

# Bases de Datos para Ingeniería

## Prueba Final

Noviembre 2024

### Indicaciones Generales:

- La duración de la prueba es de **tres (3)** horas.
- En la prueba **NO** se permite consultar material.
- Empezar cada ejercicio en una hoja nueva.
- Escribir con lápiz y de un solo lado de las hojas.
- Numerar todas las hojas. Incluir en cada hoja la cédula y el nombre. En la primer hoja, incluir la cantidad de hojas que se entregan.

### Ejercicio 1. (35 puntos)

Un gobierno departamental desea implementar un sistema de información para la gestión del transporte público. Los ómnibus son operados por empresas de transporte colectivo de pasajeros, las cuales se identifican por su nombre. Además, de cada empresa se conoce su dirección (calle y número) y una lista de teléfonos de contacto.

De los ómnibus, que se identifican por su matrícula, se conoce su marca, el año de empadronamiento y la cantidad de asientos.

Los recorridos son identificados por un nombre (ej: Ciudad Vieja- Punta Carretas) y representan una secuencia de paradas, donde para cada una de ellas se conoce su posición dentro del recorrido (por ejemplo, la parada 0 es el origen). Cada parada está identificada por sus coordenadas geográficas (latitud y longitud), y tiene un nombre que también la identifica.

Las empresas organizan los recorridos que ofrecen a través de líneas de transporte. Por ejemplo, la empresa CUTCSA realiza el recorrido “Ciudad Vieja - Punta Carretas” con la línea 117). Cada línea tiene un nombre compuesto por una serie de letras y números (ej: 2A, 117, etc.) y este nombre la identifica dentro de la empresa. Cada línea realiza un único recorrido. Además, las empresas asignan ómnibus de su flota a cada una de las líneas que ofrecen.

En las empresas trabajan empleados, quienes pueden ser conductores, guardas o personal administrativo. De cada empleado se conoce su número de identificación (CI), su nombre, y su número de teléfono. Los empleados pueden estar asociados a una o más empresas, y se registra la fecha en que ingresaron a trabajar. Los conductores tienen registrada la fecha en que vence su libreta de conducir. Algunos conductores pueden trabajar también como guardas, mientras que el personal administrativo no realiza funciones de guarda ni de conductor. Cada ómnibus con asiento para guarda debe tener asignado un guarda. Además, los conductores están asignados a ómnibus de líneas específicas.

Algunos usuarios del sistema de transporte están registrados en el sistema. Se conoce su CI, nombre y teléfono. Estos usuarios poseen una tarjeta que les permite acceder al transporte público. El sistema registra la fecha y la hora en que el usuario se sube a un ómnibus en una parada determinada y en qué línea viaja. Además, se registra el tipo de boleto que el usuario compra, si es de una hora o de dos horas.

**Se pide: Modelo Entidad-Relación completo del problema, incluyendo Restricciones No Estructurales.**

**Solución.**

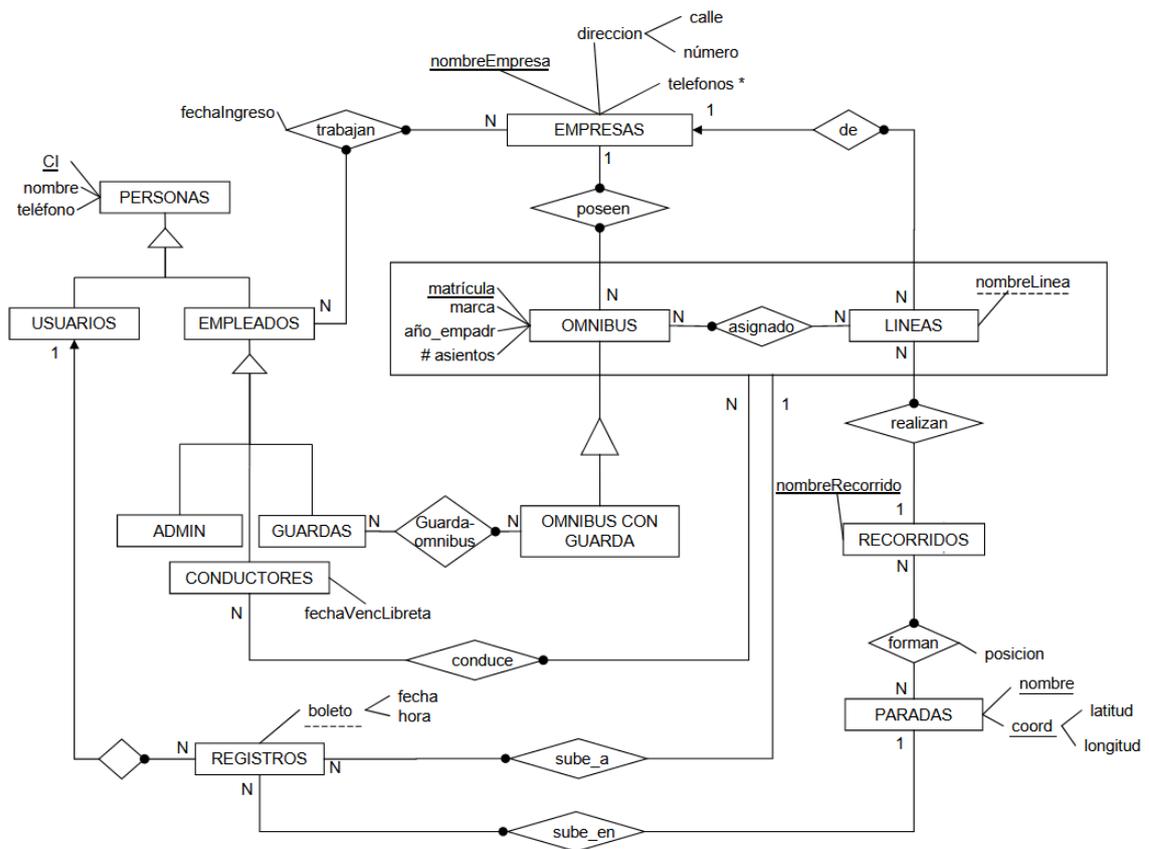


Figura 1: Solución MER

**RNEs**

1. Si un ómnibus es asignado a una línea dicha línea es de la empresa a la que pertenece el ómnibus.
2. Si un conductor conduce un ómnibus de una línea el conductor trabaja para la empresa que posee ese ómnibus
3. Si un usuario sube a un ómnibus de una línea en una parada, esa parada forma el recorrido que realiza la línea.
4. Si un guarda es guarda-omnibus de algún ómnibus con guarda entonces el guarda trabaja para la empresa que posee el ómnibus.
5. Para todo par de paradas p1 y p2 que forman un recorrido la posición es diferente.
6. El año de la fecha de vencimiento de la libreta de chofer tiene que ser mayor o igual al máximo año de ingreso de todas las empresas en las cuales trabaja.
7.  $ADMIN \cup CONDUCTORES \cup GUARDAS = EMPLEADOS$
8.  $ADMIN \cap CONDUCTORES = \emptyset$
9.  $ADMIN \cap GUARDAS = \emptyset$

**Ejercicio 2.** (30 puntos)

Considere el siguiente esquema relacional donde una empresa almacena información sobre las reuniones que ha realizado a lo largo del tiempo.

**PERSONAS (email, nombre, piso)**

Datos de cada una de las personas que asisten a reuniones, su dirección de email (que la identifica), su nombre y el número de piso del edificio donde se encuentra su espacio de trabajo.

**REUNIONES (fecha, hora, nroSala, emailOrg)**

Datos de las reuniones. Ellos son la fecha, la hora de comienzo y el número de sala donde se realiza la reunión y la dirección de email de la persona que organiza la reunión.

**SALAS (nroSala, piso, capacidad)**

Datos de las salas de reunión. Ellos son el número de sala (que la identifica), el número de piso del edificio donde está ubicada y su capacidad máxima.

**ASISTENCIA (fecha, hora, nroSala, email)**

Se registran los emails de cada una de las personas que asiste a una reunión.

Se cumplen las siguientes dependencias de inclusión:

$$\begin{aligned}\Pi_{\text{nroSala}}(\text{REUNIONES}) &\subseteq \Pi_{\text{nroSala}}(\text{SALAS}) \\ \Pi_{\text{emailOrg}}(\text{REUNIONES}) &\subseteq \Pi_{\text{email}}(\text{PERSONAS}) \\ \Pi_{\text{fecha, hora, nroSala}}(\text{ASISTENCIA}) &\subseteq \Pi_{\text{fecha, hora, nroSala}}(\text{REUNIONES}) \\ \Pi_{\text{email}}(\text{ASISTENCIA}) &\subseteq \Pi_{\text{email}}(\text{PERSONAS})\end{aligned}$$

**Nota:** Ninguna de las tablas se encuentra vacía.

**Se pide:**

**(a) Resolver en Álgebra Relacional las siguientes consultas**

- I. Devolver el email y nombre de las personas que asistieron a reuniones en todas las salas ubicadas en el piso 2.
- II. Devolver el email y nombre de las personas que han asistido a reuniones en un piso diferente al de su espacio de trabajo.

**(b) Resolver en SQL las siguientes consultas**

- I. Devolver los emails de los organizadores de reuniones realizadas en salas con capacidad mayor a 10 personas pero con menos de 5 asistentes.
- II. Devolver parejas de personas, representadas por su email, donde la primer persona no asistió a ninguna de las reuniones organizadas por la segunda.

**Solución.**

**Álgebra Relacional:**

1. Devolver el email y nombre de las personas que asistieron a reuniones en todas las salas ubicadas en el piso 2.

$$\begin{aligned}\text{SalasDelPiso2} &= \Pi_{\text{nroSala}}(\sigma_{\text{piso}=2}\text{SALAS}) \\ \text{PersonasAsistencia} &= \text{PERSONAS} * \text{ASISTENCIA} \\ \text{Resultado} &= \Pi_{\text{email, nombre, nroSala}}(\text{PersonasAsistencia}) \div \text{SalasDelPiso 2}\end{aligned}$$

2. Devolver el email y nombre de las personas que han asistido a reuniones en un piso diferente al de su espacio de trabajo.

$$\begin{aligned}\text{AsistenciaSalas} &= \text{ASISTENCIA} * \text{SALAS} \\ \text{RenPersonas} &= \rho_{\text{piso, email} \rightarrow \text{pisoTrabajo, emailPersona}}(\text{PERSONAS}) \\ \text{Resultado} &= \Pi_{\text{email, nombre}}(\text{AsistenciaSalas} \bowtie_{\text{piso!}=\text{pisoTrabajo, email}=\text{emailPersona}} \text{RenPersonas})\end{aligned}$$

**SQL:**

1. Devolver los emails de los organizadores de reuniones realizadas en salas con capacidad mayor a 10 personas pero con menos de 5 asistentes.

```
SELECT DISTINCT R.emailOrg
FROM REUNIONES R
JOIN SALAS S ON R.nroSala = S.nroSala
JOIN ASISTENCIA A ON R.fecha = A.fecha
                AND R.horaComienzo = A.horaComienzo
                AND R.nroSala = A.nroSala
GROUP BY R.emailOrg, S.capacidad, R.nroSala, R.fecha, R.horaComienzo
HAVING S.capacidad > 10
AND COUNT(A.email) < 5;
```

2. Devolver parejas de personas, representadas por su email, donde la primer persona no asistió a ninguna de las reuniones organizadas por la segunda.

```
SELECT P1.email , P2.email
FROM PERSONAS P1 JOIN PERSONAS P2 ON P1.email != P2.email
WHERE NOT EXISTS (
    SELECT 1
    FROM REUNIONES R NATURAL JOIN ASISTENCIA A
    WHERE R.emailOrg = P2.email
    AND A.email = P1.email
)
```

**Ejercicio 3.** (25 puntos)

Considere el MER de la figura 2 con sus respectivas RNEs.

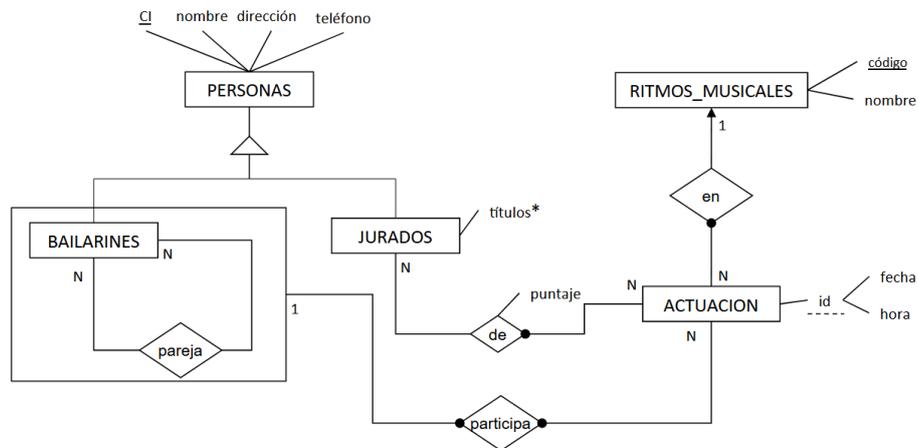


Figura 2: MER - Concurso de baile

**RNEs:**

- Un bailarín no puede formar pareja con sí mismo.  
 $\forall x \in Bailarines. \langle x, x \rangle \notin pareja$
- Las parejas, son simétricas.  
 $\forall x \in Bailarines. \forall y \in Bailarines. (\langle x, y \rangle \in pareja \rightarrow \langle y, x \rangle \in pareja)$
- Un jurado no puede participar como bailarín.  
 $Jurados \cap Bailarines \equiv \emptyset$

**Se pide:** Teniendo en cuenta el MER y las dependencias no estructurales presentadas, elaborar un esquema relacional que incluya los esquemas relación y las dependencias de inclusión.

**Solución.**

Primero modelaremos las entidades, comenzando por los ritmos musicales. Como son una entidad fuerte sin atributos multivaluados, se modela como una única tabla.

RITMOS\_MUSICALES(codigo, nombre)

Luego, como ACTUACION es una entidad débil de RITMOS\_MUSICALES y tiene un atributo estructurado, se modela de la siguiente forma:

ACTUACION(fecha, hora, codigo)

siendo fecha y hora los campos del atributo id y codigo el codigo del ritmo musical. Además, se agrega la siguiente dependencia de inclusión:

$$\Pi_{codigo}(ACTUACION) \subseteq \Pi_{codigo}(RITMOS\_MUSICALES)$$

Luego, las únicas entidades que nos faltan modelar son las de la categorización de PERSONAS, que se puede modelar de dos formas:

1. Al ser disjunta, podemos generar una tabla para las personas y agregar un atributo que sea tipo. La tabla personas quedaría:

PERSONAS(ci, nombre, direccion, telefono, tipo)

Con el atributo tipo variando en los valores 'Bailarin', 'Jurado' y 'Otro'. Además, se tiene que generar una tabla para los títulos de los jurados; la cual quedaría:

TITULOS(ciJurado, titulo)

Al usar este modelo, las dependencias de inclusión requieren del uso de selección según atributo tipo, por ejemplo,

$$\Pi_{ciJurado}(TITULOS) \subseteq \Pi_{ci}(\sigma_{tipo='Jurado'}(PERSONAS))$$

2. La otra forma de modelarlo es con una tabla por entidad:

PERSONAS(ci, nombre, direccion, telefono)

BAILARINES(ciPersona)

JURADOS(ciPersona)

Cabe destacar que se podría agregar el atributo titulo a JURADOS, pero dado que puede generar valores nulos cuando el jurado no tiene ningún título decidimos modelarlo en una tabla separada.

TITULOS(ciJurado, titulo)

Luego tenemos que agregar las siguientes dependencias de inclusión:

$$\Pi_{ciPersona}(BAILARINES) \subseteq \Pi_{ci}(PERSONAS)$$

$$\Pi_{ciPersona}(JURADOS) \subseteq \Pi_{ci}(PERSONAS)$$

$$\Pi_{ciJurado}(TITULOS) \subseteq \Pi_{ciPersona}(JURADOS)$$

Además, como JURADOS y BAILARINES son disjuntos, agregamos las siguientes dependencias:

$$\Pi_{ciPersona}(BAILARINES) \cap \Pi_{ciPersona}(JURADOS) = \emptyset$$

$$\Pi_{ciPersona}(PERSONAS) \cap \Pi_{ciPersona}(BAILARINES) = \emptyset$$

Elegimos la opción con tres tablas para modelar las PERSONAS, en particular porque las dependencias de inclusión quedan más sencillas.

Ahora pasamos a modelar las relaciones. Primero veremos la relación pareja, y como esta relación relaciona dos bailarines, necesitamos el identificador de ambos:

PAREJA(ciBailarin1, ciBailarin2)

Luego, se agregan las siguientes dependencias de inclusión:

$$\Pi_{ciBailarin1}(PAREJA) \subseteq \Pi_{ciPersona}(BAILARINES)$$

$$\Pi_{ciBailarin2}(PAREJA) \subseteq \Pi_{ciPersona}(BAILARINES)$$

Por otro lado, como un bailarín no puede ser pareja con si mismo agregamos la siguiente dependencia:

$$\Pi_{ciBailarin1,ciBailarin2}(PAREJAS) \subseteq \Pi_{ci1Bailarin1,ciBailarin2}(\sigma_{ciBailarin1 \neq ciBailarin2}(PAREJAS))$$

También como las parejas son simétricas se agrega la siguiente dependencia:

$$\Pi_{ciBailarin1,ciBailarin2}(PAREJAS) \subseteq \Pi_{ciBailarin2,ciBailarin1}(PAREJAS)$$

La siguiente relación que vamos a modelar es la de jurados y actuación, que tiene los identificadores de jurado y actuación como clave, y el atributo puntaje de la relación.

JURADOS\_DE\_ACTUACION(ciJurado, fecha, hora, codigo, puntaje)

Se agrega la dependencia:

$$\Pi_{ciJurado}(JURADOS\_DE\_ACTUACION) \subseteq \Pi_{ciPersona}(JURADOS)$$

También, como la relación es total del lado de actuación, agregamos la dependencia:

$$\Pi_{fecha,hora,codigo}(JURADOS\_DE\_ACTUACION) = \Pi_{fecha,hora,codigo}(ACTUACION)$$

Lo último que nos falta modelar es la relación participa, la cual es una relación 1 a N con totalidad del lado N, entonces, podemos agregar la clave de PAREJA a la tabla ACTUACION:

ACTUACION(fecha, hora, codigo, ciBailarin1, ciBailarin2)

Además, como existe totalidad en participa del lado de las parejas, agregamos la siguiente dependencia:

$$\Pi_{ciBailarin1,ciBailarin2}(ACTUACION) = \Pi_{ciBailarin1,ciBailarin2}(PAREJAS)$$

Entonces, las tablas del esquema relacional son:

RITMOS\_MUSICALES(codigo, nombre)

ACTUACION(fecha, hora, codigo, ciBailarin1, ciBailarin2)  
 PERSONAS(ci, nombre, direccion, telefono)  
 BAILARINES(ciPersona)  
 JURADOS(ciPersona)  
 PAREJA(ciBailarin1, ciBailarin2)  
 TITULOS(ciJurado, titulo)  
 JURADOS\_DE\_ACTUACION(ciJurado, fecha, hora, codigo, puntaje)

Dónde se cumplen las dependencias de inclusión:

$\Pi_{codigo}(ACTUACION) \subseteq \Pi_{codigo}(RITMOS\_MUSICALES)$   
 $\Pi_{ciPersona}(BAILARINES) \subseteq \Pi_{ci}(PERSONAS)$   
 $\Pi_{ciPersona}(JURADOS) \subseteq \Pi_{ci}(PERSONAS)$   
 $\Pi_{ciJurado}(TITULOS) \subseteq \Pi_{ciPersona}(JURADOS)$   
 $\Pi_{ciPersona}(BAILARINES) \subsetneq \Pi_{ciPersona}(JURADOS)$   
 $\Pi_{ciPersona}(PERSONAS) \subsetneq \Pi_{ciPersona}(BAILARINES)$   
 $\Pi_{ciBailarin1}(PAREJA) \subseteq \Pi_{ciPersona}(BAILARINES)$   
 $\Pi_{ciBailarin2}(PAREJA) \subseteq \Pi_{ciPersona}(BAILARINES)$   
 $\Pi_{ciBailarin1,ciBailarin2}(PAREJAS) \subseteq \Pi_{ci1Bailarin1,ciBailarin2}(\sigma_{ciBailarin1 \neq ciBailarin2} PAREJAS)$   
 $\Pi_{ciBailarin1,ciBailarin2}(PAREJAS) \subseteq \Pi_{ciBailarin2,ciBailarin1}(PAREJAS)$   
 $\Pi_{ciJurado}(JURADOS\_DE\_ACTUACION) \subseteq \Pi_{ciPersona}(JURADOS)$   
 $\Pi_{fecha,hora,codigo}(JURADOS\_DE\_ACTUACION) \subseteq \Pi_{fecha,hora,codigo}(ACTUACION)$   
 $\Pi_{fecha,hora,codigo}(ACTUACION) \subseteq \Pi_{fecha,hora,codigo}(JURADOS\_DE\_ACTUACION)$   
 $\Pi_{ciBailarin1,ciBailarin2}(ACTUACION) = \Pi_{ciBailarin1,ciBailarin2}(PAREJAS)$

#### Ejercicio 4. (10 puntos)

- (a) Según la literatura de Calidad de Datos vista en el curso ¿en que consiste un Modelo de Calidad de Datos? Explique brevemente qué es y para que sirve.

##### Solución.

Los Modelos de Calidad de Datos son artefactos que permiten definir y representar los aspectos relevantes sobre la calidad de datos en un dominio en particular. Un Modelo de Calidad está compuesto por dimensiones que representan las cuestiones de calidad de datos relevantes para un determinado escenario o conjunto de escenarios, donde cada Dimensión captura una faceta a alto nivel de la calidad. A su vez, los Factores representan aspectos particulares de cada dimensión. Usualmente también se definen Métricas, que pueden verse como instrumentos que definen la forma de medir un factor de calidad y Métodos de Medición, que son los procesos que implementan las métricas y generan valores.

- (b) Mencione escenarios en los que podría ser adecuado el uso de una base de datos de grafos.

##### Solución.

Las bases de datos de grafos son particularmente útiles en escenarios donde las relaciones entre los elementos del dominio cumplen un rol central. Escenarios donde los tipos o números de relaciones pueden cambiar con el tiempo, y se necesita flexibilidad para adaptarse a estas modificaciones son ideales para este tipo de bases. También suelen ser adecuadas cuando las consultas frecuentes requieren navegar múltiples niveles de relaciones, recorrer caminos, vecindades, o patrones complejos de conexión. Suelen ser particularmente útiles para representar datos con estructura en forma de árbol, jerarquía o red compleja, dado que estas estructuras son difíciles de modelar eficientemente en bases de datos relacionales. Por último, y debido a que tienen un mayor nivel de flexibilidad en la definición de sus esquemas que las bases relacionales, son adecuadas para modelar realidades donde las entidades tienen cierto nivel de heterogeneidad.