

GESTIÓN DE DATOS 2020



Trabajo Práctico TP Frba Viajes - 1C 2020

Curso: K3151

Turno: Martes y Jueves (Noche)

Grupo: FELICES_PASCUAS

Nro. Grupo: 25

Integrantes:

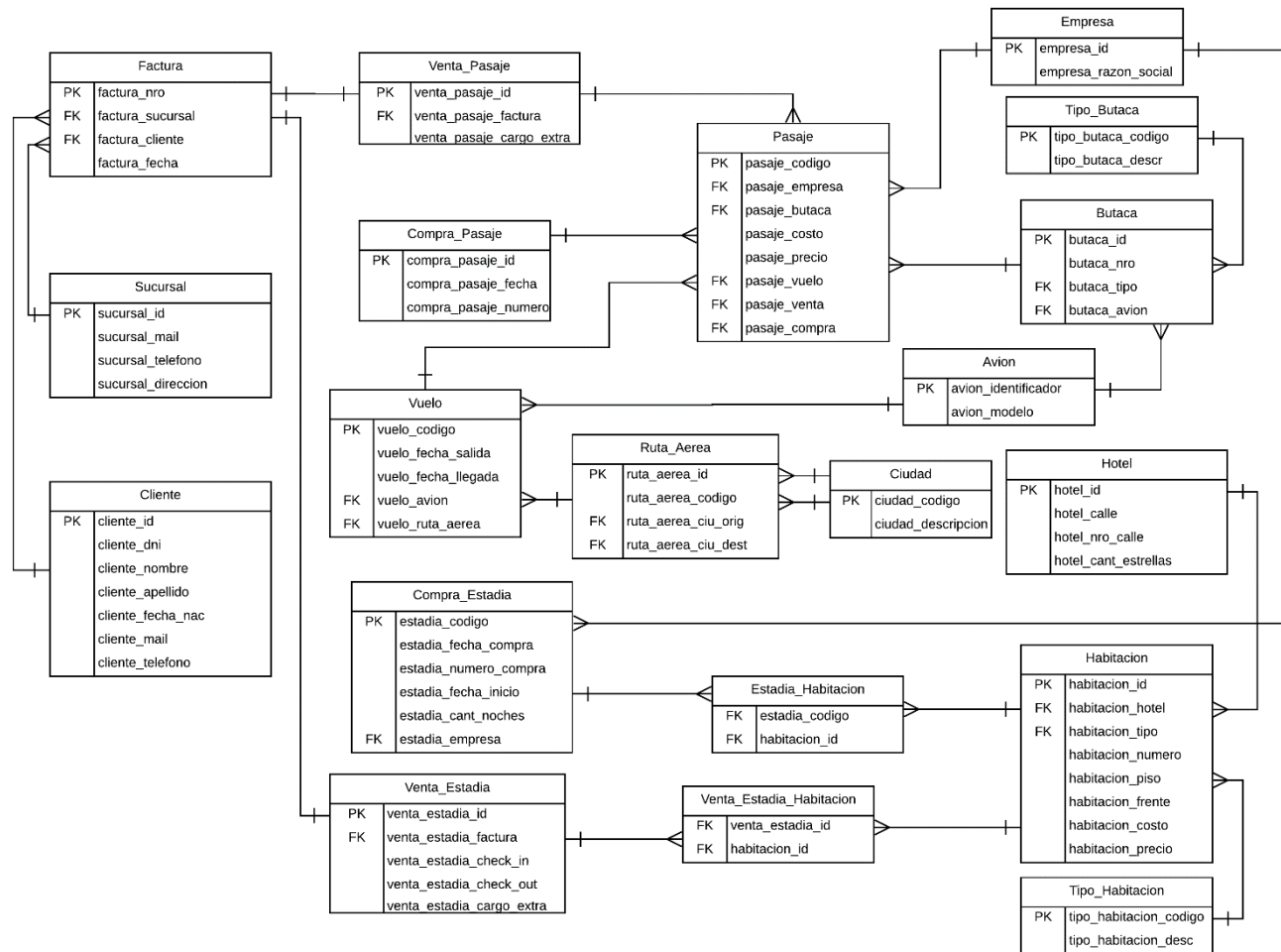
- Vitelli, Maximiliano 158.008-5
- Montenegro, Agustín 150.757-6
- Sorbelli, Luciana 141.977-8

INDICE

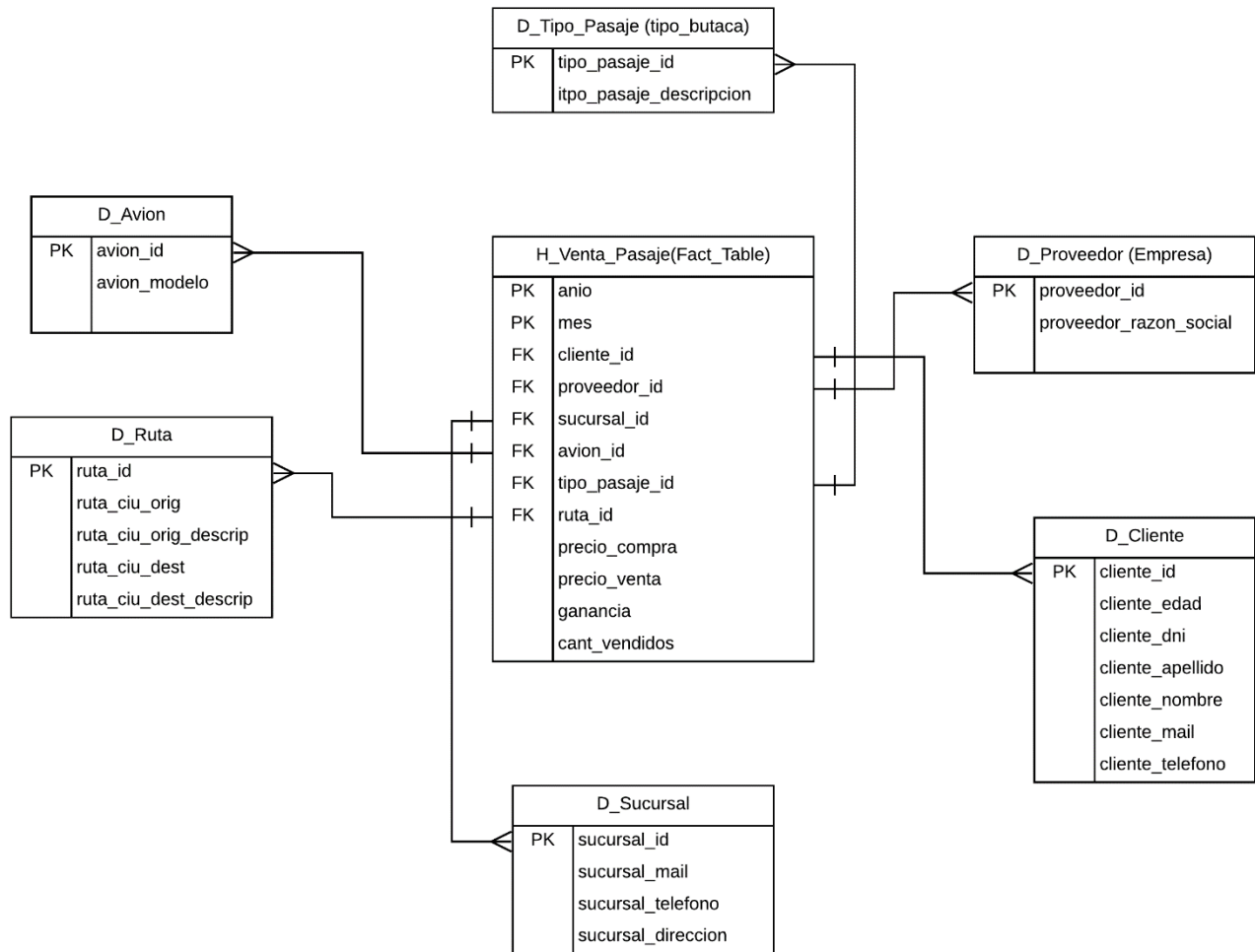
DER.....	3
MODELO OLTP.....	3
MODELO BI PASAJE.....	4
MODELO BI ESTADÍA.....	5
MIGRACIÓN DE DATOS.....	6
MODELO OLTP	6
TABLAS.....	6
MODELO BI	9
TABLAS.....	9
VISTAS.....	12
DESICIONES GENERALES.....	12
DESICIONES GENERALES SOBRE EL MODELO BI.....	13

DER

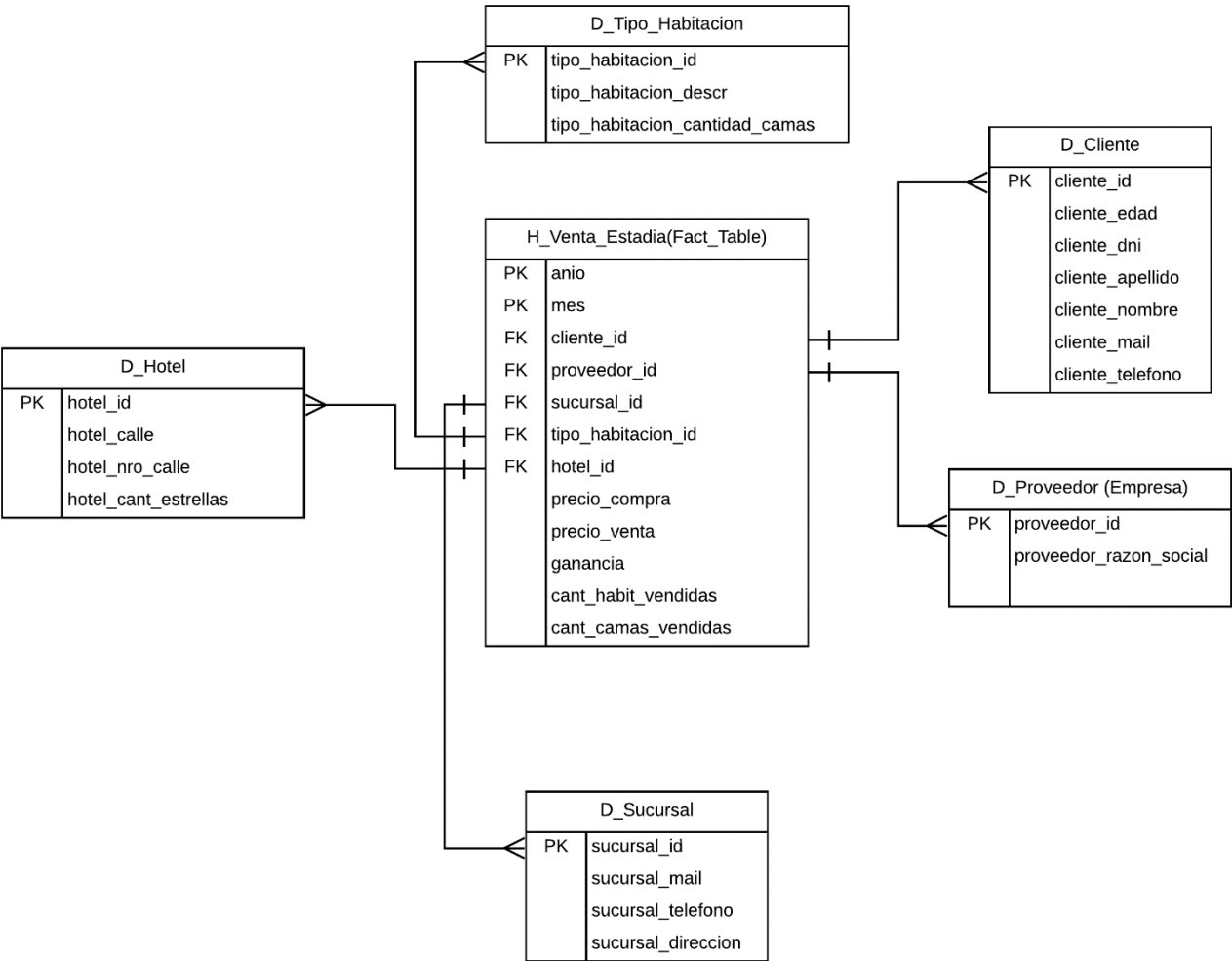
Modelo OLTP



Modelo BI - Pasaje



Modelo BI – Estadía



Migración de Datos

Modelo OLTP

Tablas

Explicamos las tablas más relevantes:

Cliente: Tabla que representa el modelado de todos los clientes. Llamamos cliente en nuestro sistema de venta a las entidades que pueden comprar pasajes o estadías ya adquiridas previamente por nosotros. Se identifican unívocamente por un campo ID. Se optó por un ID porque existían DNI repetidos en la tabla maestra y no era posible utilizar este campo como PK. Se almacena:

- el DNI,
- nombre y apellido,
- fecha de nacimiento,
- su dirección de correo electrónico y
- su teléfono de contacto.

Compra Estadía: Entidad que tiene la información de la compra de las estadías por parte de nuestra agencia respecto de la empresa que emite las estadías. En nuestro sistema se almacenan los siguientes datos de las compras de las estadías:

- código de estadía comprada (clave primaria de la tabla),
- fecha en la cual se realizó la compra,
- número de compra,
- fecha de inicio de la estadía comprada,
- cantidad de noches de la estadía comprada,
- empresa a la cual le pertenece la estadía comprada.

Compra Pasaje: Tabla que representa la compra de los pasajes. Así como sucede con las estadías, los pasajes adquiridos por nuestra agencia deben ser previamente comprados a una empresa que emite estos pasajes. Dentro de la tabla se encuentran los datos de esa compra que son los siguientes:

- identificador de la compra del pasaje (clave primaria de la tabla),
- fecha de compra del pasaje y
- número de compra.

Estadia Habitación: Tabla intermedia (o entidad relacionante) que relaciona las estadías con sus respectivas habitaciones que pueden tener las mismas. Los datos que se persisten son:

- Identificador perteneciente a la estadía comprada,
- Identificador perteneciente a la habitación.

Factura: Entidad de vital importancia, que representa los datos de facturación de todas las ventas por parte de nuestra agencia de viajes. Esta tabla está relacionada con las ventas de los pasajes y las estadías (por separado). Partiendo de la estrategia/decisión que tomamos que, al facturar, no se pueden vender estadías y pasajes en la misma factura, los datos almacenados son:

- número de factura (clave primaria de la tabla),
- sucursal donde se emitió la factura,
- cliente perteneciente a nuestra agencia al cual se emitió la factura y
- fecha de emisión de la factura.

Habitación: Tabla que contiene la información de las distintas habitaciones existentes dentro del alcance de nuestro sistema. Los campos de esta tabla son:

- identificador de la habitación (clave primaria de la tabla),
- hotel donde se encuentra la habitación,
- el tipo de habitación,
- el número de habitación,
- el piso del hotel donde se encuentra la habitación,
- booleano que indica si la habitación posee frente,
- costo de la habitación (precio de compra),
- precio de venta de la habitación a nuestros clientes (precio de venta en la factura).

Inconsistencia: Tabla en la que guardamos los datos duplicados tanto de estadías como de pasajes. Detectamos que por cada operación de compra/venta hay dos filas, una que tiene toda la información de la venta (factura, fecha, etc) y la otra que tiene todo idéntico pero los campos de venta en NULL. Los datos de una de estas filas los insertamos en esta tabla. Los campos son:

- identificador de inconsistencia (clave primaria),
- tipo de inconsistencia, clave foránea a la tabla Tipo_Inconsistencia,
- un texto con los datos repetidos, separados por coma

Tipo_Inconsistencia: Tabla para diferenciar las inconsistencias. Los campos que tiene son:

- Identificador de inconsistencia (clave primaria)
- Detalle de la inconsistencia o descripción de la tabla
- Los nombres de columnas o *headers* (en la tabla maestra), por si se quisiera recuperar esta información.

Pasaje: Entidad que representa los boletos que nuestra empresa adquiere y vendemos a nuestros clientes. Los campos de la tabla son:

- Código de pasaje (clave primaria),
- Empresa que nos vende el pasaje (clave foránea)
- Identificador de butaca (clave foránea)
- Costo del pasaje (precio de compra)
- Precio de la venta del pasaje a nuestros clientes
- Identificador de compra (clave foránea)
- Identificador de venta (clave foránea), si es que el pasaje fue vendido
- Identificador de vuelo

Pasaje Anomalo: Información inconsistente de pasajes. Son los registros que encontramos en la tabla maestra de pasajes distintos que hacían referencia a una misma butaca y fueron vendidos a más de una persona. Decidimos conservar el boleto del cliente que compró primero, siempre y cuando se haya vendido dicho pasaje. Tiene los mismos campos que la tabla Pasaje.

Venta Pasaje: Entidad en la que se registran las ventas a nuestros clientes de los pasajes. Los campos son:

- Identificador de venta (clave primaria),
- Identificador de factura (clave foránea),
- Cargo extra por la venta

Venta Estadía: Entidad en la que se registra las ventas a nuestros clientes de las estadías. Los campos son:

- Identificador de venta (clave primaria),
- Identificador de factura (clave foránea),
- Cargo extra por la venta
- Fecha de *check_in*
- Fecha de *check_out*

Venta Estadía Habitación: Tabla intermedia (o entidad relacionante) que relaciona las ventas de estadías con sus respectivas habitaciones. De esta manera, se pueden fragmentar las ventas de estadías para no tener que vender todo el paquete que compramos en una sola operación. Los datos que se persisten son:

- Identificador perteneciente a la venta de la estadía (clave foránea),
- Identificador perteneciente a la habitación (clave foránea)

Modelo BI

Tablas

D_Cliente: es la dimensión con la información detallada de nuestros clientes. Se substituyó la fecha de nacimiento por la edad. Se tomó la decisión de mantener toda la tabla en vez de tener una dimensión por cada dato de interés del cliente, como, por ejemplo: edad. De esta manera no se pierden los datos pero es necesario un JOIN más para hacer consultas relacionadas con esta tabla. Su clave primaria es el identificador y sus campos son:

- el DNI,
- nombre y apellido,
- edad (calculada a partir de su fecha de nacimiento),
- su dirección de correo electrónico y
- su teléfono de contacto.

D_Avion: es la dimensión con los datos del avión. Sus campos son:

- identificador (clave primaria),
- modelo

D_Hotel: dimensión que guarda los datos del hotel. La agregamos porque nos pareció útil poder distinguir en las consultas los hoteles para tener estadísticas de esto. Sus campos son:

- identificador (clave primaria),
- nombre de la calle,
- numero de domicilio,
- cantidad de estrellas

D_Proveedor: dimensión relacionada con la empresa que nos vende tanto los pasajes como las estadías. Sus campos son:

- identificador (clave primaria),
- razón social

D Ruta: dimensión que registra las rutas aéreas con sus respectivos par de ciudades, de manera desnormalizada. Tomamos la decisión de conservar esta dimensión con respecto a tener una dimensión Ciudad porque con un simple JOIN se pueden obtener los mismos filtros. Sus campos son:

- identificador (clave primaria),
- identificador de la ciudad de origen,
- descripción de la ciudad origen,
- identificador de la ciudad destino,
- descripción de la ciudad destino

D Sucursal: dimensión que detalla dónde facturamos las ventas. Sus campos son:

- identificador (clave primaria),
- dirección,
- correo electrónico,
- teléfono de contacto

D Tipo Habitación: dimensión que registra los tipos de habitación con los que operamos y su descripción. Decidimos asignarle a cada tipo un número de camas que representa para facilitar estadísticas. Sus campos son:

- identificador (clave primaria),
- descripción,
- cantidad de camas que tiene cada habitación de ese tipo

D Tipo Pasaje: dimensión que describe los tipos de butaca/pasaje que vendemos y su descripción. Sus campos son:

- identificador (clave primaria),
- descripción

H_Venta_Estadía: tabla de hecho que agrupa todas las operaciones de venta de estadías dadas por las siguientes dimensiones: año, mes, cliente, proveedor, sucursal, tipo_habitacion, hotel. Consideramos pre-calcular atributos relevantes para el negocio como: precio de compra, precio de venta, ganancia, cantidad de habitaciones vendidas, cantidad de camas vendidas. Sus campos son:

- año,
- mes,
- cliente,
- proveedor,
- sucursal,
- tipo de habitación
- hotel,
- precio de compra,
- precio de venta,
- ganancia,
- cantidad de habitaciones vendidas,
- cantidad de camas vendidas

H_Venta_Pasaje: tabla de hecho que agrupa todas las operaciones de ventas de pasajes dadas por las siguientes dimensiones: año, mes, cliente, proveedor, sucursal, avión, tipo_pasaje, ruta aérea. Consideramos pre-calcular atributos relevantes para el negocio como: precio de compra, precio de venta, ganancia, cantidad de pasajes vendidos. Dada la combinación (group by) de dimensiones de esta tabla no se agrupan muchas filas al hacer el insert. Se mantiene una relación de 1 a 1 con la tabla de venta de pasajes. Sus campos son:

- año,
- mes,
- cliente,
- proveedor,
- sucursal,
- avión,
- tipo de pasaje (es decir, tipo de butaca),
- ruta aérea,
- precio de compra,
- precio de venta,
- ganancia,
- cantidad de pasajes vendidos

Vistas

Creamos algunas vistas de ejemplo para mostrar cómo serían las consultas en las tablas de hecho.

V Camas Vendidas Marzo 2018: consulta que retorna la cantidad de camas vendidas en el mes de Marzo de 2018.

V Cantidad Pasajes Ejecutiva: consulta que retorna la cantidad de pasajes vendidos de la clase “Ejecutiva”.

V Ganancia Total Estadías: consulta que retorna la ganancia total de todas las estadías vendidas.

Decisiones generales

- Como primary key de tablas que no tenían un identificador en la Maestra, utilizamos la función ***“select row_number() over (order by (select NULL))”***. Lo que hace esta función es darte el número de fila que estás insertando. Así quedaría como en esta query de la tabla Empresa.

	empresa_id	empresa_razon_social
1	1	China Southern Airlines
2	2	AccorHotels
3	3	American Airlines
4	4	Ryanair
5	5	British Airways
6	6	Choice Hotels International Inc
7	7	United Airlines
8	8	Shanghai Jin Jiang International Hotel Group Inc.
9	9	Emirates
10	10	China Eastern Airlines
11	11	Best Western Hotels & Resorts
12	12	Marriott International
13	13	Delta Airlines
14	14	easyJet
15	15	Hilton

Decidimos optar por esta estrategia en vez de una **CONSTRAINT PK AUTOINCREMENT** ya que de esta manera perdemos el control de la PK porque incrementaría infinitamente y se vería afectada la performance.

- Con respecto a los pasajes vendidos a distintos clientes de una misma butaca y un mismo vuelo, la estrategia que utilizamos para persistir la información fue: insertar ambos registros en la tabla Pasaje.

Luego, actualizamos y eliminamos de dicha tabla la fila que correspondía al cliente que compró en segundo lugar (por orden de fecha).

- Para unificar en la tabla Pasaje las operaciones de compra/venta y resolver el problema de que haya 2 filas por cada código de pasaje lo que hicimos fue realizar una consulta a la tabla maestra agrupando por PASAJE_CODIGO, insertar dichos registros y luego volver a consultar para actualizar los vendidos.

Decisiones generales sobre Modelo BI

- **Dimensión tiempo:** se optó por no tener una dimensión tiempo aunque sí existe como atributo de ambas tablas de hecho. Se utilizó el mes y año de las operaciones.
- **Tablas de hecho:** para nuestro modelo BI utilizamos 2 tablas de hecho para tener en cada una información más relacionada con la operación, de esta manera se evitó tener valores nulos como Foreign Keys.
- **Compras:** en nuestro modelo sólo utilizamos de la compra el precio o costo, no vimos necesario o relevante ningún otro dato.