

Especificaciones del proyecto a entregar para la Evaluación a  
Título de Suficiencia 2025-2

## Bases de Datos Avanzadas

Nombre: \_\_\_\_\_ Boleta: \_\_\_\_\_

### Instrucciones Generales y Criterios de Evaluación

- **Objetivo:** Evaluar la capacidad del alumno para diseñar, implementar y analizar soluciones de datos complejas, a partir de bases de datos no relacionales, tecnologías OLAP, bancos de datos y diversos modelos NoSQL.
- **Modalidad:** Examen práctico individual. El alumno deberá presentar su computadora con el entorno de Docker pre-configurado para la revisión en vivo.
- **Entorno de Trabajo:** Toda la solución debe estar orquestada y ser ejecutable mediante un único comando `docker-compose up --build`.
- **Entrega:** Se deberá compartir un enlace a una carpeta de Google Drive con todo el proyecto y un documento PDF con la solución al correo [ghurtadoa@ipn.mx](mailto:ghurtadoa@ipn.mx).
- **Revisión:** La revisión del proyecto será de forma presencial el lunes 14 de julio de 2025 de las 7:00 a las 10:00 hrs., en la sala 1110, preguntar por el profesor Gabriel Hurtado Avilés. No se considerarán revisiones fuera del tiempo asignado. La valoración está sujeta al criterio del profesor, la correcta ejecución de las implementaciones y la calidad de las justificaciones. Incluir en el reporte la solución de ambas partes del proyecto, correspondiendo al 50% del valor de la calificación del ETS, si cumple con los requerimientos adecuadamente. El restante 50% del ETS será el examen escrito presencial programado según el calendario oficial.  
Cualquier duda con respecto al desarrollo del proyecto, se resolverá por correo electrónico ([ghurtadoa@ipn.mx](mailto:ghurtadoa@ipn.mx)) del 9 al 11 de julio de 2025.

# Parte 1: Sistema de Análisis de Riesgo Sísmico (5 Puntos)

## Escenario

Se debe desarrollar una solución de Business Intelligence completa para analizar el impacto de los eventos sísmicos en México. La solución deberá integrar datos de sismos, demografía y economía para permitir un análisis multidimensional.

## Fuentes de Datos Oficiales

Es obligatorio utilizar los siguientes conjuntos de datos públicos:

- **Datos de Sismos:** Catálogo histórico del Servicio Sismológico Nacional (SSN).
  - <http://www2.ssn.unam.mx:8080/catalogo/>
- **Datos de Población:** Censo de Población y Vivienda 2020 del INEGI (Resultados por localidad ITER, archivo `iter_00_cpvt2020_csv.zip`).
  - <https://www.inegi.org.mx/datosabiertos/>
- **Datos Económicos:** Censos Económicos 2019 del INEGI (Datos nacionales, archivo `ce2019_nac_csv.zip`).
  - <https://www.inegi.org.mx/datosabiertos/>

## Requerimientos

1. [1.5 Puntos] **Diseño e Implementación del Data Warehouse y ETL:**
  - Diseñe y justifique un **esquema en estrella** para el DWH.
  - Implemente un proceso **ETL** (usando una herramienta o script) para limpiar, transformar y cargar los datos en un DWH sobre **PostgreSQL**. El DWH final dockerizado debe estar completamente poblado y listo para ser consultado una vez que los servicios se levanten.
  - Entregue los scripts SQL de creación y poblado, y el flujo/script ETL.
2. [2.0 Puntos] **Implementación del Cubo OLAP:**
  - Utilice un **servidor OLAP dedicado** (ej. **Pentaho BI Server** con JDK 8) y **Schema Workbench** para diseñar y publicar un cubo OLAP. *No se aceptarán simulaciones con vistas SQL.*
  - Entregue el archivo de esquema Mondrian (`.xml`).
  - Demuestre con capturas de pantalla el análisis del cubo en **Saiku Analytics**.



- Incluya el código en lenguaje **MDX** para resolver las consultas de negocio. Su código debe demostrar explícitamente operaciones de navegación y análisis sobre el cubo, tales como **Drill-down, Roll-up, Slice, Dice y Pivot**.

3. [1.5 Puntos] **Análisis de Negocio y Documentación Analítica:**

- Su solución (cubo OLAP) debe ser capaz de responder a las siguientes **Consultas de Interés para el Negocio**:
  - Calcular un **Índice de Riesgo Sísmico** por estado (ej. frecuencia  $\times$  magnitud promedio).
  - Estimar la **Población Total en Riesgo** para los 5 estados con mayor actividad sísmica.
  - Cuantificar el **Impacto Económico Potencial** (Producción Bruta Total) en las zonas de alto riesgo.
  - Mostrar la tendencia de **Número de sismos por año** para la última década.
- **Documentación Analítica (como parte de la solución):** En su documento, incluya un análisis comparativo del **plan de mantenimiento** de una BD transaccional vs. el DWH, y discuta las **capacidades y limitaciones** de su arquitectura de DWH.

# Parte 2: Sistema de Gestión de Motores (5 Puntos)

## Escenario

Se requiere modelar el sistema de diseño, fabricación y ensamblaje de motores de una empresa, basándose en el siguiente diagrama Entidad-Relación.

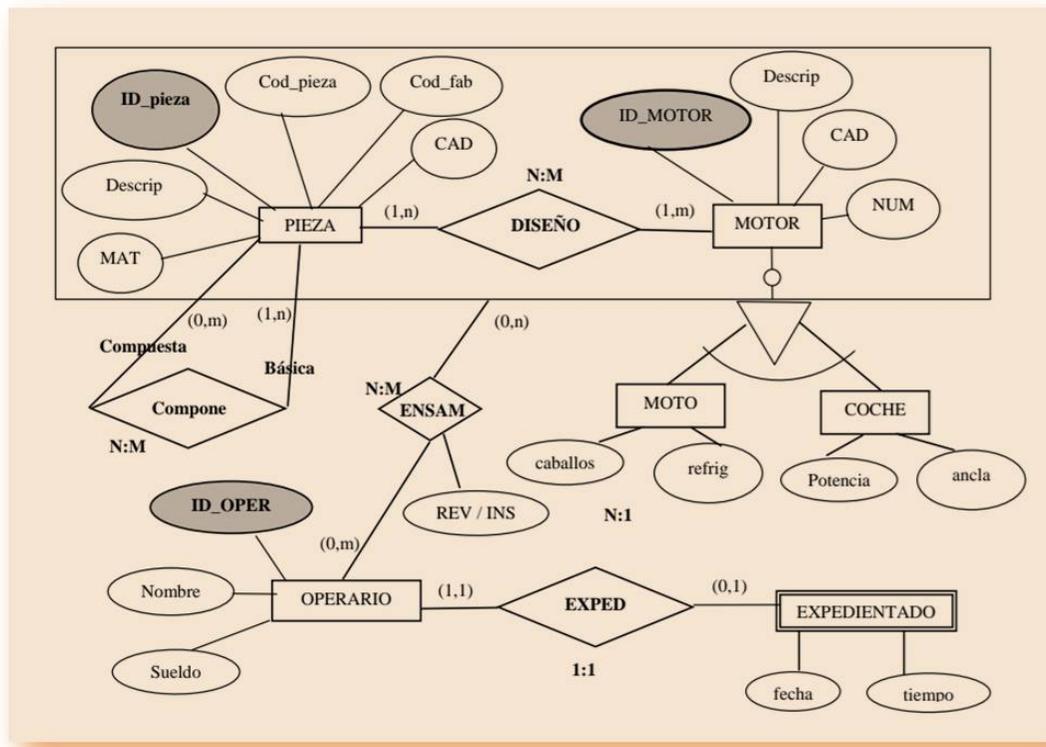


Figura 1: Modelo Entidad-Relación para el sistema de gestión de motores.

## Requerimientos

Realice la transformación del escenario a los siguientes modelos NoSQL. Para cada modelo, debe plantear **3 preguntas de negocio** relevantes y desarrollar las consultas para resolverlas.

### 1. [1.25 Puntos] Modelo Relacional-Objetual (PostgreSQL):

- Diseñe la transformación al modelo relacional-objetual. Emplee características de PostgreSQL como **tipos de datos compuestos (CREATE TYPE)** y/o **herencia de tablas**.

- Incluya un diagrama de clases o un esquema que represente su propuesta.
- Incluya el código SQL de creación de tipos/tablas, inserción de datos y la resolución de sus 3 consultas de negocio.

2. [1.25 Puntos] **Modelo Semiestructurado (Documentos y XML):**

- Diseñe la transformación a un modelo documental en **MongoDB**. Justifique sus decisiones de embebido vs. referencia.
- Incluya el documento **JSON Schema** para una colección principal (ej. 'motores').
- Incluya un documento **XML Schema (XSD)** que represente la estructura de un documento de su colección principal.
- Incluya el código en la sintaxis del Shell de MongoDB para la creación, poblado y la resolución de sus 3 consultas de negocio.

3. [1.25 Puntos] **Modelo Clave-Valor (Redis):**

- Diseñe una solución utilizando un modelo Clave-Valor para un caso de uso apropiado (ej. caché para especificaciones de piezas, seguimiento de estado de ensamblaje en tiempo real).
- Defina la estructura de claves y el tipo de dato de Redis a utilizar.
- Incluya los comandos de Redis para poblar los datos y para resolver sus 3 consultas de negocio.

4. [1.25 Puntos] **Modelo Orientado a Grafos (Neo4j):**

- Diseñe la transformación al modelo orientado a grafos, empleando el SGBD **Neo4j**.
- Incluya una imagen del grafo de su solución propuesta (un ejemplar con las relaciones posibles).
- Incluya el código de las operaciones de creación en lenguaje **Cypher**.
- Incluya las sentencias en Cypher que resuelvan sus 3 consultas de negocio.