

Escuela de Ingeniería en Computación
Bachillerato en Ingeniería en Computación
Sede Interuniversitaria de Alajuela

Prof. Alberto Shum Chan
Base de Datos II
Semestre II, 2024

Proyecto #2

Tema: Replicación y análisis de datos

Entrega en el TecDigital:

- Deben presentar un archivo .PDF
- Adjuntar scripts y código fuente incluidos en el .PDF
- Adjuntar los scripts en un archivo .SQL
- Adjuntar los archivos de Tableau
- Forma de trabajo en grupos de 3 personas.
- Se agendará cita para revisión de la funcionalidad y dominio de lo implementado

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, que permita al estudiante experimentar con ambientes de apoyo a la toma de decisiones.

Objetivos específicos:

- Implementar un ambiente de replicación que permita separar el modelo multidimensional de la instancia transaccional.
- 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*.

Instrucciones:

- El segundo proyecto consiste en montar y cargar una base de datos para el alquiler de películas.
- Establecer una réplica (i.e., instancia esclava) de la base de datos en que se mantenga una copia actualizada de la base de datos.
- En el servidor maestro se debe implementar un modelo multidimensional que será replicado a la instancia esclava para realizar operaciones OLAP y visualización de estadísticas utilizando *Tableau*.
- La definición original de la base de datos y los datos de prueba están en un archivo SQL producido por el utilitario `pg_dump` de PostgreSQL y pueden ser fácilmente restaurados con el utilitario `pg_restore`.
- A continuación, se describen con más detalle las operaciones que se deben realizar.

1- Base de datos para alquiler de películas

Se usará la base de datos descrita en las páginas [PostgreSQL Sample Database](#), y [Load PostgreSQL Sample Database](#). Dicha base de datos consiste de 15 tablas para el control de préstamos de películas:

actor	almacena datos sobre actores y actrices
film	almacena datos de interés sobre diferentes películas
film_actor	almacena la asociación entre películas y actores
category	codifica los diferentes tipos de género de películas
film_category	almacena la asociación entre películas y categorías
store	contiene la diferente información sobre las tiendas
inventory	almacena los datos del inventario físico de discos con películas
rental	datos sobre los diferentes alquileres que se realizan. El alquiler de una película se realiza a un cliente en una fecha particular. El campo <code>return_date</code> es solo informativo, no debe ser tomado en cuenta para el cálculo del monto a pagar por el cliente.
payment	almacena los pagos de los diferentes clientes cuando retornan las películas alquiladas. El campo <code>payment_date</code> permite calcular el monto a cobrar por el alquiler de una película (ie. Sacando la diferencia entre la fecha de alquiler y la fecha de retorno real de la película). Vamos a asumir que no hay multas por entrega tardía de las películas.
staff	almacena los datos de los empleados
customer	almacena los datos de los clientes
address	almacena los datos de las direcciones de los clientes, empleados o tiendas
city	codifica el nombre de las ciudades
country	codifica el nombre de los países

- Todos los procedimientos y funciones almacenadas deben contener documentación completa que incluya una descripción de parámetros, salidas y bloques relevantes.

1.2. Seguridad: (5 puntos)

- ❖ Cree los siguientes roles
 - **EMP**: solo tiene el derecho de ejecutar los siguientes procedimientos almacenados; no puede leer ni actualizar ningún objeto de la base de datos
 - registrar un alquiler
 - registrar una devolución
 - buscar una película
 - **ADMIN**: tiene el derecho de un empleado más el derecho de ejecutar los siguientes procedimientos almacenados; no puede leer ni actualizar ningún objeto de la base de datos
 - insertar un nuevo cliente
- ❖ Cree los siguientes usuarios
 - **video**: no login, dueño de todas las tablas y de todos los procedimientos creados
 - **empleado1**: un usuario con rol **EMP**
 - **administrador1**: un usuario con rol **ADMIN**
- ❖ Seguridad en procedimientos almacenados los procedimientos almacenados deben correr usando las credenciales de su dueño, **video**.

2. Réplicas (30 puntos)

Se deben establecer una réplica o instancia esclava de la base de datos. No es necesario que la réplica esté en una máquina distinta. Para efectos del proyecto, la réplica puede estar en otra instancia en la misma estación de trabajo. No deben usar contenedores (Docker)

Aunque la réplica contiene una copia completa y actualizada de la base de datos, solo algunas de las tablas serán requeridas para implementar el **modelo estrella** para visualización que se describe a continuación.

Deben documentar detalladamente todo el proceso de configuración de la réplica, así como las pruebas realizadas.

3. Modelo multidimensional (20 puntos)

Con el fin de analizar la información acumulada de los alquileres se debe implementar un modelo multidimensional e implementar un *dashboard* para consultar dicho modelo (5 puntos).

Las medidas de interés son:

- número de alquileres
- monto total cobrado por alquileres

Comentado [ASC1]: Se debe hacer en la réplica para efectos de no recargar la instancia transaccional

Comentado [ASC2]: Recordar que es modelo estrella

Las dimensiones de interés son:

- Película (Film): jerarquía de categoría, filme incluyendo actores.
- Lugar (Address): jerarquía de país y ciudad
- Fecha (Rental): jerarquía de año, mes y día
- Sucursal (Store)

A partir de esta información se debe diseñar y construir las diferentes tablas del modelo multidimensional (estrella) que será la base para las visualizaciones. **(10 puntos)**

La documentación del modelo multidimensional (dimensiones, tabla de hechos y medidas) debe realizarse como lo propone la referencia.

Se deben desarrollar procedimientos almacenados para alimentar con datos el modelo multidimensional a partir de la réplica de la base de datos de alquileres **(15 puntos)**. Todos los procedimientos y funciones almacenadas deben contener documentación completa (detallada anteriormente).

4. Visualización y acceso por medio de interfaz gráfica al modelo estrella (30 puntos)

Se debe diseñar e implementar un *dashboard* que resuma información importante sobre los datos utilizando el software *Tableau*. Las consultas que debe solventar el dashboard se listan a continuación:

- ❖ Para una sucursal (a seleccionar por el usuario), grafique el número de alquileres realizados y el monto cobrado por mes, sin importar el año **(5 puntos)**.
- ❖ Graficar para un año (a seleccionar por el usuario) los montos cobrados por alquileres por mes **(5 puntos)**.
- ❖ Para una categoría de película (a seleccionar por el usuario), graficar el número de alquileres y el monto cobrado por año **(5 puntos)**.
- ❖ Para los 10 actores con más alquileres, graficar los montos totales de alquileres por año (a seleccionar por el usuario). Incluya la opción de todos los años **(5 puntos)**.
- ❖ Despliegue un mapa de ciudades que presente por año el monto de alquiler total representado por el tamaño del punto sobre la ciudad. **(10 puntos)**

Para cada uno de los puntos anteriores se debe documentar el proceso de construcción de cada visualización (imágenes y explicación).

Para cada visualización se debe hacer una interpretación de los datos mostrados.

Referencias

[1] Reeves, L (2009). A Manager Guide's to Data Warehousing (Capítulo 7).

El capítulo está disponible en el TecDigital o el libro se puede bajar de

<https://www.pdfdrive.com/a-managers-guide-to-data-warehousing-e23005825.html>

Comentado [ASC3]: La estrella se construye y se le da mantenimiento del lado del servidor MAESTRO y mediante replicación se sincroniza con el servidor ESCLAVO (permitir lectura – para hacer el OLAP y visualización). Se hace una replicación completa de toda la BD del servidor MAESTRO.

