



Instituto Politécnico Nacional

Escuela Superior de Cómputo



Programa académico / Plan de estudios

Ingeniería en Sistemas Computacionales / 2020

Unidad de aprendizaje

Bases de Datos

Práctica 5: "Edición de valores de una base de datos relacional empleando DML"

Objetivo: Desarrollar habilidades en la manipulación de datos mediante sentencias DML en bases de datos relacionales, aplicando técnicas de inserción, actualización y consulta a diferentes escalas. Los estudiantes automatizarán el poblado de bases de datos mediante scripts y comprenderán el impacto del volumen de datos en el rendimiento. Opcionalmente, podrán explorar arquitecturas de Data Warehouse para puntos adicionales.

Ejercicio 1: Diseño del Modelo de Datos

Diseñe un modelo de datos relacional de un dominio real (hospital, e-commerce, sistema académico, red social o bancario).

Requisitos:

- Mínimo 6 tablas relacionadas
- Restricciones: PRIMARY KEY, FOREIGN KEY, UNIQUE, NOT NULL, CHECK, DEFAULT
- Tipos de datos apropiados

Documentación necesaria:

- Diagrama Entidad-Relación con cardinalidades
- Diccionario de datos con descripción de tablas y columnas
- Script DDL completo con todas las restricciones e índices

Entregable: Documento con ERD, diccionario de datos y script DDL comentado.

Ejercicio 2: Poblado de Base de Datos en Tres Niveles

Desarrolle scripts automatizados para poblar su base de datos en tres niveles de carga. Use Python, Java, PHP u otro lenguaje.

Nivel 1: Poblado Leve (Desarrollo)

- 50-100 registros en tablas principales
- 100-1000 registros en tablas secundarias
- Datos realistas y coherentes

- Respeto de integridad referencial
- Uso de transacciones

Nivel 2: Poblado Moderado (Pre-producción)

- 5,000-10,000 registros en tablas principales
- 50,000-500,000 registros en tablas secundarias
- Técnicas de batch insert
- Desactivación temporal de índices
- Monitoreo de progreso y tiempo

Nivel 3: Poblado Masivo (Producción)

- 100,000-1,000,000 registros en tablas principales
- 1,000,000-10,000,000 registros en tablas secundarias
- Uso de COPY/LOAD DATA INFILE
- Carga paralela
- Métricas detalladas de rendimiento

Tabla comparativa obligatoria:

Métrica	Nivel 1	Nivel 2	Nivel 3
Total de registros			
Tiempo de ejecución			
Registros/segundo			
Uso de memoria (MB)			
Tamaño BD (MB)			

Entregable: Scripts de poblado para los tres niveles, documentación de técnicas utilizadas, capturas de ejecución y tabla comparativa.

Ejercicio 3: Operaciones DML Avanzadas

Implemente operaciones DML complejas que demuestren dominio de la manipulación de datos:

3.1 Consultas SELECT (mínimo 10)

Incluya al menos uno de cada tipo:

- JOINS múltiples (3+ tablas)
- Subconsultas correlacionadas
- Funciones de agregación con GROUP BY y HAVING
- Window functions (RANK, ROW_NUMBER, PARTITION BY)
- Operaciones de conjuntos (UNION, INTERSECT, EXCEPT)
- Common Table Expressions (CTEs)

- Consultas con CASE
- Análisis temporal con fechas
- Expresiones regulares o búsqueda de texto

3.2 Operaciones INSERT

- INSERT con subconsultas
- INSERT múltiple
- INSERT con valores calculados
- INSERT con manejo de duplicados (UPSERT)

3.3 Operaciones UPDATE

- UPDATE con JOIN
- UPDATE condicional con CASE
- UPDATE masivo
- UPDATE con subconsultas

3.4 Operaciones DELETE

- DELETE con subconsultas
- DELETE con JOIN
- Soft delete (marcado lógico)
- Archivado antes de eliminación

3.5 Transacciones

Implemente escenarios con:

- BEGIN/COMMIT/ROLLBACK
- SAVEPOINTS
- Niveles de aislamiento
- Manejo de bloqueos (FOR UPDATE)
- Control de errores y rollback

Entregable: Archivo SQL con todas las operaciones DML, documentación de cada operación y resultados de ejecución.

Ejercicio 4: Implementación con Docker

Dockerice su solución completa para garantizar portabilidad y reproducibilidad.

Componentes requeridos:

1. **Contenedor de base de datos:**
 - PostgreSQL 15+ o MySQL 8+
 - Volúmenes para persistencia
 - Scripts DDL de inicialización automática
 - Configuración optimizada
2. **Contenedor de aplicación:**
 - Scripts de poblado (tres niveles)
 - Dependencias instaladas
 - Variables de entorno configurables
 - Script de entrada para seleccionar nivel
3. **Contenedor de administración (opcional):**
 - pgAdmin4 o Adminer
 - Interfaz web para consultas

Archivos necesarios:

- Dockerfile - Imagen de la aplicación
- docker-compose.yml - Orquestación de contenedores
- entrypoint.sh - Script de entrada configurable
- README.md - Instrucciones de uso

Estructura del proyecto:

```
practica5/
├── data/
│   ├── sql/
│   │   ├── ddl/
│   │   └── dml/
│   └── scripts/
│       ├── poblar_leve.py
│       ├── poblar_moderado.py
│       └── poblar_masivo.py
├── Dockerfile
├── docker-compose.yml
├── entrypoint.sh
├── requirements.txt
└── README.md
```

Comandos de ejecución esperados:

bash

Poblado leve

docker-compose up

Poblado moderado

NIVEL_POBLADO=moderado docker-compose up

Poblado masivo

NIVEL_POBLADO=masivo docker-compose up

Entregable: Dockerfile, docker-compose.yml, scripts de configuración, README con instrucciones, capturas de contenedores en ejecución y evidencia de poblado exitoso.

Ejercicio 5: Data Warehouse (Puntos Adicionales - Opcional)

ESTE EJERCICIO ES OPCIONAL

Implemente un Data Warehouse básico siguiendo el modelo dimensional para análisis OLAP.

5.1 Diferencias OLTP vs OLAP

CARACTERÍSTICA	OLTP	OLAP
PROPÓSITO	Operaciones diarias	Análisis y reportes
DISEÑO	Normalizado (3FN)	Desnormalizado (Estrella)
CONSULTAS	Simples, frecuentes	Complejas, agregadas
VOLUMEN	Transacciones actuales	Histórico masivo
ACTUALIZACIONES	Constantes	Periódicas (ETL)

5.2 Modelo Dimensional

Diseñe un esquema estrella con:

Tabla de Hechos:

- Métricas numéricas (cantidad, monto, descuento)
- Claves foráneas a dimensiones
- Granularidad definida

Tablas de Dimensión (mínimo 4):

- dim_tiempo (fecha, mes, trimestre, año)
- dim_producto (nombre, categoría, marca)
- dim_cliente (nombre, segmento, región)
- dim_vendedor (nombre, región, sucursal)

5.3 Proceso ETL

Implemente las tres fases:

1. **Extract:** Leer datos de BD OLTP, identificar registros nuevos
2. **Transform:** Limpiar, validar, calcular métricas, enriquecer datos
3. **Load:** Insertar en dimensiones, insertar en hechos, actualizar agregados

5.4 Consultas OLAP

Implemente mínimo 5 consultas analíticas:

- Análisis de ventas por dimensión temporal
- Drill-down y roll-up por jerarquías
- Análisis comparativo entre períodos
- Top N productos/clientes
- Análisis de tendencias

Entregable: Script DDL del data warehouse, proceso ETL implementado, consultas OLAP, documentación comparativa OLTP vs OLAP, diagramas del modelo dimensional.

Fecha de Entrega

La fecha límite para la entrega de esta práctica es el **martes 16 de diciembre de 2025**. No se aceptarán entregas fuera de tiempo y forma.