

---

# Proyecto de Base de Datos

Profesor: Teofilo Chambilla, Brenner Ojeda  
Curso: Base de Datos I

Ciclo 2024.1

---

## 1. Introducción

El proyecto del curso de Bases de Datos I consiste en construir una base de datos de tamaño “razonable” para un problema del mundo real a elección. Se debe diseñar el modelo conceptual, el modelo de relacional y el modelo físico usando PostgreSQL. En la base de datos construida se debe ingresar datos reales y/o sintéticos, escribir consultas interactivas y modificaciones, explorar algunos otros aspectos como restricciones, índices, optimización, y crear un programa para manipular dicha base de datos (*Puntos extras para aquellos que desarrollen un front-end simple para poner la base de datos en la Web*). El punto clave inicial es identificar el dominio que se quiere administrar con la base de datos. Se aconseja tomar un problema con el que sientan afinidad o les divierta. Recuerde que estarán el resto del semestre trabajando en él. Ejemplos: un hobby, un sitio Web favorito, materiales de otro curso, un proyecto de investigación, la administración de un pequeño negocio, etc. Ojalá sea un dominio donde los datos reales estén presente. Respecto de las dimensiones: el diagrama de E/R debiera tener al menos 5 entidades y un número similar de relaciones. Se debe considerar restricciones de multiplicidad y participación, dominio de atributo (enteros, decimales, strings, etc.). Si encuentran dificultad para encontrar su dominio de datos, miren cualquier sitio Web de shopping. Ese es un tamaño adecuado para su proyecto.

## 2. Entregables

**Hito 1:** Fecha de entrega: 17 de Mayo 23:59hrs (aula virtual), se penalizará las entregas después de la fecha indicada (-2pt por cada día de retraso).

- **Descripción de requerimientos:** Suponga que Ud. es el cliente que necesita la base de datos; entonces debe escribir un documento de especificación de requerimientos que Ud. le entregará a la empresa desarrolladora que construirá su base de datos. Debe ser breve y claro. Indique posibles dificultades. Indique de dónde se obtendrán los datos. Debe indicar que datos existen, que significan, que reglas semánticas valen se aplica entre ellos, posibles restricciones, cómo se cargan y modifican (estáticamente una vez, se modifica todos los días, etc.), cómo se consultan (que consultas quieren hacer), que volúmenes se esperan, que velocidades de respuesta se esperan para las operaciones más comunes.
- **Modelo E/R:** Contextualizar la base de datos en el modelo E/R (conciso, pero no tan conciso). No olvide indicar los atributos de llave para las entidades e incluir las flechas indicando la multiplicidad y la doble línea indicando la participación. Si hay entidades débiles o relaciones “Is-A”, asegúrese de anotarlas apropiadamente indicando la cobertura y participación.
- **Modelo Relacional:** Producir el conjunto de relaciones basado en su modelo E/R. Especifique llaves para todas las relaciones. Si el modelo no contempla todas las llaves, en algunos casos **las llaves que Ud. especifique deberán estar basadas en la semántica de los datos**. El modelo debe estar normalizado (veremos que significa esto) y las restricciones de integridad –si existen– debidamente especificadas.

**Hito 2:** Fecha de entrega: 28 de Junio 23:59hrs (aula virtual), se penalizará las entregas después de la fecha indicada (-2pt por cada día de retraso).

- **Implementación en PostgreSQL:** Creación de la estructura de la base de datos en PostgreSQL. Hay algunas decisiones de implementación que deben tomarse. Registro de los datos. Implementación de consultas en SQL. Generación de vistas de usuario.
- **Optimización y Experimentación:** Seleccionar un subconjunto de las operaciones que se considere más críticas en términos de eficiencia y trabajar con PostgreSQL para lograr un mejor tiempo de respuesta para dichas operaciones. Esto incluye crear los índices adecuados para los campos adecuados. Analizar y estudiar los tiempos de actualización y de consulta cuando uno usa un índice u otro. Elaborar cuadros estadísticos de las consultas elegidas con y sin índices.

### 3. Formato de Informe

El informe se entregará en formato Latex y seguirá la siguiente estructura:

#### 1. Requisitos

- 1.1. Introducción
- 1.2. Descripción general del problema/organización/empresa
- 1.3. Necesidad/usos de la base de datos
- 1.4. ¿Cómo resuelve el problema hoy?
  - 1.4.1. ¿Cómo se almacenan/procesa los datos hoy?
  - 1.4.2. Flujo de datos
- 1.5. Descripción detallada del sistema
  - 1.5.1. Objetos de información actuales
  - 1.5.2. Características y funcionalidades esperadas
  - 1.5.3. Tipos de usuarios existentes/necesarios
  - 1.5.4. Tipos de consulta, actualizaciones
  - 1.5.5. Tamaño estimado de la base de datos
- 1.6. Objetivos del proyecto
- 1.7. Referencias del proyecto
- 1.8. Eventualidades
  - 1.8.1. Problemas que pudieran encontrarse en el proyecto
  - 1.8.2. Limites y alcances del proyecto

#### 2. Modelo Entidad-Relación

- 2.1. Reglas semánticas
- 2.2. Modelo Entidad-Relación
- 2.3. Especificaciones y consideraciones sobre el modelo

#### 3. Modelo Relacional

- 3.1. Modelo Relacional
- 3.2. Especificaciones de transformación
  - 3.2.1. Entidades
  - 3.2.2. Entidades débiles

- 3.2.3. Entidades superclase/subclases
  - 3.2.4. Relaciones binarias
  - 3.2.5. Relaciones ternarias
- 3.3. Diccionario de datos
- 4. Implementación de la base de datos **(Corresponde al Hito 2)**
  - 4.1. Creación de Tablas en PostgreSQL
  - 4.2. Carga de datos
  - 4.3. Simulación de Datos Faltantes
- 5. Optimización y Experimentación
  - 5.1. Consultas SQL para el experimento
    - 5.1.1. Descripción del tipo de consultas seleccionadas
    - 5.1.2. Implementación de consultas en SQL
  - 5.2. Metodología del experimento
  - 5.3. Optimización de consultas
    - 5.3.1. Planes de índices para Consulta 1
    - 5.3.2. Planes de índices para Consulta 2
    - 5.3.3. Planes de índices para Consulta 3
    - 5.3.4. Planes de índices para Consulta 4
  - 5.4. Plataforma de Pruebas
  - 5.5. Medición de tiempos
    - 5.5.1. Sin índices
    - 5.5.2. Con índices
  - 5.6. Resultados
    - 5.6.1. Consulta 1
    - 5.6.2. Consulta 2
    - 5.6.3. Consulta 3
    - 5.6.4. Consulta 4
  - 5.7. Análisis y Discusión
- 6. Conclusiones
- 7. Anexos ( Debe anexar un vídeo de la experimentación con 1 millón de registros con y sin índices )

## 4. Consideraciones

Considerar que la “experimentación y optimización” debe realizarse mediante 2 o 3 consultas “genéricas” con un nivel aceptable de complejidad en las consultas propuestas. Dichas consultas se deben realizar en cuatro contextos de 1 000(mil) datos, 10 000(diez mil) datos, 100 000(cien mil) datos y 1 000 000 (un millón) de datos almacenados en la base de datos (deberían tener 4 dumps de su proyecto).

(Extra) ¿Cuál sería la complejidad operacional si escalamos los datos por encima del millón?, realice una comparativa respecto a la cantidad de datos del párrafo anterior. ¿Es suficiente la arquitectura Cliente-Servidor para procesar millones de datos?

## **5. Rúbrica del Proyecto**

Ver archivo anexo 1 (Rubrica Proyecto I - Base de datos I 2022-1)