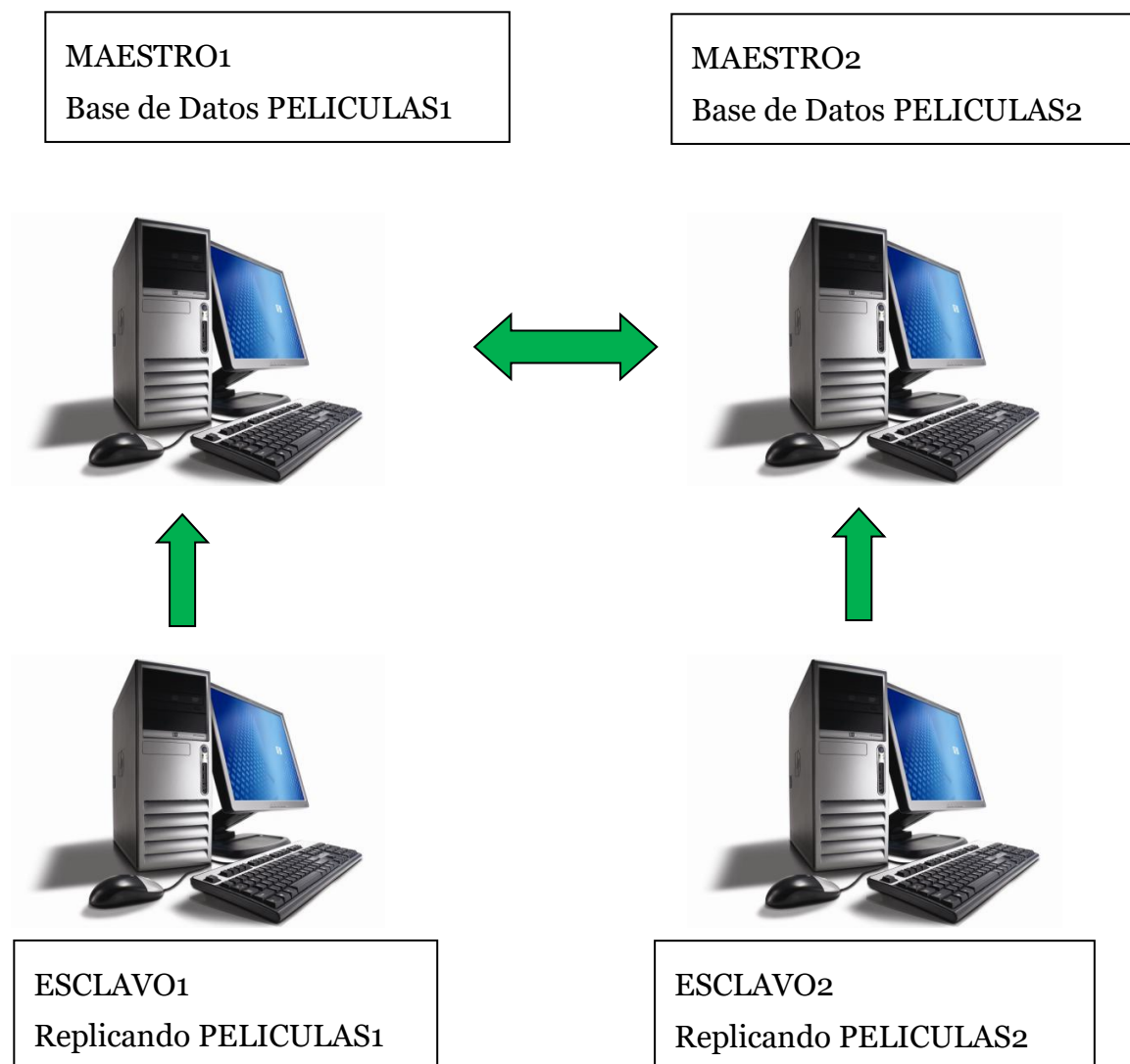
AMBOS ALUMNOS DEBEN ENTREGAR EL FICHERO EN LA PLATAFORMA.

Introducción

El contenido de esta práctica versa sobre la Replicación de Bases de Datos con PostgreSQL e introducción a las bases de datos distribuidas. Concretamente se va a utilizar los servicios de replicación de bases de datos que tiene PostgreSQL. Para ello se utilizará PostgreSQL 15.x con soporte para replicación. **Se prohíbe el uso de cualquier otro programa externo a PostgreSQL para realizar la replicación, como puede ser Slony.**

También se va a diseñar e implementar una pequeña base de datos distribuida. Una base de datos distribuida es una base de datos lógica compuesta por varios nodos (equipos) situados en un lugar determinado, cuyos datos almacenados son diferentes; pero que todos ellos forman una base de datos lógica. Generalmente, los datos se reparten entre los nodos dependiendo de donde se utilizan más frecuentemente.

El escenario que se pretende realizar se muestra en el siguiente esquema:



Se van a necesitar 4 postgres: 2 maestros y 2 esclavos. Cada maestro puede ser un ordenador de cada miembro del grupo con una base de datos de PELICULAS en

concreto (PELICULAS₁ y PELICULAS₂). Dentro de cada maestro se puede instalar una máquina virtual, que se corresponderá con el esclavo que se encarga de replicar la base de datos que tiene cada maestro, es decir, hace una copia o backup continuo de la base de datos PELICULAS₁ o de la base de datos PELICULAS₂. También es posible no usar otra máquina virtual para el esclavo y levantar otra estancia de postgres para guardar la réplica en un clúster diferente. Hay más de una solución.

Se debe de entregar el cuestionario que posea como los siguientes puntos para poder ser reproducido el trabajo por cualquier persona que pueda leerlo.

Cuestión o. Configurar el fichero de Error Reporting and Logging de PostgreSQL para que aparezcan recogidas las sentencias SQL DDL (Lenguaje de Definición de Datos) + DML (Lenguaje de Manipulación de Datos) generadas en dicho fichero. No se pide activar todas las sentencias. No activar la duración de la consulta. También se debe de configurar el log para que en el comienzo de la línea de registro de la información del log ("line prefix") aparezcan los DNI's de los miembros del equipo, el nombre del host con su puerto y la **fecha y hora** de la operación.

Cuestión 1. Configuración de cada uno de los nodos maestros de la base de datos de PELICULAS1 y PELICULAS2 por medio del **módulo postgres_fdw** que se puedan recibir y realizar consultas sobre las tablas de la base de datos que no tienen implementadas localmente.

Cuestión 2. Insertar datos diferentes en cada una de las bases de datos del MAESTRO1 y del MAESTRO2. Realizar una consulta sobre el MAESTRO1 que permita obtener el número total de géneros, críticas y visualizaciones de cada película usando toda la base de datos distribuida (MAESTRO1 + MAESTRO2). Explicar cómo se resuelve la consulta y su plan de ejecución.

Consulta

Plan de Ejecución

Explicaciones y Comentarios

Cuestión 3. Configuración completa de los equipos para estar en modo de replicación. Configuración del nodo maestro y del nodo esclavo. Tipos de nodos maestros, diferencias en el modo de funcionamiento y tipo elegido. Tipos de nodos esclavos, diferencias en el modo de funcionamiento y tipo elegido, etc.

Cuestión 4. Comentar los tipos de operaciones que se pueden realizar en cada tipo de nodo. Además, provocar situaciones de caída de los nodos y observar los mensajes obtenidos por cada nodo. Realizar las acciones correctoras necesarias para volver el sistema a un estado normal de operaciones. Comentar lo observado y realizado en esta cuestión.

Cuestión 5. Si el nodo MAESTRO2 se quedase inservible, ¿Qué acciones habría que realizar para poder usar completamente la base de datos en su modo de funcionamiento normal? ¿Cuál sería la nueva configuración de los nodos que quedan?

Cuestión 6. Según el método propuesto por PostgreSQL, ¿podría haber inconsistencias en los datos entre la base de datos del nodo maestro y la base de datos del nodo esclavo? ¿Por qué?

Cuestión 7. Conclusiones.

Se deben mostrar evidencias de que ha funcionado el sistema, por eso se deben de entregar los logs y ficheros de configuración de cada equipo utilizados en la práctica.

Bibliografía

- Capítulo 19: Server Setup and Configuration.
- Capítulo 20: Server Configuration.
- Capítulo: 21.1. The pg_hba.conf File
- Capítulo 26: Backup and Restore.
- Capítulo 27: High Availability, Load Balancing, and Replication.
- Appendix F: Additional Supplied Modules. F.33. Postgres_fdw