

MEMORIA TRABAJO BASES DE DATOS

CURSO 2023/2024



FACULTAD DE CIENCIAS MATEMÁTICAS

Autores: Noelia Cárdenas Salamanca, Sergio Castell Mateo, Marcos Escribano
Sánchez y Ana Lara Villalba

Madrid, noviembre de 2023

Índice

1. Introducción. Enunciado del proyecto	2
2. AS-IS del estado de la aplicación actual. Modelo E/R preliminar	3
2.1. Entidades y atributos	3
2.2. Asociaciones y cardinalidades. Jerarquías	3
3. Entrevista	4
4. Modelo TO-BE objetivo. Modelo E/R definitivo	5
4.1. Entidades y atributos	5
4.2. Asociaciones y cardinalidades. Jerarquías	5
5. Diseño lógico de la base de datos. Transformaciones y normalización	7
5.1. Fase de diseño lógico estándar: Transformaciones	7
5.2. Fase de diseño lógico específico: Normalización	9
5.2.1. Primera forma normal (1FN)	9
5.2.2. Segunda forma normal (2FN)	9
5.2.3. Tercera forma normal (3FN)	10
5.2.4. Modelo relacional normalizado	10
6. Diseño Físico	12
6.1. Consideraciones previas a la migración de datos	12
6.2. Migración de datos	12
6.3. Vistas. Consultas	13
7. Resumen de las decisiones tomadas	14

1. Introducción. Enunciado del proyecto

Se nos presenta la siguiente situación para el proyecto:

La agencia TRAVELIN se dedica a la reserva de viajes, alojamientos y excursiones a lo largo de toda España. Fue fundada hace 3 años, y para gestión encargó a una consultora TI la realización de un sistema de información.

La empresa dispone de una aplicación web donde los clientes reservan alojamiento y excursiones, así como una gran cantidad de sucursales repartidas por el territorio nacional.

Debido a su gran momento de expansión el director general ha detectado que el sistema se quedó obsoleto, y necesitan que alguien remodele su plataforma de reserva. En una primera fase, el consultor de nuestro equipo de desarrolladores ha tenido varias reuniones con el área de negocio y le han trasladado las siguientes premisas que debería cumplir el nuevo aplicativo:

1. *El principal objetivo es maximizar el beneficio. Para ello se propone:*
 - a. *Optimizar las rutas correspondientes a las excursiones reservadas.*
 - b. *Abrir nuevos canales (venta, atención usuarios, RRSS, difusión, etc.).*
 - c. *Sacar beneficio a los datos mediante previsiones para adaptar la demanda o facilitar el reporte.*
2. *Solucionar algunos de los problemas detectados:*
 - a. *Han detectado que hay muchos hacker que robotizan la aplicación.*
 - b. *Tienen una gran cantidad de registros falsos.*
 - c. *Existen clientes duplicados en la BD.*
 - d. *Hay una tasa elevada de clientes que abandonan debido al tiempo de respuesta de la web.*
 - e. *Demasiados clientes anulan rutas, y no se puede reorganizar a tiempo suponiendo coste asociado por autocares vacíos.*
 - f. *Las sucursales se quejan del tiempo elevado de consulta de los datos de los clientes.*
 - g. *La Alta Dirección no dispone de reporte.*
 - h. *Ha habido filtrado de datos, que está poniendo en riesgo la GDPR.*

También nos ha facilitado el esquema actual de la BD de la que dispone la aplicación, con datos de ejemplo para que podamos analizar el enfoque.

Después de analizar el contexto, el jefe de proyecto de tu equipo propone realizar las siguientes etapas:

- 1°. *Realizar un assesment AS-IS para identificar el estado actual de la aplicación y la BD mediante técnicas de reingeniería que permitan pasar del modelo físico al modelo E/R.*
- 2°. *Agendar una cita con el Director General para que el analista-programador de vuestro equipo resuelva todas las cuestiones que hayan surgido al resto de miembros del equipo.*
- 3°. *Definir un modelo TO-BE objetivo que resuelva los principales problemas de negocio y cumpla las expectativas de los objetivos (modelo E/R, etc.).*
- 4°. *Realizar el diseño lógico estándar de la BD (transformaciones).*
- 5°. *Realizar el diseño lógico específico (normalización).*
- 6°. *Realizar migración de los datos al nuevo esquema.*
- 7°. *Construir vistas que encapsulen el acceso a los datos, como por ejemplo accesos específicos para la web y los usuarios de sucursal, acceso a la información de usuario, etc.*
- 8°. *Construir consultas avanzadas que permitan tener un reporte a dirección general, rutas más rentables, perfil del usuario de cada ruta, optimizar rutas, etc.).*
- 9°. *Securizar la BD.*
- 10°. *Optimizar la BD (diseño físico).*

2. AS-IS del estado de la aplicación actual. Modelo E/R preliminar

Exclusivamente con la base de datos provista y la información que de las tablas podíamos inferir, creamos el modelo E/R. A continuación vamos a explicar el modelo propuesto, junto con algunas de las decisiones de diseño que hemos tomado, teniendo en cuenta que nos hemos ceñido a la información dada por las tablas.

2.1. Entidades y atributos

Cada tabla tiene su entidad y hemos mantenido todos y cada uno de los atributos asociados al tratarse de un primer acercamiento al modelo. Algunas consideraciones de interés sobre atributos son que 'Teléfono' aparece como multivaluado porque, como se ve en la tabla, puede darse que una sucursal tenga más de un número de teléfono; y que, más allá de las claves típicas de ID, entidades cruciales y con información sensible como 'USUARIOS.WEB' cuentan con más de una.

2.2. Asociaciones y cardinalidades. Jerarquías

Explicamos brevemente las asociaciones y sus respectivas cardinalidades, a la par que las jerarquías y el tipo que son:

1. 'ACUDE' hace referencia a los clientes que van a la sucursal física. Como puede ser que no vaya nadie o vaya mucha gente a una sucursal, tenemos la cardinalidad 0:N por esa parte. Y como un cliente mínimo ha tenido que ir a una sucursal (por definición de cliente) y puede haber ido a muchas, entonces la cardinalidad es 1:N.
2. 'REALIZA' es relación débil en existencia, pues la reserva debe ser hecha por el usuario web o cliente para decir que existe. La reserva se relaciona unívocamente con el usuario o cliente (1:1), y el usuario o cliente puede hacer ninguna o muchas reservas (0:N).
3. 'RESERVA' es la entidad padre de una jerarquía con las entidades ALOJAMIENTO y RUTA, pues por el contexto intuimos que la agencia ofrece estos dos servicios a reservar, a gusto del consumidor si desea uno, el otro, o ambos. Es total por lo recién expuesto, y disjunta porque es imposible una ruta que a la vez haga de alojamiento.
4. 'CUESTA' relaciona el alojamiento y la ruta con sus respectivos precios. Un alojamiento debe tener mínimo un precio establecido y, como se ve en la tabla correspondiente, los precios de un mismo alojamiento varían con las fechas, de modo que la cardinalidad es 1:N; y un precio (como valor numérico) puede tenerlo ningún o varios alojamientos, de modo que la cardinalidad es 0:N. Las cardinalidades para las rutas son iguales por los mismos razonamientos.
5. 'SE EFECTÚA' relaciona las rutas con los autocares. Un autocar puede no hacer ninguna ruta (está estacionado por el motivo que sea) o puede hacer varias (0:N), mientras que una ruta la debe hacer mínimo un autocar (carece de sentido dejarla desierta, más aún pudiéndose reservar) o pueden seguirla varios (1:N).
6. 'HACE' es relación débil en existencia, puesto que las paradas de las rutas, más allá de su localización física, requieren del autocar que en ellas se detenga para que propiamente sean paradas de las rutas. El autocar en ruta para en al menos una parada (el destino), lo normal es que lo haga en varias (1:N), de modo que cada parada recibe un autocar durante la ruta (1:1).
7. 'REALIZADA' relaciona la entidad RUTA con la entidad RUTA_CLIENTE. Dentro del equipo hemos visto que relacionar esta última entidad no era algo inmediato, y podía verse desde distintos ángulos, de modo que esta es la visión que hemos tomado: una ruta X puede entrar en la ruta planeada de ningún o muchos clientes (0:N), y la ruta concreta de un cliente puede coincidir con varias predeterminadas, pero mínimo con una (1:N).

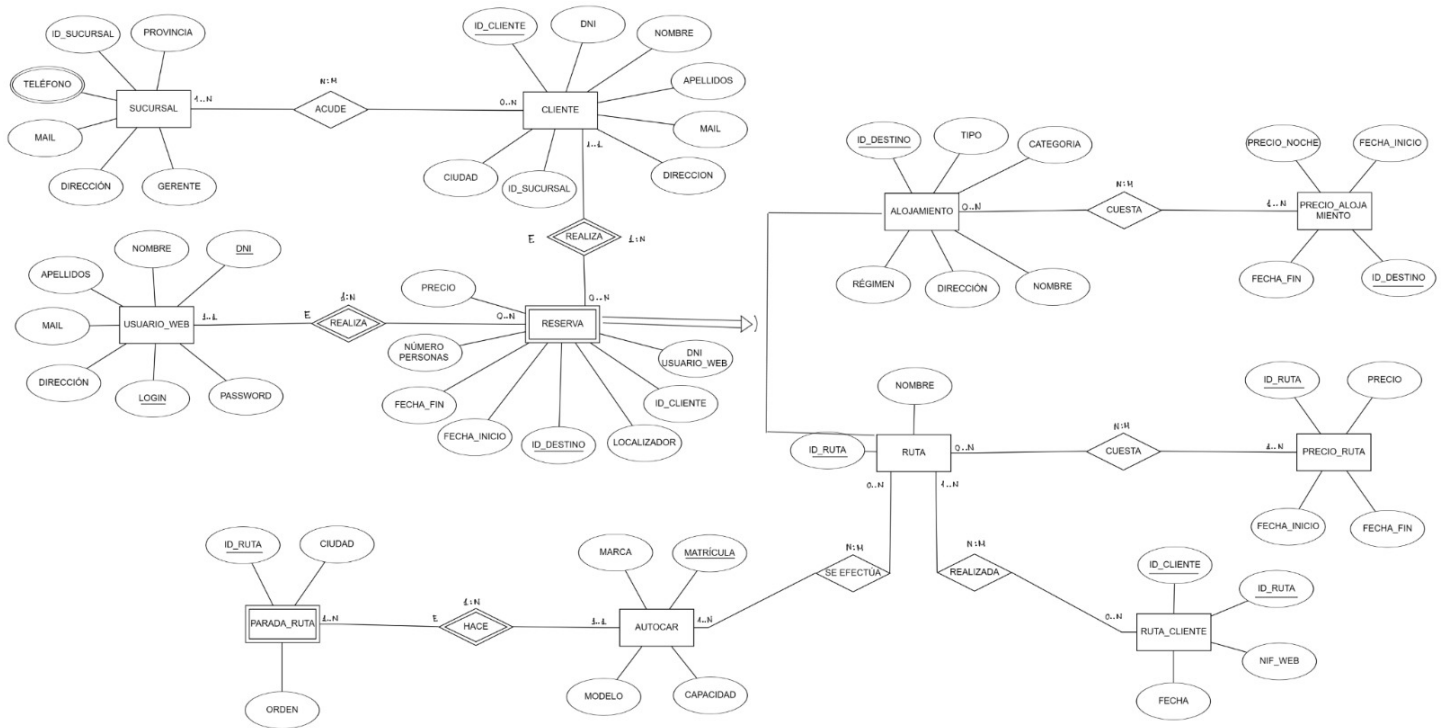


Figura 1: Modelo E/R preliminar

3. Entrevista

Para entender mejor el contexto de negocio y poder así ajustar adecuadamente nuestra propuesta de mejora, nuestro equipo agendó una cita con el Director General para entender mejor los intereses tras esta mejora y resolver dudas para encaminar un modelo TO-BE objetivo.

La analista-programadora nos trasladó las siguientes conclusiones:

- Las sucursales sin dirección indicada, ocurre que las que no tienen gerente están en proceso de cierre, y las que lo tienen son nuevas y están por abrirse.
- Es de gran interés que tanto el cliente que accede por la web como el que acude físicamente a una sucursal, tengan un perfil web.
- Las contraseñas de los usuarios deben ser seguras, además de tener un doble factor de verificación: teléfono y correo.
- En pos de maximizar el rendimiento, los precios, como se aprecia en los datos provistos, varían: suben en verano y bajan en invierno, atendiendo así a las variaciones de la demanda.

Como cierre, nos comentó que el Director incidió en que se optimizasen las rutas (ergo, el rendimiento) y que la seguridad aumentase.

4. Modelo TO-BE objetivo. Modelo E/R definitivo

Lo concluido de la reunión con el Director, junto con varias decisiones de diseño que a continuación se explicarán, dieron lugar a que estableciésemos el siguiente modelo E/R definitivo, nuestro modelo TO-BE. Pasamos a comentar las decisiones de diseño que hemos tomado, tomando como referencia el modelo de la sección anterior y exponiendo los cambios realizados.

4.1. Entidades y atributos

1. *(La aparición de las entidades FÍSICO y ONLINE, además de la variación en los atributos, se detallan en el subapartado siguiente, al explicar la aparición de la nueva jerarquía. Esto mismo le ocurre a RESERVA con ALOJAMIENTO y RUTA.)*
2. En vista de ofrecer un mejor servicio para el cliente y, en consecuencia, maximizar el beneficio, planteamos la creación de varias entidades:
 - 'ATENCIÓN_CLIENTE' ofrece un teléfono y correo para atender las dudas y consultas de los clientes, mejorando así su satisfacción con la agencia. El atributo Teléfono, no multivaluado porque consideramos que un solo teléfono para la atención al cliente es lo justo y necesario, es clave.
 - 'REDES_SOCIALES' presenta un canal moderno para publicitar la agencia. Esta tendrá presencia con un usuario en diversas redes sociales. El atributo Username, en referencia al nombre de usuario de la agencia en la red social correspondiente, es clave.
 - 'CAMPAÑA' es la principal apuesta para la mejora de los beneficios. Planteamos un sistema de cupones de uso limitado (por fechas concretas) que rebajen el precio de rutas y/o alojamientos. En particular, estos descuentos se darían en fechas señaladas: Semana Santa, San Valentín, Navidades... El atributo Cupón, al tomarlo como un código único, es clave.
3. La entidad RUTA_CLIENTE, viendo que el usuario online realiza una reserva, que no es más que una ruta y/o alojamiento, por lo cual la entidad ONLINE y RUTA están asociadas, consideramos que solo iba a aportar información redundante (viendo que sus atributos se repiten con respecto a los de las dos entidades recién mencionadas) y prescindimos totalmente de ella. Por un razonamiento parecido, la asociación directa entre ONLINE y RESERVA hizo que prescindieramos totalmente de ID_CLIENTE y DNI_USUARIO_WEB se transformó en solo la clave de la entidad CLIENTE, DNI.
4. Han variado las claves en distintos puntos del modelo:
 - Localizador pasa a ser clave, pues identifica unívocamente las reservas. Las reservas de alojamientos tienen como clave principal (Id_destino, Localizador), pues tenemos así la información de qué cliente se hospedará en qué destino, y análogamente le ocurre a las reservas de rutas: (Id_ruta, Localizador). Nótese que esta decisión de diseño refleja que ambos servicios son independientes: aunque pueda resultar que lo más razonable es contratar una ruta y alojamiento en el punto deseado, es perfectamente posible que un cliente solo quiera el alojamiento y se encargue del transporte, por ejemplo.
 - Fecha_inicio, junto con Id_destino o Id_ruta, según corresponda, forman la clave principal en las respectivas relaciones de precios. Esto se debe a que solo las ID no pueden ser clave, y el añadido de la fecha indica además cuándo cambia el precio. Orden pasa a ser clave, formando así la clave principal (Id_ruta, Orden), pues sin Orden como clave, la Id_ruta no puede ser clave principal por sí sola (lógico, para las varias paradas se repite su valor en las respectivas ocurrencias).

4.2. Asociaciones y cardinalidades. Jerarquías

1. El comentario del Director sobre que interesa que todos los clientes tengan presencia web nos llevó a generar una siguiente jerarquía con CLIENTE. Primeramente, notamos que los clientes físicos y los usuarios web compartían bastantes atributos, de modo que nos pareció una decisión natural de por sí (incluimos además el atributo multivaluado Teléfono a petición del Director, como se indicó en el apartado de la Entrevista). La jerarquía es total (los clientes solo pueden ser físicos u online) y solapada, que precisamente es nuestro interés, pues queremos que los físicos sean también online. Esto último lo conseguimos por medio de la disposición que tienen FÍSICO y ONLINE y sus asociaciones: el cliente físico acude a sucursal, y la sucursal registra a online (es decir, los empleados de la sucursal toman los datos de los clientes y les crean un perfil online, a través del cual realizan las reservas de los clientes en el momento). Por tanto, queda reflejado en el modelo nuestra intención de que todo cliente

tenga presencia online y el resto de asociaciones se harán con la entidad ONLINE directamente.

La asociación ACUDE es idéntica a la del modelo inicial, y la asociación REGISTRA relaciona SUCURSAL y ONLINE, de manera que una sucursal puede registrar en la web a ningún o varios clientes (0:N), mientras que el cliente nuevo que acude a una sucursal recibe allí sus datos web para siempre (1:1).

- Las nuevas entidades creadas en este modelo TO-BE (ATENCIÓN_AL_CLIENTE, REDES_SOCIALES y CAMPAÑA), todas se relacionan con la entidad ONLINE (o sea, usuarios web, que llamaremos directamente clientes):

- ATENCIÓN_AL_CLIENTE se relaciona con ONLINE mediante CONTACTA. El servicio de atención al cliente puede atender a ningún o muchos clientes (0:N), y un cliente puede no requerir de dicha atención, o acudir a ella en múltiples ocasiones (0:N).
- REDES_SOCIALES se relaciona con ONLINE mediante UTILIZA. Las redes sociales pueden ser visitadas por ningún cliente o por muchos (0:N), y dichos clientes pueden no usar redes sociales, o usar varias (0:N).
- CAMPAÑA se relaciona con ONLINE mediante OPTA. En una campaña de promoción puede no participar nadie, o participar muchos clientes (0:N), y un mismo cliente puede no aprovechar ninguna campaña, o beneficiarse de varias (0:N).

Así pues, tenemos un modelo E/R sobre el que cimentar nuestra nueva propuesta de base de datos.

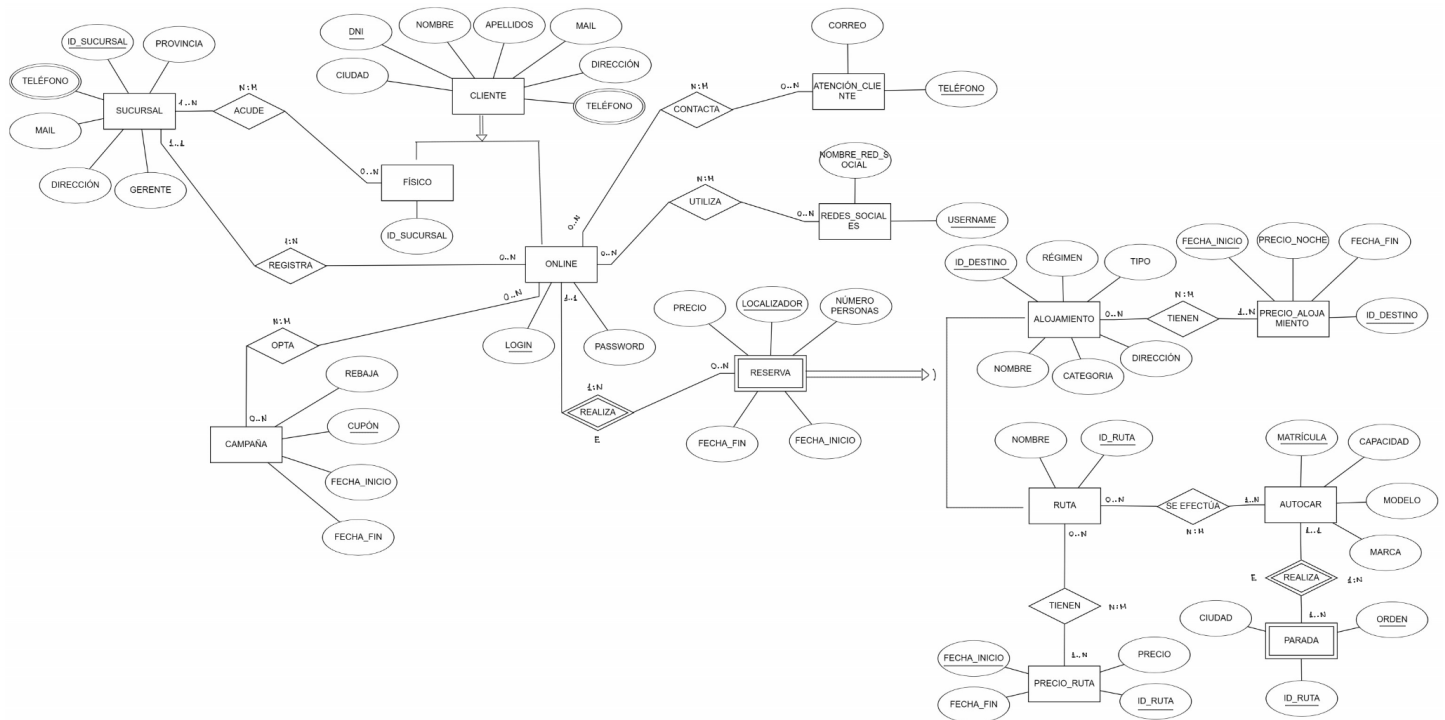


Figura 2: Modelo E/R definitivo

5. Diseño lógico de la base de datos. Transformaciones y normalización

5.1. Fase de diseño lógico estándar: Transformaciones

Por medio de las reglas de transformación, obtendremos el modelo relacional que surge del modelo E/R que acabamos de plantear. Teniendo en cuenta alguna decisión de diseño expuesta más adelante, sería así:

Físico: (DNI, Nombre, Apellidos, Mail, Dirección, Ciudad, Id_sucursal)

Online: (DNI, Nombre, Apellidos, Mail, Dirección, Ciudad, Login, Password, Id_sucursal)

Teléfono: (DNI, Teléfono)

Sucursal: (Id_sucursal, Mail, Dirección, Provincia, Gerente)

Teléfono: (Id_sucursal, Teléfono)

Acude: (DNI, Id_sucursal)

Atención_cliente: (Correo, Teléfono)

Contacta: (DNI, Login, Teléfono)

Redes_sociales: (Nombre_red_social, Username)

Utiliza: (DNI, Login, Username)

Campaña: (Cupón, Rebaja, Fecha_inicio, Fecha_fin)

Opta: (DNI, Login, Cupón)

Reserva: (Localizador, Precio, Número_personas, Fecha_inicio, Fecha_fin, DNI, Login)

Alojamiento: (Localizador, Id_destino, Tipo, Categoría, Nombre, Dirección, Régimen)

Ruta: (Localizador, Id_ruta, Nombre)

Precio_alojamiento: (Id_destino, Fecha_inicio, Fecha_fin, Precio_noche)

Tienen: (Localizador, Id_destino, Fecha_inicio)

Precio_ruta: (Id_ruta, Fecha_inicio, Fecha_fin, Precio)

Tienen: (Localizador, Id_ruta, Fecha_inicio)

Autocar: (Matrícula, Capacidad, Modelo, Marca)

Se_efectúa: (Localizador, Id_ruta, Matrícula)

Parada: (Id_ruta, Orden, Ciudad, Matrícula)

Las decisiones de diseño se ciñen a las dos jerarquías, a la relación FÍSICO y a una cuestión de notación:

- Para la jerarquía con CLIENTE como padre, se decidió optar por la vía en la que las entidades hijas sean las que se transformen en relaciones y no CLIENTE, teniendo que heredar ambas todos los atributos de CLIENTE. La transformación que deja a las 3 como relaciones iba a suponer una tabla más, y con ello información duplicada; y dejar a CLIENTE como única relación haría perder esa filosofía de que todos los clientes son online en definitiva, además de que habría muchos atributos y bastantes null (pensando en Id_sucursal por aquellos clientes que hayan accedido a la web desde sus casas). Puede parecer que estas dos relaciones tienen también muchos atributos, y encima la mayoría repetidos, pero podemos prescindir totalmente de la relación FÍSICO, decisión justificada en el siguiente punto.
- Para la jerarquía con RESERVA como padre, se decidió que esta y sus entidades hijas fuesen relaciones, principalmente porque RUTA y APARTAMENTO son bastante ajenas como para hacer que el padre sea él solo relación, y que las hijas fueran relaciones solas implica tener muchos atributos, y que la mayoría estén repetidos.
- A la vista de que la asociación REALIZA, debido a la transformación, se reduce a añadir el atributo Id_sucursal a la relación ONLINE. Nuestro equipo se percató de que ese atributo

puede identificar a los clientes que hayan ido a sucursal (aparecerá la ID correspondiente) de los que hayan accedido a la web por cuenta propia (aparecerá null). Por tanto, la relación FÍSICO pasa a ser enteramente redundante y decidimos prescindir de ella, lo cual va en pos de la filosofía de que todos los clientes deben tener presencia online.

Físico: (DNI, Nombre, Apellido, Mail, Dirección, Ciudad, Id_sucursal)

- Antes de dar por finalizado el punto, nos pareció apropiado renombrar ciertas relaciones a fin de que en el resultado final (diseño físico) se distingan mejor entre sí:
 - Al desaparecer FÍSICO, la relación ONLINE pasa a llamarse CLIENTE de manera natural.
 - La primera relación TELÉFONO pasa a ser TELÉFONO_CLIENTE.
 - La segunda relación TELÉFONO pasa a ser TELÉFONO_SUCURSAL.
 - La relación ACUDE pasa a ser VISITA_SUCURSAL_CLIENTE.
 - La relación CONTACTA pasa a ser CLIENTE_ATENDIDO.
 - La relación UTILIZA pasa a ser CLIENTE_VISITA_RRSS.
 - La relación OPTA pasa a ser CLIENTE_USA_CUPÓN.
 - La primera relación TIENEN pasa a ser TEMPORADA_ALOJAMIENTO.
 - La primera relación TIENEN pasa a ser TEMPORADA_RUTA.
 - La relación SE_EFECTÚA pasa a ser PASAJEROS_AUTOCAR.

Tenemos así el modelo relacional final:

Cliente: (DNI, Nombre, Apellidos, Mail, Dirección, Ciudad, Login, Password, Id_sucursal)

Teléfono_cliente: (DNI, Teléfono)

Sucursal: (Id_sucursal, Mail, Dirección, Provincia, Gerente)

Teléfono_sucursal: (Id_sucursal, Teléfono)

Visita_sucursal_cliente: (DNI, Id_sucursal)

Atención_cliente: (Correo, Teléfono)

Cliente_atendido: (DNI, Login, Teléfono)

Redes_sociales: (Nombre_red_social, Username)

Cliente_visita_RRSS: (DNI, Login, Username)

Campaña: (Cupón, Rebaja, Fecha_inicio, Fecha_fin)

Cliente_usa_cupón: (DNI, Login, Cupón)

Reserva: (Localizador, Precio, Número_personas, Fecha_inicio, Fecha_fin, DNI, Login)

Alojamiento: (Localizador, Id_destino, Tipo, Categoría, Nombre, Dirección, Régimen)

Ruta: (Localizador, Id_ruta, Nombre)

Precio_alojamiento: (Id_destino, Fecha_inicio, Fecha_fin, Precio_noche)

Temporada_alojamiento: (Localizador, Id_destino, Fecha_inicio)

Precio_ruta: (Id_ruta, Fecha_inicio, Fecha_fin, Precio)

Temporada_ruta: (Localizador, Id_ruta, Fecha_inicio)

Autocar: (Matrícula, Capacidad, Modelo, Marca)

Pasajeros_autocar: (Localizador, Id_ruta, Matrícula)

Parada: (Id_ruta, Orden, Ciudad, Matrícula)

5.2. Fase de diseño lógico específico: Normalización

Con el objetivo de eliminar las posibles anomalías que aparezcan en las distintas tablas, aplicaremos a continuación la teoría de la normalización, la cual detecta y corrige tales errores. Nuestra finalidad será llegar a la Tercera Forma Normal (3FN). A partir del modelo relacional anterior, estudiaremos si tales tablas están o no en dicha forma.

5.2.1. Primera forma normal (1FN)

Como todos los atributos de nuestro modelo son simples, todas las relaciones (tablas) están en 1FN.

5.2.2. Segunda forma normal (2FN)

Una tabla está en 2FN si: está en 1FN y se da una de las siguientes posibilidades:

- (I) La clave principal es simple.
- (II) No hay atributos no principales.
- (III) Cada atributo no principal depende completamente de la clave.

Procedamos a analizar nuestras tablas:

- Siguiendo (I), las relaciones SUCURSAL, ATENCIÓN_CLIENTE, REDES_SOCIALES, CAMPAÑA, RESERVA, AUTOCAR están en 2FN.
- A partir de (II), las relaciones TELÉFONO_CLIENTE, TELÉFONO_SUCURSAL, VISITA_SUCURSAL_CLIENTE, CLIENTE_ATENDIDO, CLIENTE_VISITA_RRSS, CLIENTE_USA_CUPÓN, TEMPORADA_ALOJAMIENTO, TEMPORADA_RUTA y PASAJEROS_AUTOCAR están en 2FN.
- Por (III), en las tablas CLIENTE, ALOJAMIENTO y RUTA ninguna de la claves forman dependencia completa, de hecho, todas las claves candidatas implican todos los atributos. Por tanto las nuevas tablas 2FN son la original y una nueva tabla formada solo por las claves.
Por último, revisamos PRECIO_ALOJAMIENTO y PRECIO_RUTA que sí generan nuevas tablas. Como Id_destino, Fecha_inicio no implica Precio, hay dependencia completa con el atributo Precio. Sin embargo, Id_destino no implica Fecha_fin aunque Fecha_inicio sí implica Fecha_fin. Por tanto, no hay dependencia completa en este caso y tendremos que generar una nueva tabla, con nombre Temporadas_precios y que consta de los atributos (Fecha_inicio, Fecha_fin) y las tablas originales quedarían igual pero sin el atributo Fecha_fin.
- Un caso especial es el de la relación PARADA, ya que hemos tenido que tomar decisiones de diseño para comprobar si está o no en 2FN. En este caso, hemos decidido que para un Id_ruta concreto pueden aparecer más de una matrícula, es decir, pese a que asociaremos los autobuses con más plazas a las rutas más concurridas, puede que la demanda de una ruta sea tan grande que se necesite más de un autobús para satisfacerla y de ahí que para un id puedan aparecer más de una matrícula. Teniendo esto en cuenta, Id_ruta y Orden no implican Ciudad ni Matrícula, por lo que habrá dependencia completa y, por ello, la tabla está en 2FN. En este punto procederíamos igual que para las relaciones del punto (II).

5.2.3. Tercera forma normal (3FN)

Una tabla está en 3FN si: está en 2FN y se da una de las siguientes posibilidades:

- (I) No hay más de un atributo no principal.
- (II) Cada atributo no principal no depende transitivamente de la clave.

Procedamos a analizar nuestras tablas:

- Teniendo en cuenta la primera posibilidad, las relaciones TELÉFONO_CLIENTE, TELÉFONO_SUCURSAL, VISITA_SUCURSAL_CLIENTE, ATENCIÓN_CLIENTE, CLIENTE_ATENDIDO, REDES_SOCIALES, CLIENTE_VISITA_REDES_SOCIALES, CLIENTE_USA_CUPÓN, RUTA, PRECIO_ALOJAMIENTO, PRECIO_RUTA, TEMPORADA_ALOJAMIENTO, TEMPORADA_RUTA, TEMPORADAS_PRECIOS y PASAJEROS_AUTOCAR están en 3FN.
- Para las tablas restantes, tenemos que analizar la existencia o no de dependencia transitiva con la clave. En SUCURSAL, vemos que los atributos no principales Mail, Dirección, Provincia y Gerente no implican Id_sucursal. Como además no hay dependencia funcional entre ninguno de los anteriores atributos, llegamos a la conclusión de que no hay dependencia transitiva de ninguno de los atributos con la clave, por lo que la tabla está en 3FN. Para la relación CAMPAÑA, como Fecha_inicio y Fecha_fin tienen dependencia con la clave y, además, el otro atributo no principal de la relación, Rebaja, no tiene implicaciones con ningún otro atributo ni con la clave, entonces no hay dependencia transitiva de ningún atributo no principal con la clave y, por tanto, la tabla está en 3FN. Para la relación PARADA, como los atributos no principales Precio, Número_personas, Fecha_inicio, Fecha_fin, DNI y Login no tienen dependencias entre ellos ni con la clave Localizador, entonces la tabla está en 3FN. Por último, la relación RESERVA ya está en 3FN porque no hay dependencias entre ninguno de los atributos no principales ni tampoco entre los atributos no principales con la clave Localizador. La única relación que no está en 3FN es AUTOCAR, ya que el atributo no principal Modelo depende transitivamente de Matrícula a través de Capacidad y Marca (hay dependencia funcional de Modelo con Capacidad y Marca y no la hay con la clave Matrícula). Para que la relación esté en 3FN tenemos que generar las nuevas tablas Capacidad_Modelo de la forma (Capacidad, Modelo) y Marca_Modelo, formada por (Marca, Modelo). El resto de atributos no principales no tienen dependencia entre ellos ni con la clave, por lo que no hay que hacer más tablas y con esta modificación, AUTOCAR está en 3FN.

5.2.4. Modelo relacional normalizado

Una vez acabado el proceso de normalización del modelo relacional, lo presentamos explícitamente, pues será el modelo sobre el cual partamos en la fase de diseño físico.

Cliente: (DNI, Nombre, Apellidos, Mail, Dirección, Ciudad, Login, Password, Id_sucursal)

Datos_Cliente: (DNI, Login)

Teléfono_cliente: (DNI, Teléfono)

Sucursal: (Id_sucursal, Mail, Dirección, Provincia, Gerente)

Teléfono_sucursal: (Id_sucursal, Teléfono)

Visita_sucursal_cliente: (DNI, Id_sucursal)

Atención_cliente: (Correo, Teléfono)

Cliente_atendido: (DNI, Login, Teléfono)
Redes_sociales: (Nombre_red_social, Username)
Cliente_visita_RRSS: (DNI, Login, Username)
Campana: (Cupón, Rebaja, Fecha_inicio, Fecha_fin)
Cliente_usa_cupón: (DNI, Login, Cupón)
Reserva: (Localizador, Precio, Número_personas, Fecha_inicio, Fecha_fin, DNI, Login)
Alojamiento: (Localizador, Id_destino, Tipo, Categoría, Nombre, Dirección, Régimen)
Datos_alojamiento: (Localizador, Id_destino)
Ruta: (Localizador, Id_ruta, Nombre)
Datos_ruta: (Localizador, Id_ruta)
Precio_alojamiento: (Id_destino, Fecha_inicio, Precio_noche)
Temporada_alojamiento: (Localizador, Id_destino, Fecha_inicio)
Precio_ruta: (Id_ruta, Fecha_inicio, Precio)
Temporada_ruta: (Localizador, Id_ruta, Fecha_inicio)
Temporadas_precios: (Fecha_inicio, Fecha_fin)
Autocar: (Matrícula)
Capacidad_modelo: (Capacidad, Modelo)
Marca_modelo: (Marca, Modelo)
Pasajeros_autocar: (Localizador, Id_ruta, Matrícula)
Parada: (Id_ruta, Orden, Ciudad, Matrícula)

6. Diseño Físico

6.1. Consideraciones previas a la migración de datos

Una vez llegado a este punto del proyecto, nuestro equipo ha tomado una serie de consideraciones para llevar a cabo la desnormalización de algunas tablas. Volviendo a `CLIENTE`, `RUTA` y `ALOJAMIENTO`, que estaban ya en 2FN, obtuvimos dos tablas a partir de una: la misma tabla y otra formada únicamente por las claves. Una de las mencionadas consideraciones es que la tabla que contiene solo las claves es redundante, por lo que eliminaremos todas las tablas de esta forma. Al realizar la normalización de las relaciones `PRECIO_RUTA` y `PRECIO_ALOJAMIENTO`, obtenemos unas tablas que pierden el atributo `Fecha_fin`, por lo que `Fecha_inicio` queda falta de contexto (aunque nosotros sepamos que los precios, si varían, lo hacen semestralmente, la información queda algo desligada). Además, tenemos una tabla más por solo perder un atributo, así que desnormalizamos para tener una única tabla formada por los demás atributos de la relación y el mencionado `Fecha_inicio`, ya que así reducimos el número de tablas y, por tanto, se aumenta en la optimización de la información. Por último, la concurrencia no será un problema por la razón previamente argumentada.

Sobre particionamiento de las tablas, no lo hemos visto necesario. Por un lado, el particionamiento horizontal no será de gran utilidad puesto que las tablas no son excesivamente grandes y en cuanto a la partición vertical, como las tablas son de pocas columnas, tampoco será utilizado. Además, para aquellas relaciones que sí cuentan con más columnas, como los datos son del mismo tipo (por ejemplo se refieren a datos personales o a datos de reservas) tampoco sufrirán dicho particionamiento vertical.

6.2. Migración de datos

Previa a la migración, el data scientist y el arquitecto de datos preguntaron por un par de cuestiones respecto a ciertos datos restantes en el modelo provisto que consideramos imprescindibles para finalizar el proyecto: los buses que realizan qué rutas, las ubicaciones de las nuevas sucursales, y un popurrí de datos sobre clientes que no podían quedar vacíos. Al equipo se nos proporcionaron los primeros y segundos, y para conseguir los terceros simplemente se hizo a todos los clientes, incluidos los que fueron solo a sucursal, que cuando entraron a la web, forzosamente tuvieran que iniciar sesión de vuelta (si no registrarse enteramente) pero ahora pidiendo un (y no más) teléfono de contacto; DNI, mail y dirección auténticas; y una contraseña segura: mínimo 8 caracteres, al menos una mayúscula y un número/símbolo.

Con todo listo para el migrado, el arquitecto consideró que sobre las direcciones (también viendo cómo vienen dadas estas en los datos aportados), podrían ser completas para ahorrarnos la atributo `Ciudad` en `CLIENTE` y `Provincia` en `SUCURSAL`.

En el propio migrado tuvieron que tomarse ciertas decisiones de diseño, en el contexto del programa MySQL Workbench:

- En la entrevista se nos indicó que la sucursal sin dirección ni gerente estaba cerrando, así que para no perder información sencillamente dejamos la entidad con ambos atributos en null, para denotar que ya no existe (también se decidió no borrarla físicamente porque hay una instancia de cliente que acudió a ella).
- Aunque haya bastantes claves que son duplas o ternas, algunos elementos que las conforman no pueden ser nulos o incluso deben ser únicos de por sí: DNI, Login, Mail y Password son *not null* y *unique* en `CLIENTE`, por ejemplo, pues es lógico que un mail o una contraseña sean obligatorios y cada usuario deba tener los suyos. Sin embargo, esto solo ocurre a veces: en `RUTA`, Localizador puede ser nulo (caso extraño, pero puede que ningún cliente haya elegido cierta ruta, o simplemente sea nueva) y no puede ser ya único porque puede haber

varios null (por lo demás, entendemos que un localizador es único al alojamiento/ruta individual que se reserve); y Id_ruta es lógico que se repita (varia gente, localizadores, harán la misma ruta).

- Más allá de las nuevas relaciones que hemos añadido (RRSS, atención al cliente, campaña), han ido apareciendo otras nuevas que permiten obtener información de nuevos contextos y que en este punto lo hacen patente: véanse Temporada_alojamiento y Temporada_ruta, permiten conocer qué destinos son más o menos populares con la temporada como variable. Estas nuevas tablas ahora mismo se encuentran vacías precisamente por su novedad (en el caso particular de RUTA y consecuentemente RESERVA, la entidad padre, se ponen placeholders en los localizadores por la ausencia de estos en los datos suministrados. Por su delicadeza y la no urgencia de añadirlos, así los dejamos por el momento. De hecho, los datos de precio, número de personas y fecha final para las rutas en RESERVA se han añadido por la lógica de la tabla PRECIO_RUTA y por suposiciones verosímiles respectivamente),
- En la sección de claves ajenas, se usa *On Update Cascade*, pues es casi algo usado por convención, y *On Delete Restrict* para prevenir errores humanos, además de que no preveemos un gran uso del borrado en el contexto de negocio que nos encontramos. En esta misma hay algunas que quedan pendientes de desbloquear por razones varias, principalmente falta de datos.

6.3. Vistas. Consultas

En pos de la securización de la BD, se presentan varias vistas en el archivo .sql presentado. La primera muestra un caso sencillo de acceso por parte de los empleados de la sucursal a la información identificativa de un cliente (por ejemplo, para poder hacerles una reserva cómodamente, sin tener que escudriñar por la tabla). La segunda y tercera presentan la información que un cliente recibiría en la web si buscara rutas y alojamientos, respectivamente. Sobre consultas, se indica un primer ejemplo sencillo de un caso real: el empleado de una sucursal busca la dirección de un cliente porque se ha mudado y hay que actualizarla. Los otros dos atañen cuestiones de optimización: la primera muestra los destinos que mayores beneficios han aportado, con datos identificativos del destino y el localizador de la reserva (se ven tendencias claras, que pueden llevar a dejar de ofertar hoteles como el de por su bajo rendimiento); y la segunda muestra las rutas más populares.

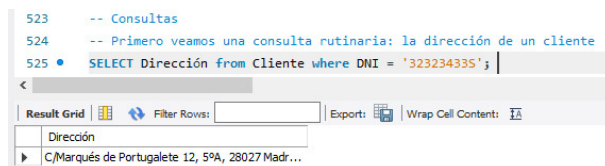


Figura 3: Consulta de prueba

```

526 -- Veamos qué destinos son los que generan mayores beneficios
527 • SELECT Alojamiento.Localizador, Id_destino, Nombre, Dirección,
528 Régimen, Precio_euros from Reserva join Alojamiento
529 on Reserva.Localizador = Alojamiento.Localizador
530 order by Precio_euros DESC;

```

Localizador	Id_destino	Nombre	Dirección	Régimen	Precio_euros
IIDK53	9	Hotel Estrella	C/Alonso Martínez 5, Madrid	AD	2399
KDLAFIL34	9	Hotel Estrella	C/Alonso Martínez 5, Madrid	AD	2000
88831F	6	Hotel Centrol	C/Alcala 4, Madrid	AD	1994
JDBN4	6	Hotel Centrol	C/Alcala 4, Madrid	AD	1874
ADF3LK33	10	Hotel Estrella	C/Alonso Martínez 5, Madrid	PC	1039
KKEI2233	2	Hotel Torredembarra	C/Torredembarra 1, Barcelona	MP	1000
KKD2443	2	Hotel Torredembarra	C/Torredembarra 1, Barcelona	MP	993
KDKJ33	12	Hotel la cala	C/Cala 3, La Pineda Tarragona	AD	876

Figura 4: Primera consulta

```

531 -- Veamos ahora qué rutas son las más populares
532 • SELECT Id_ruta, Nombre, count(Localizador) as num from Ruta
533 group by Id_ruta,Nombre;

```

Id_ruta	Nombre	num
1	Mediterráneo	9
2	De Los Caballeros	5
3	Cantábrico	3
4	Centro	0

Figura 5: Segunda consulta

7. Resumen de las decisiones tomadas

Para finalizar la memoria de la práctica, vamos a escribir un resumen con todas las consideraciones que hemos ido tomando a lo largo de su realización. Las primeras decisiones tomadas estuvieron influenciadas por la entrevista realizada con el CEO de la empresa, en la que se nos comentó que la empresa buscaba unificar los clientes que iban a las sucursales con aquellos que se registraban vía web y asimismo, que había sucursales que iban a cerrar por lo que en la base de datos no se tenía datos sobre ellas o que teníamos que adaptarnos a según en qué momento del año nos encontrábamos en relación a los precios. También se nos pidió que se optimizaran las rutas y aumentara la seguridad de la página; procesos que hemos realizado en los apartados 6.2 y 6.3.

Una vez realizado el modelo E/R preliminar y teniendo en cuenta las observaciones del CEO de la empresa, obtuvimos el modelo E/R definitivo, en el que consideramos que para aumentar los beneficios, tener más publicidad y ofrecer ayuda a los consumidores era necesario crear un apartado de atención al cliente, redes sociales y elaborar una serie de ofertas que atraigan potencialmente a nuevos clientes. Además, al estudiar nuestro modelo nos dimos cuenta de que algunos atributos eran redundantes y decidimos eliminarlos o cambiarlos de lugar y también se realizó el cambio de algunas claves.

En cuanto a la elaboración del modelo relacional, como ya están expuestas las decisiones adoptadas, basta con destacar que eliminamos la relación de FÍSICO al parecernos poco significativa.

Por último, en la parte del diseño físico se han mencionado las razones por las que hemos decidido desnormalizar algunas tablas, hemos mejorado la seguridad de la web mediante la introducción de restricciones en las contraseñas y que no se pueda escribir cualquier DNI al azar. Por último, hemos llevado a cabo una serie de consultas que consideramos interesantes como pueden ser los destinos que mayor beneficio aportan o cuáles son las rutas más populares.