

# Trabajo Práctico de Bases de Datos 2024

**Materia:** Bases de Datos

**Universidad:** UNSAM

**Profesores:**

Denise Martin

Diego Mirarchi

Pablo Núñez Monzon

---

## Lineamientos Generales del Trabajo Práctico

**1. Propósito:**

- Desarrollar competencias en el diseño de bases de datos, implementación de esquemas SQL y generación de consultas avanzadas.
- Facilitar la comprensión de los procesos de análisis, limpieza de datos, selección de motor de base de datos, y documentación.
- Proponer soluciones basadas en buenas prácticas para optimización, automatización y organización de datos.

**2. Entregables:**

- Documentación completa del análisis, diseño, y justificación de decisiones técnicas. Incluyendo diagramas y resultados obtenidos.
- Diagrama Entidad-Relación (DER) con justificación de entidades y relaciones.
- Scripts SQL utilizados para la creación de esquemas y consultas.
- Reporte final con análisis de consultas, rendimiento, y propuesta teórica de procedimientos, triggers y automatización.
- Presentación formal para portfolio.

**3. Fecha de Entrega:** viernes 22 de noviembre, 23:00 hs

<b>Índice</b>	<b>1</b>
Parte 1: Análisis y Limpieza de Datos	4
Parte 2: Selección y Justificación del Motor de Base de Datos	
Parte 3: Diseño de Modelo de Datos y Creación del Esquema	5
Parte 4: Carga y Manipulación de Datos	5
Parte 5: Procedimientos, Triggers y Automatización	7
Parte 6: Buenas Prácticas, Documentación y Presentación del TP	8

## **Objetivo General**

En la Universidad Nacional de San Martín, la necesidad de una gestión de datos efectiva se vuelve cada día más urgente. Con el crecimiento de la institución y el aumento de estudiantes, la universidad ha lanzado una licitación para desarrollar una nueva base de datos que permita capturar, organizar y analizar la información de los alumnos de manera integral.

La dirección académica ha identificado que la falta de visibilidad en los datos actuales limita su capacidad para tomar decisiones informadas y estratégicas. Actualmente, la información de los estudiantes – que incluye datos personales, intereses, y desempeño académico – se encuentra dispersa y poco accesible, lo cual dificulta evaluar tendencias, identificar necesidades, y realizar mejoras en la experiencia educativa.

Con esta iniciativa, la universidad busca implementar una solución que no solo centralice la información de sus estudiantes, sino que también ofrezca herramientas analíticas que permitan explorar y entender el perfil de cada alumno y del cuerpo estudiantil en general. El objetivo es claro: dotar a la institución de una base de datos robusta, eficiente y accesible que facilite la toma de decisiones informadas y promueva la innovación en los servicios académicos.

Para lograr esto, se espera que el equipo a cargo del proyecto desarrolle una base de datos relacional que permita gestionar y visualizar información clave, mejorando así la capacidad de análisis y optimización de procesos. Esta nueva plataforma de datos busca, en última instancia, ser una herramienta estratégica para impulsar la calidad educativa y el bienestar de los estudiantes de la Universidad Nacional de San Martín.

## Parte 1: Análisis y Limpieza de Datos

**Objetivo:** Asegurar que los datos proporcionados en el archivo inicial estén organizados y normalizados para su inclusión en la base de datos.

### 1. Análisis de los Datos Existentes:

- **Inspección de la Información:** Examinar los datos proporcionados en el archivo adjunto, identificando datos clave y campos incompletos.
  - **Organización y Normalización de Datos:**
    - Estandarizar el formato de nombres, roles y grupos que aparecen en el archivo adjunto y dividir campos según sea necesario (por ejemplo, separar nombre y apellido, validar información y estandarizar los nombres de las localidades y materias, eliminar duplicados, etc).
  - **Documentación Inicial:** Elaborar un informe que describa
    - Pasos de limpieza y organización realizados (data preprocessing).
    - Proponer soluciones para los datos incompletos o faltantes.
- 

## Parte 2: Selección y Justificación del Motor de Base de Datos

**Objetivo:** Seleccionar y justificar el motor de base de datos relacional que mejor se adapte a las necesidades del proyecto, complementando con una evaluación comparativa con una alternativa no relacional.

### 1. Investigación y Selección de Motor Relacional:

- Elegir un motor de base de datos relacional (por ejemplo, MySQL o PostgreSQL) y documentar las razones detrás de esta decisión, explicando:
  - **Ventajas del motor seleccionado:** Describir las características específicas que hacen que el motor relacional sea adecuado para este proyecto.
  - **Cualidades en términos de rendimiento, organización y escalabilidad:** Explicar cómo el motor relacional facilita la organización de datos estructurados y consultas complejas.
  - **Condicionamientos o limitaciones:** Mencionar posibles limitaciones, como requerimientos de infraestructura o restricciones de licencia que pudieran condicionar en la elección del motor.

### 2. Evaluación Comparativa con una Base de Datos No Relacional:

- **Descripción Alternativa con una Base de Datos No Relacional:** Proponer y describir cómo funcionaría hipotéticamente la misma solución utilizando un motor de base de datos no relacional (por ejemplo, MongoDB).
- **Justificación de la Alternativa No Relacional:** Explicar los beneficios y limitaciones de una base de datos no relacional en este contexto, considerando aspectos como flexibilidad en el almacenamiento, escalabilidad, y tipos de consulta.
- **Análisis Comparativo:** Resumir las ventajas y desventajas de ambos enfoques (relacional vs. no relacional), proporcionando una visión general de cuándo y por qué podría ser preferible cada uno para esta solución.

---

### Parte 3: Diseño de Modelo de Datos y Creación del Esquema

**Objetivo:** Crear un modelo E-R y un esquema SQL que estructuren y soporten la información de la base de datos.

#### 1. Diseño del Modelo Entidad-Relación:

- Crear un diagrama E-R que capture la información organizada en entidades como alumnos, materias, hobbies, ubicación y cualquier otra entidad que considere necesaria.
- Documentar y justificar cada entidad y relación en el modelo, asegurándose de alcanzar la tercera forma normal (3NF), de ser posible.

#### 2. Implementación del Esquema en SQL:

- Crear las tablas en SQL según el diseño E-R
  - Aplicar restricciones de integridad y claves foráneas.
  - Documentar las consultas utilizadas, incluir capturas de pantalla de la consola con las respuestas obtenidas.
- 

### Parte 4: Carga y Manipulación de Datos

**Objetivo:** Cargar los datos en el esquema creado y desarrollar consultas avanzadas para extraer información relevante.

#### 1. Carga Inicial de Datos:

- Importar los datos organizados en las tablas correspondientes.
- Verificar la integridad de los datos y la conformidad con las restricciones.

#### 2. Consultas Básicas y Reportes:

- Desarrollar consultas para obtener información relevante como se detalla a continuación:

- a) Crear un reporte que muestre, por cada localidad, el hobby predominante y la cantidad de alumnos que lo practican.

Localidad	Hobby más Común	Total Alumnos
San Martín	Fútbol	8
Villa Ballester	Lectura	5
San Miguel	Música	7
Tres de Febrero	Videojuegos	3

- b) Generar un análisis que incluya la cantidad de materias en curso por los alumnos y que detalle su experiencia previa en bases de datos (dividida entre relacional y no relacional). Top 5

Alumno	Materias en curso	Experiencia BD Relacional	Experiencia BD No Relacional
Alumno A	5	Sí	No
Alumno B	3	No	Sí
Alumno C	4	Sí	Sí
Alumno D	2	No	No

- c) Identificar la materia más popular en cada grupo de estudio y el porcentaje de alumnos de cada grupo que está inscrito en esas materias.

Grupo	Materia	Alumnos Inscritos	% del Grupo
Grupo 1	BD	5	100%
Grupo 2	Algoritmos y Estructuras	4	80%
Grupo 3	Seminario de Sistemas	2	40%
Grupo 4	Programación Avanzada	3	60%

- d) Crear un reporte que identifique alumnos con experiencia significativa en temas de bases de datos o actividades tecnológicas, clasificándolos como posibles mentores para el resto de los compañeros. TOP 5

Alumno	Experiencia BD Relacional	Experiencia BD No Relacional	Actividad
Alumno C	Sí	Sí	D. de Software
Alumno A	Sí	No	Base de datos
Alumno B	No	Sí	Redes

### 3. Consultas Avanzadas y Optimización:

- Desarrollar consultas avanzadas para:
  - a) Listar la cantidad de alumnos por localidad junto con el número de materias promedio en las que están inscritos, lo cual permitiría analizar la carga académica en diferentes áreas geográficas.

Localidad	Promedio de materias	Total Alumnos
San Martin	3,2	8
Tres de Febrero	2.5	5
Villa Ballester	4.1	6
San Miguel	3.8	7

### Interpretación:

Con este reporte, podríamos interpretar que en Villa Ballester, los estudiantes están inscritos en más materias en promedio, lo que podría indicar una mayor carga académica en esa zona. Esto puede llevar a decisiones como:

- Ofrecer más cursos virtuales para la localidad de Tres de Febrero, donde hay pocos estudiantes y menos materias inscritas.
  - Revisar la oferta académica para zonas con alta demanda como Localidad Villa Ballester.
- b) Generar un reporte que muestre qué alumnos pueden actuar como mentores de otros (match), basándose en afinidades de intereses y experiencia en temas específicos (como bases de datos). Se muestran alumnos junto a posibles mentores con experiencia en áreas similares.

Alumno	Potencial Mentor	Interés Común	Experiencia Mentor BD Relacional	Experiencia Mentor BD No Relacional
Alumno A	Alumno C	Música Rock	Sí	No
Alumno B	Alumno D	Fútbol	No	Sí
Alumno E	Alumno A	Videojuegos	Sí	Sí

- Evaluar/considerar el rendimiento de las consultas con grandes volúmenes de datos y proponer ajustes como por ejemplo: uso de índices de ser necesario, incluir justificación en cada caso.

---

## Parte 5: Procedimientos, Triggers y Automatización

**Objetivo:** Analizar y proponer soluciones teóricas para la automatización de procesos mediante procedimientos, triggers y estrategias de optimización.

### 1. Análisis Teórico de Procedimientos Almacenados:

- **Investigación:** Investigar el concepto de procedimientos almacenados en bases de datos, su propósito y beneficios en la automatización de tareas recurrentes.
- **Propuesta Teórica:**
  - Proponer un procedimiento teórico para registrar automáticamente la inscripción de un nuevo alumno en una materia.
  - Explicar cómo este procedimiento podría facilitar la gestión de datos y reducir la carga manual.
- **Documentación:** Redactar una breve explicación sobre el funcionamiento del procedimiento y su implementación ideal en el motor de base de datos seleccionado.

## 2. Análisis Teórico de Triggers:

- **Investigación:** Estudiar el concepto de triggers en bases de datos, cuándo se usan y por qué son útiles para mantener la integridad de los datos.
- **Propuesta de Trigger Teórico:**
  - Proponer un trigger que registre un log automáticamente cuando se modifique la información personal de un alumno, explicando su utilidad en auditorías de datos.
- **Documentación:** Describir el trigger en detalle, incluyendo cuándo se activaría, qué acciones ejecutaría y por qué mejoraría la gestión de la base de datos.

## 3. Propuesta de Estrategias de Automatización:

- **Investigación de Automatización en Bases de Datos:** Investigar técnicas básicas de automatización en bases de datos, como los índices de optimización, el uso de vistas, y la programación de tareas (backups, actualizaciones, etc.).
- **Propuesta de Estrategia Teórica:** Proponer al menos una estrategia de automatización para la base de datos de alumnos, por ejemplo: una rutina de backup o cómo implementar índices y vistas para mejorar el rendimiento.
- **Documentación:** Incluir esta propuesta en la documentación final, explicando cómo la automatización y optimización seleccionada contribuirían a la eficiencia y seguridad de los datos en un entorno real.

---

## Parte 6: Buenas Prácticas, Documentación y Presentación del TP

**Objetivo:** Documentar el proceso completo y presentar el proyecto en un formato profesional y orientado al cliente.

### 1. Buenas Prácticas y Documentación Completa:

- Documentar el proceso completo de desarrollo de este trabajo práctico, incluyendo:
  - Organización inicial de datos, diseño del modelo, elección del motor y justificación.
  - Ejecución y explicación de consultas y procedimientos implementados.
  - Las queries utilizadas y capturas de la consola con los resultados obtenidos.
- Explicar las buenas prácticas aplicadas (ej. uso de índices, normalización y backups)

### 2. Presentación final del documento:

- Preparar una presentación clara, concisa y profesional que incluya:
  - Descripción del modelo y el esquema de la base de datos.
  - Justificación del motor seleccionado y sus beneficios de la solución.
  - Ejemplos de reportes o análisis que pueden extraerse de la base de datos.