



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN
IIC2413 - BASES DE DATOS

Proyecto semestral Etapa 2: SQL y Manejo de Datos

Administrativo

- El proyecto es individual.
- **Fecha de entrega:** Miércoles 14 de mayo de 2025, 23:59.
- **Fecha y hora máxima de entrega atrasada:** sábado 17 de mayo de 2025, 23:59.
- **Entregables:** Scripts de SQL para creación de tablas, carga de datos desde CSV y consultas.
- **Lugar de entrega:** Directorio Sites/E2/ de su cuenta personal del servidor del curso (bdd1.ing.puc.cl). **NO SE EVALUARÁN OTROS DIRECTORIOS.**
- **Consultas exclusivamente por ISSUES.** Revise DISCUSSIONS en GitHub antes de preguntar.

Objetivo de la etapa 2

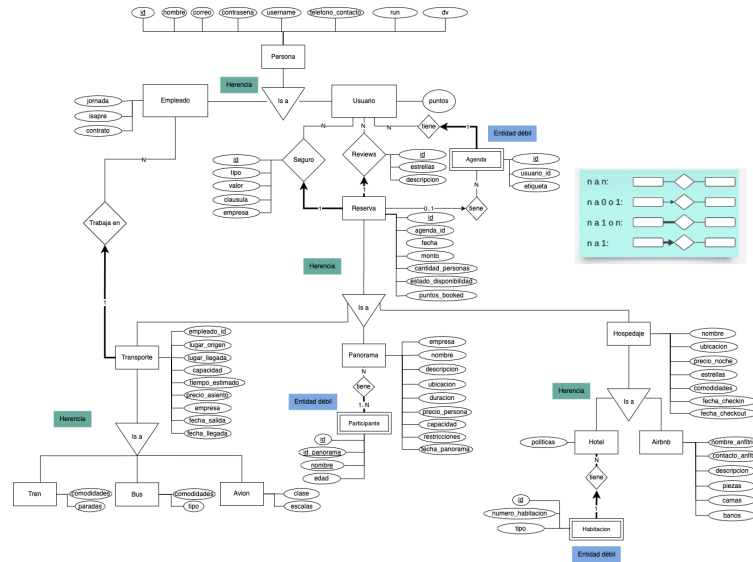
En esta etapa se deben trasladar los datos actuales desde archivos CSV a una base de datos local montada en PostgreSQL. Para ello debes utilizar archivos SQL y, con este mismo lenguaje, realizar consultas con tal de extraer información útil para la plataforma **Booked.com**.

1. Contexto

Dado tu desempeño como programador junior en **Booked.com**, se te ha solicitado trasladar la información de archivos CSV a PostgreSQL. Deberás desarrollar consultas SQL que permitan obtener información clave.

2. Tareas a realizar

A continuación se entrega el modelo E/R con el cual se deben realizar las tareas indicadas



2.1. Creación de Esquema y Carga de Datos (10 pts)

■ Desarrollar scripts en SQL que:

- Crear, en la base de datos, el Esquema Relacional acorde al modelo entregado. Para ello deben crear el archivo `archivos/crear_tablas.sql` con las sentencias DDL de creación de cada tabla con sus respectivos atributos, llaves primarias, foráneas, dominios y restricciones de integridad. **Los nombres de los atributos deben ser los indicados en el modelo (6 pts)**
- Crear en el archivo `poblar_tablas.sql` las sentencias SQL para la carga de datos en las tablas correspondientes desde los archivos CSV y luego ejecuten la carga. Para esto pueden a) crear una tabla auxiliar por cada set de datos sin restricciones de integridad y luego traspasar los datos a la instancia que tiene el esquema con las restricciones o b) directamente ir leyendo los CSV y cargando los datos a la instancia con el esquema que contiene las restricciones. En ningún caso se puede manipular los datos manualmente antes de cargarlos (con excel, datagrip, etc.) todas las sentencias de carga desde los CSV a la instancia debe estar en el archivo `crear_tablas.sql` (3 pts)
- Las tuplas que no cumplan con las restricciones de integridad deben ser registradas en los archivos indicados en el punto 7 del enunciado. El DBMS, no permitirá la carga de las tuplas que violen las Restricciones de integridad. (1 pts)

2.2. Consultas SQL (18 pts)

- Crear las siguientes consultas SQL en archivos según el siguiente formato `1a.sql`, `2a.sql`, `2b.sql`, etc. (algunas consultas tienen partes a y b, otras solo a).
- Cada consulta deberá incluir una breve descripción comentada al inicio.
- Las consultas deben estar dentro del directorio `Sites/E2/consultas/`
- Deben usar los nombres de los atributos (originales y nuevos) según se indica
- Las consultas son las siguientes:

1. Reservas mensuales

- a) Entregue la `cantidad_reservas` por mes y el `monto_total` acumulado en el mes. Se deben de entregar los atributos `mes` (formato MM-YYYY), `cantidad_reservas` y `monto_total`. (1.7 pts)

2. Hospedajes disponibles

- a) Debes entregar los hospedajes que están disponibles, indicando el `nombre_hospedaje`, su ubicación, el `precio_noche` y su calificación en estrellas. Se deben ordenar las tuplas basado en el número de estrellas, de forma descendente. El output debe incluir los atributos `nombre_hospedaje`, `ubicación`, `estrellas` y `precio_noche`. (1.3 pts)
- b) Se debe refinar la búsqueda anterior a los hoteles que tengan más de una política registrada. El output debe estar ordenado por cantidad de estrellas de forma descendente y, en caso de empate, por precio de forma ascendente. Te piden solo incluir los atributos `nombre_hospedaje`, `ubicación`, `estrellas` y `precio_noche`. (1.7 pts)

3. Calificación panoramas

- a) Se debe obtener el promedio de estrellas de los panoramas según su nombre y la cantidad de reviews que tienen. Debes obtener los atributos `nombre_panorama`, `prom_estrellas` y `cant_reviews`. (1.3 pts)
- b) Se debe realizar una consulta que obtenga el promedio de estrellas, cantidad de reviews y la última descripción hecha para cada panorama. El output debe incluir `nombre_panorama`, `prom_estrellas`, `cant_reviews` y `ult_comentario`. (1.7 pts)

4. Temporada alta de panoramas

- a) Se quiere obtener los meses con mayor cantidad de panoramas, por lo que debes entregar los top 3 meses con mayor cantidad de panoramas, incluyendo la cantidad de personas que participaron en panoramas y el monto total obtenido. Te piden solo incluir los atributos `mes` (MM-YYYY), `cantidad_panoramas`, `cantidad_participantes` y `monto_ganado`. (1.7 pts)

5. Usuarios más fieles

- a) Debes obtener los top 5 usuarios con mayor cantidad de puntos, mostrando la cantidad de reservas futuras a partir del 27 de mayo. Te piden incluir los atributos **nombre_usuario**, **puntos** y **cantidad_reservas**. (1.7 pts)

6. Actividad usuarios

- a) La empresa quiere poder visualizar los usuarios que, en total, tengan más de 70 reservas o más de 1000 puntos. Se debe obtener el **nombre_usuario**, la **cantidad_reservas** y los **puntos**. Se deben entregar los datos ordenados por puntos de forma decreciente. (1.7 pts)

7. Paseo de fin de semestre (5.2 pts)

Los ayudantes de Bases de datos están organizando un viaje de fin de semestre para gastarse todas las UT ganadas. Para ello los ayudantes de bienestar (Cata Bienestar y Jorge Bienestar) están organizando un viaje a Italia, todo a cargo de Lucas Viajero. Como los gustos de los ayudantes son un poco excentricos han decidido pedirle a Booked que les genere el viaje completo. Para ello pasarán una semana entre 1 y 6 de agosto en la ciudad de Corleone, Sicilia (Italia es un destino tradicional de los ayudantes de BBDD). Como ser ayudante tiene restricciones importantes de gastos se quedarán en un Airbnb 'La familia' administrado por Connie Corleone. Entre los panoramas que quieren agendar están una cata de vinos en el local 'Vino de Mesa Italiano', cenarán en el restaurant 'El príncipe di Corleone', visita a 'The God Father's House' en Via Candelora, 25, Corleone y para compensar visitarán también el centro del 'Movimiento Antimafia (CIDMA)'. El resto del tiempo lo pasarán en la playa en Palermo.

Por las mismas restricciones presupuestarias el viaje lo realizarán en un charter de 'AeroPeor' desde Santiago (UTC-4) a Palermo (UTC+2) el 1 de agosto a las 3 am y volviendo el 6 de agosto a la misma hora en un viaje de 20 horas que hace escala en Rio de Janeiro y Casablanca, el piloto es Paulie Gatto y los viajes internos estarán a cargo de la empresa de buses 'Viaja con respeto' y el conductor Luca Brasi.

Los participantes del viaje son, además de los nombrados, Martina Tattaglia, Tomás Barzini, Vincenzo Martino, Agustino Beckerini, Consuelo Inostrozini, Ignacio Garridelli, Olivia Llanini, Paula Contessa, Sofia Retamalini. Se entrega el archivo **viaje.xlsx** que incluye más información del viaje. Se deben generar las sentencias necesarias para insertar todos los datos mencionados anteriormente (los datos pueden estar en 'duro' *harcodear* en el archivo **.sql**

- a) Se debe crear los alojamientos, transportes, personas, participantes, aerolineas, etc. (1.5 pts)
- b) Generar una consulta que indique el costo total del viaje. (1.5 pts)

- c) Noticia de última hora, el piloto del charter renunció (ya no es empleado) y no será Paulie, sino que el mismo Luca Brasi. Escriba la sentencia SQL que refleje estos cambios. (1.5 pts)
- d) Sofía indica que hay un error en su nombre ya que está sin tilde en la i, aparece como Sofia y es Sofía. Escriba la sentencia que modifica este lamentable error. (0.7 pts)

2.3. Readme (2 pts)

- Se tiene que mencionar las restricciones de integridad identificadas en el enunciado y por qué son necesarias, estas restricciones incluyen llaves foráneas, datos no nulos, datos únicos y llaves primarias. (1.5 pts)
- Incluir pasos a seguir para la ejecución y corrección de la entrega. (0.5 pts)

3. Formato

Para el **formato de las tablas** en la base de datos de PostgreSQL, tienen que **asignar los nombres según los definidos en el diagrama E/R**. Para las **consultas** se tiene que llamar a los **atributos por los especificados**. El incumplimiento del formato indicado para archivos (etiquetado y comentarios) conllevará descuentos automáticos.

Todos los archivos deben ir dentro de la carpeta ./Sites/E2/

Esta tarea será mayormente corregida por un corrector automático, es decir que es muy importante seguir el formato de las carpetas, archivos y resultados. No agregar cosas adicionales como guiones bajos, espacios antes o después de las palabras.

4. IC y Reglas de negocio

1. De personas y sus datos

- Existe una entidad persona que registra los datos de todos los miembros del sistema de viajes (Empleados y Usuarios) y contiene los datos comunes a todos ellos (nombre, correo, contraseña, nombre de usuario, teléfono de contacto, run y dv).
- Las personas se identifican de forma única por su correo electrónico, el cual debe ser ingresado obligatoriamente al momento de registrarse y no puede repetirse en el sistema. Asimismo, su nombre de usuario también debe ser único. Además, cada persona debe tener registrados su RUN y dígito verificador.
- Una persona puede tener estar registrada en el sistema como empleado o usuario, pudiendo tener ambos tipos de registro al mismo tiempo, debido a que al ser

empleado y usuario tienen correos distintos, la misma persona puede tener dos cuentas diferenciándose por el mail.

- Un usuario es una persona registrada que puede realizar acciones como crear agendas, tomar reservas, dejar reseñas y contratar seguros. Cada usuario está asociado a una única persona del sistema y cuenta con un saldo de “puntos booked” que parte desde cero, que acumula al realizar reservas.
- Con cada reserva, un usuario suma “puntos booked” que puede canjearlos por dinero en reservas futuras.
- Un empleado es una persona registrada con funciones asignadas como conductor de transporte. Cada empleado está vinculado a una única persona, y debe estar correctamente identificado en el sistema.
- Los empleados pueden trabajar exclusivamente en jornadas diurnas, nocturnas.
- Los empleados siempre cuentan con un contrato full time o part time.
- La empresa les exige a todos los empleados reportar en qué seguro médico se encuentran inscritos, es decir, su isapre (Más vida, Colmena, Consalud, Banmédica, Fonasa).
- Cualquier dato que sea eliminado o modificado tiene que reflejar este cambio en toda la base de datos.

2. Agenda

- Cada agenda pertenece a un usuario y esta contiene una etiqueta como título al crear la agenda. La agenda puede contener múltiples reservas.

3. Reserva

- Al igual que en la etapa anterior, existe una entidad de reservas, donde una reserva se considera “No disponible” cuando está presente en una agenda de usuario, lo que significa que está reservada, por otro lado, una reserva se considera “Disponible” si no ha sido agregada por un usuario a su agenda, por lo que pueden existir reservas sin agenda. Por último, una reserva solo puede estar presente como máximo en una agenda.
- Las reservas deben contener una fecha que representa el momento que se realizó la reservación, estado disponibilidad según corresponda, el monto total pagado, el número de personas y los puntos “booked” obtenidos, que también deben ser mayores a cero.
- Cada reserva del usuario puede ser de transporte, panorama u hospedaje.
- Los usuarios pueden dejar una única reseña por reserva, donde deben indicar una calificación entre 1 y 5 estrellas, además de un comentario opcional.

- Un usuario puede asociar uno o más seguros a una reserva específica. Cada seguro está vinculado de forma exclusiva a una sola reserva. Los seguros tienen un valor asociado que representa el total que paga el usuario para asegurar a todas las personas de la reserva, un tipo dependiendo a qué servicio se busca cubrir (por ejemplo Seguro de Interrupción del Viaje, Seguro de Asistencia Médica, etc), una cláusula que detalla las condiciones del seguro, y una empresa que es la que provee el seguro.
- Ambos tipos de hospedajes tienen las siguientes características nombre de hospedaje, ubicación, precio por noche, estrellas, comodidades, fecha check in, fecha check out. Las reservas a hospedajes tienen un nombre según el hospedaje, ubicación, precio por noche, calificación en estrellas entre 1 y 5, una lista de comodidades que ofrece el hospedaje, y fechas de check in (fecha llegada) y check out (fecha abandono). Cada hospedaje está asociado exclusivamente a una reserva y deben incluir su ubicación, el precio por noche, calificación en estrellas y fecha de check-in.
- Además, hay dos tipos de hospedajes, los hoteles y los airbnb. Un hotel tiene una lista de políticas sobre la normativa del hotel, y un airbnb tiene el nombre y número de contacto del anfitrión, descripción del airbnb, cantidad de piezas, camas y baños.
- A diferencia del airbnb, el hotel cuenta con un registro de las habitaciones, donde cada habitación tiene un número único dentro del hotel y tipo según el tamaño de la misma (Sencilla, Doble, Matrimonial, Triple, Cuádruple, Suite). La habitación solo pertenece a un hotel.
- Los panoramas tienen un nombre que representa la actividad, empresa que realiza el panorama, descripción de la actividad, ubicación de donde se realiza, duración en horas, precio por persona que participa, capacidad máxima de participantes, una lista de restricciones para participar y fecha del panorama.
- Cada panorama se reserva por un usuario y este puede añadir participantes indicando su nombre y edad.
- Cada transporte está relacionado con una única reserva y debe contar con un empleado conductor asignado. Las características asociadas a un transporte son lugar de origen, lugar de llegada, capacidad, tiempo estimado, precio por asiento, empresa, fecha de salida y fecha de llegada.. Debe registrarse su tiempo estimado de viaje, precio por asiento (mayor a cero), y la fecha de salida.
- Existen tres tipos de transportes. El tren, el cual tiene una lista de comodidades y una de paradas. El bus, que tiene una lista de comodidades y un tipo que varía según el bus (cama, normal y semi-cama). Por último, el avión que tiene su clase dependiendo del tipo de avión y una lista de escalas.
- Las calificaciones (estrellas) para hospedajes y reviews son discretas.

- Los montos/precios para las reservas y seguros deben ser positivos.

5. Flujo de la tarea

1. **Conexión a la base de datos:** Conéctate al servidor del curso (bdd1.ing.puc.cl) utilizando tu usuario UC como nombre de la base de datos, mediante PostgreSQL.
2. **Análisis de estructura:**
 - Estudia cuidadosamente el diagrama Entidad-Relación (E/R) proporcionado en el enunciado.
 - Revisa los archivos CSV disponibles en la carpeta `./Sites/E2/csv/`.
 - Determina cómo se llevarán los datos de los CSV a las tablas de la base de datos.
3. **Creación de tablas:**
 - Escribe el archivo `crear_tablas.sql`, el cual al ser ejecutado debe crear todas las tablas necesarias en la base de datos.
 - Asegúrate de incluir todas las restricciones de integridad al crear cada tabla: llaves primarias, foráneas, unicidad, no nulos y restricciones de negocio.
4. **Carga de datos:**
 - Escribe el archivo `poblar_tablas.sql` para cargar los datos desde los CSV a la base de datos.
 - Usa tablas temporales para facilitar la validación previa a la inserción definitiva.
 - Los datos que no cumplan con las restricciones de integridad deben ser registrados en los archivos `nombreTabla_descartados.csv` dentro de la carpeta `descartados/`.
5. **Verificación de integridad:**
 - Asegúrate de que los datos insertados respetan todas las reglas de integridad.
 - Ejecuta pruebas para validar relaciones entre tablas y restricciones específicas del modelo de negocio.
6. **Consultas SQL:**
 - Desarrolla las 12 consultas solicitadas, etiquetadas estrictamente como `1a.sql`, `1b.sql`, `2a.sql`, etc.
 - Guarda cada una en el directorio `./Sites/E2/consultas/`.
 - Incluye una descripción como comentario al inicio de cada archivo.
 - Verifica que cada consulta retorne resultados correctos y en el formato esperado.

7. Creación del README:

- Incluye las restricciones de integridad que implementaste y una breve justificación.
- Detalla los pasos necesarios para ejecutar los scripts correctamente.

8. Verificación final:

- Asegúrate de que todos los archivos se encuentren en las ubicaciones correctas según la estructura solicitada.
- Ejecuta los scripts mediante `psql -f archivo.sql` para validar su funcionamiento automático.

9. Entrega:

- Sube todos los archivos al directorio `Sites/E2/` de tu cuenta personal en el servidor del curso.
- Verifica que no haya archivos fuera de lugar o con nombres incorrectos.

6. Recomendaciones

- Asegura la correcta ejecución de los scripts¹ SQL en PostgreSQL.
- Verifica la precisión y eficiencia de las consultas SQL.
- Mantén claridad y estructura en tus entregables.

¹Utilizar comando `psql -f archivo.sql`.

7. Estructura de los archivos y formatos

La estructura de carpetas debe seguir el siguiente formato:

Sites

```

|-- E2
|   |-- consultas
|       |-- 1a.sql
|       |-- 2a.sql
|       |-- 2b.sql
|       |-- 3a.sql
|       |-- 3b.sql
|       |-- 4a.sql
|       |-- 5a.sql
|       |-- 6a.sql
|       |-- 7a.sql
|       |-- 7b.sql
|       |-- 7c.sql
|       |-- 7d.sql
|
|   |-- descartados
|       |-- personas_descartados.csv
|       |-- usuarios_descartados.csv
|       |-- empleados_descartados.csv
|       |-- agendas_descartados.csv
|       |-- reservas_descartados.csv
|       |-- reviews_descartados.csv
|       |-- seguros_descartados.csv
|       |-- transportes_descartados.csv
|       |-- buses_descartados.csv
|       |-- trenes_descartados.csv
|       |-- aviones_descartados.csv
|       |-- panoramas_descartados.csv
|       |-- participantes_descartados.csv
|       |-- hospedajes_descartados.csv
|       |-- hoteles_descartados.csv
|       |-- habitaciones_descartados.csv
|       |-- airbnb_descartados.csv
|
|   |-- csv
|       |-- agenda_reserva.csv
|       |-- habitaciones.csv
|       |-- participantes.csv

```

```
| | |-- personas.csv
| | |-- review_seguro.csv
| |
| |-- archivos
| | |-- crear_tablas.sql
| | |-- poblar_tablas.sql
| | |-- README.md
```

Es importante que sigan esta estructura y el nombre de los atributos indicados en cada consulta ya que la corrección la requiere. **Si no se sigue el formato, se descontará puntaje .**