Universidade Vigo

ESCOLA SUPERIOR DE ENXEÑARÍA INFORMÁTICA

Grupo 2_3 - RENFE

Alumna/o: Daniel Jorge Fernández Iglesias
Alumna/o: Xavier Iglesias Casal
Alumna/o: Carlos Piñeiro García

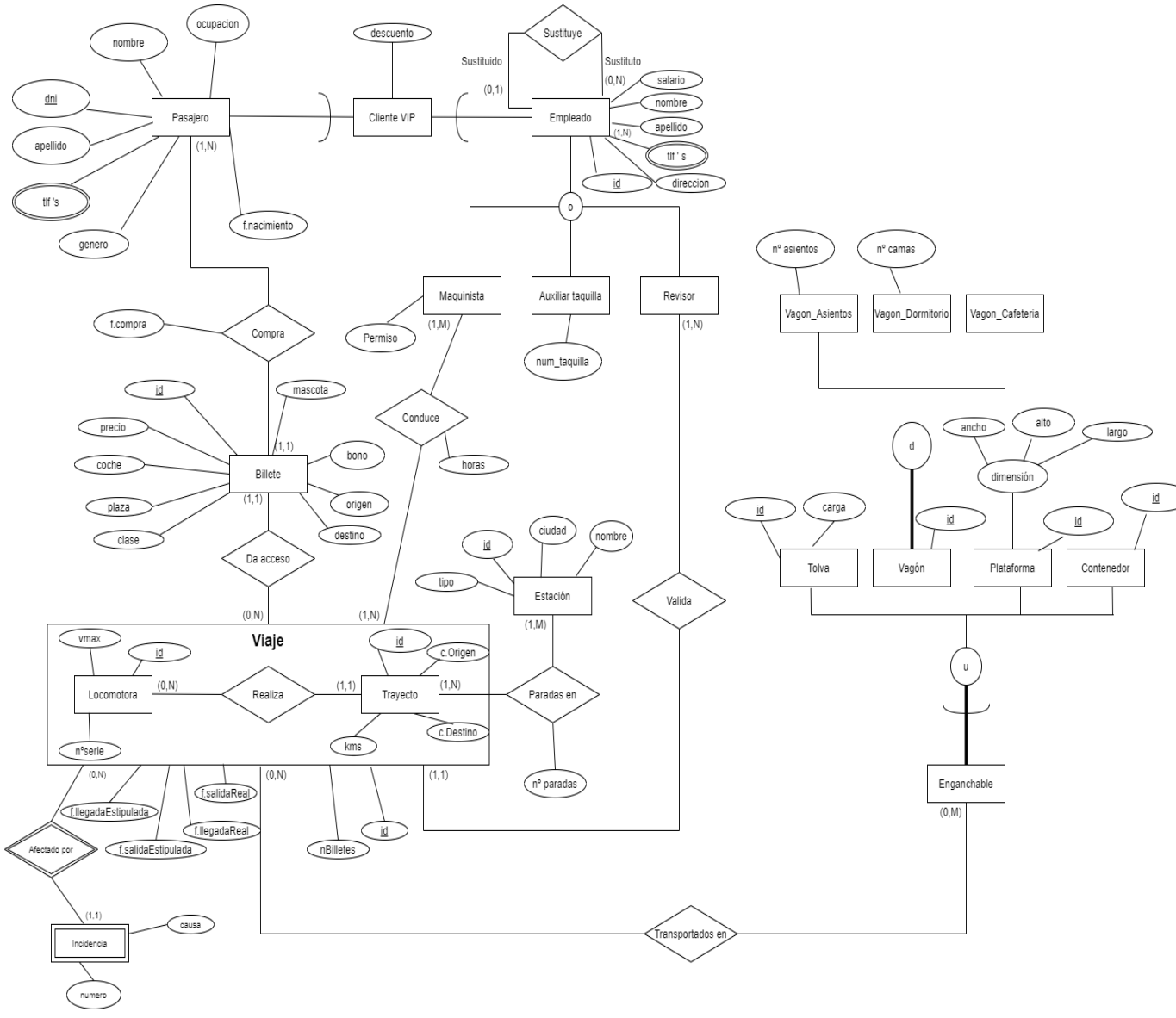
ÍNDICE

1. Descripción del proceso

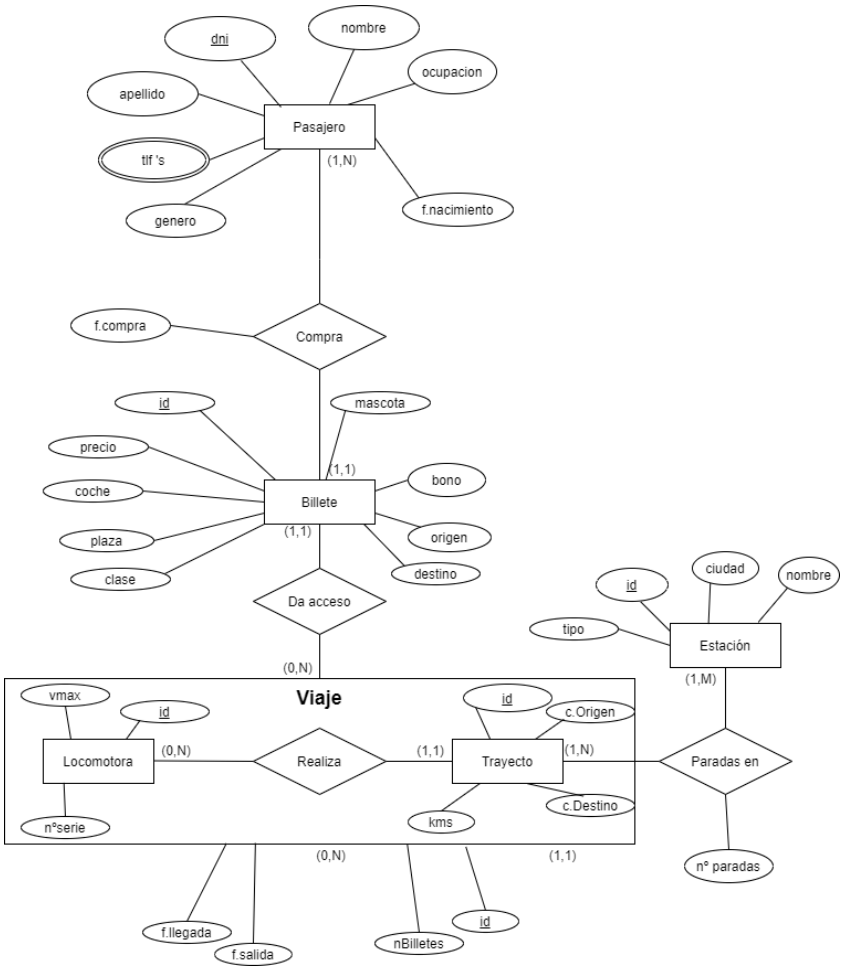
1.1. Descripción textual de la organización

Se trata de una empresa de trenes como RENFE. Su interés principal es vender billetes a pasajeros. Estos billetes dan permiso de acceso a un viaje, que tiene sus propios parámetros. Estos parámetros, como la fecha, la distancia o el origen, serán muy útiles para poder identificar temporadas de viaje. También serán relevantes datos referentes al billete, como la cantidad que se compran de una vez o la antelación con la que se adquieren. El pasajero dará, del mismo modo, información relevante para analizar. Juntando toda esta información, deberíamos ser capaces de identificar correctamente las temporadas altas y bajas de viaje, así como el perfil de pasajero que esperar. Todo esto será crucial para vender más billetes.

1.2. Modelo E/R original correspondiente al sistema OLTP preexistente



1.3. Modelo E/R con las entidades que almacenan información de interés para el mercado de datos



1.3.1. Explicar la necesidad de incorporación de nuevos atributos a los preexistentes en la organización, en su caso.

Se han añadido nuevos atributos “género”, “ocupación” y “f.nacimiento” al pasajero para poder determinar mejor el tipo de cliente del que se trata.

Se han añadido nuevos atributos “mascota”, “bono”, “origen” y “destino” para, los primeros, tener un mejor control sobre las variaciones que puede sufrir el precio del billete. Los siguientes, para tener más a mano la información relativa al viaje al que da acceso dicho billete.

Se ha añadido un atributo “kilómetros” al viaje para ver de manera más directa la distancia que supone el mismo.

Se ha añadido un atributo “tipo_ciudad” a la estación para poder determinar más rápidamente en qué situación geográfica (playa, montaña) se encuentra.

1.4. Descripción de las fuentes externas

FUENTE	**Ciudad**	
Descripción	**Listado completo de todas las ciudades donde hay estaciones de RENFE**	
ATRIBUTO	**TIPO**	**DESCRIPCIÓN**
id_ciudad	**int**	
código	**text**	**Identificador numérico asignado a cada estación**
nombre	**text**	**Nombre de la ciudad**
latitud	**text**	**Latitud de la estación**
longitud	**text**	**Longitud de la estación**

direccion	**text**	**Dirección de la estación**
c.p.	**text**	**Código postal en el que se encuentra la estación**
población	**text**	**Ciudad de la estación**
provincia	**text**	**Provincia de la estacíon**
pais	**text**	**País de la estación**
cercanías	**text**	**Si posee cercanías o no**
feve	**text**	**Si posee ferrocarril o no**

FUENTE	**Días festivos**	
Descripción	**Listado de todos los días festivos que ha habido desde el año 2020**	
ATRIBUTO	**TIPO**	**DESCRIPCIÓN**
provincia	**text**	**Nombre de la provincia**
enero	**text**	**Días festivos en enero**
febrero	**text**	**Días festivos en febrero**
marzo	**text**	**Días festivos en marzo**
abril	**text**	**Días festivos en abril**
mayo	**text**	**Días festivos en mayo**
junio	**text**	**Días festivos en junio**
julio	**text**	**Días festivos en julio**
agosto	**text**	**Días festivos en agosto**
septiembre	**text**	**Días festivos en septiembre**
octubre	**text**	**Días festivos en octubre**
noviembre	**text**	**Días festivos en noviembre**
diciembre	**text**	**Días festivos en diciembre**

1.5. Descripción textual de la actividad a modelar
Proponemos usar el modelo de almacenes de datos para poder identificar las temporadas altas y bajas de viajes. También tener conocimiento de todos los viajes para ver cuáles son más concurridos. Proponemos analizar la venta de billetes para poder planificar y gestionar de una manera efectiva la empresa, conociendo los horarios o fechas importantes y prever sobre ello.

2. Selección de la granularidad

2.1. Descripción del gránulo de la actividad a modelar
Gránulo mayor: se quiere tener información de las ventas en temporadas concretas, analizar el comportamiento de los pasajeros en función de la duración del viaje o analizar la velocidad de los trenes para identificar donde se producen retrasos.

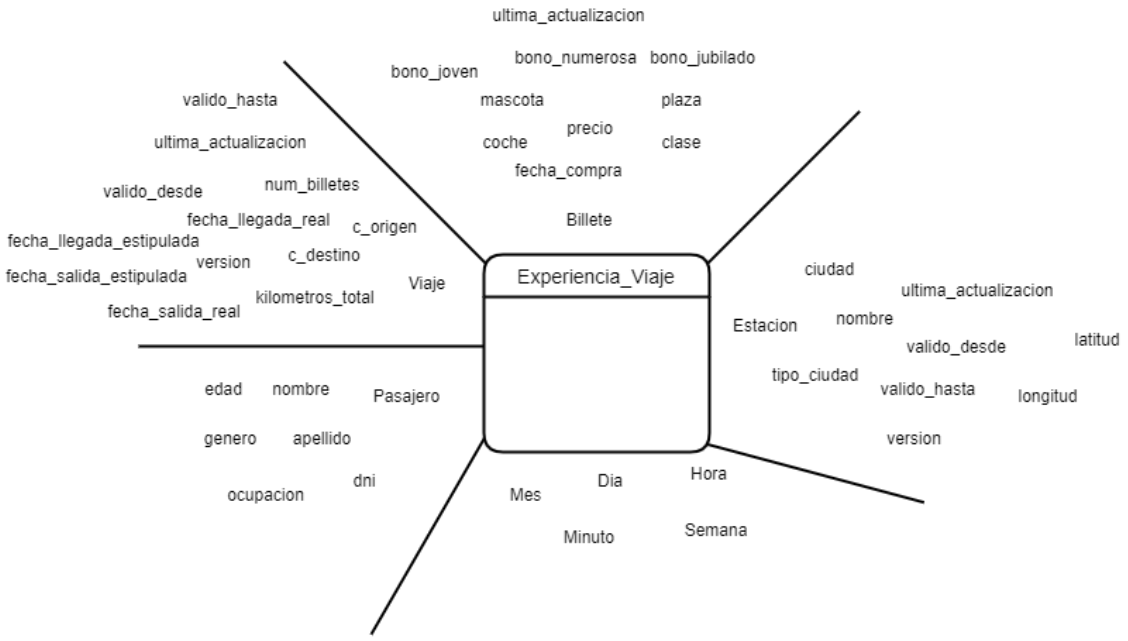
2.2. Razonar por qué no se ha optado por una granularidad mayor o menor
No se ha optado por un gránulo menor porque sí nos interesa almacenar información a nivel de billete, para saber a dónde se viaja en qué temporadas y así. Una granularidad mayor, con datos más generales, nos haría perder fechas importantes de viajes, dificultando la definición de temporadas.

3. Identificación de las dimensiones

3.1. Descripción textual de las dimensiones (a nivel general, sin detalle de los atributos)
Se han definido diferentes dimensiones que consideramos necesarias para saber cómo aumentar los beneficios de las ventas de billetes.
Dimensión tiempo: en esta dimensión se trabaja a nivel de día. Es necesario conocer cuándo se han comprado los billetes y en qué día se van a utilizar. Esto será útil para poder definir las máximas temporadas de viaje

Dimensión pasajero: este es el cliente. La información que se obtiene de él es necesaria para conocer el tipo de clientes que hay en relación a los viajes que se realizan
Dimensión viaje: este es el "lugar" en el que se realiza la acción. El producto son los billetes, y estos otorgan acceso a los viajes. De los viajes queremos saber el origen y el destino, así como su distancia. Esto podrá identificar los trenes más interesantes para el público.
Dimensión billete: el producto que queremos vender. De él es útil conocer el precio, los bonos que se le aplican y las fechas de venta, salida y llegada.
Dimensión estación: aquí se analiza la información de los lugares de viajes, tanto de salida como de llegada. Esto nos permite ver desde dónde y hacia dónde se realizan más viajes.

3.2. Diagrama de las dimensiones

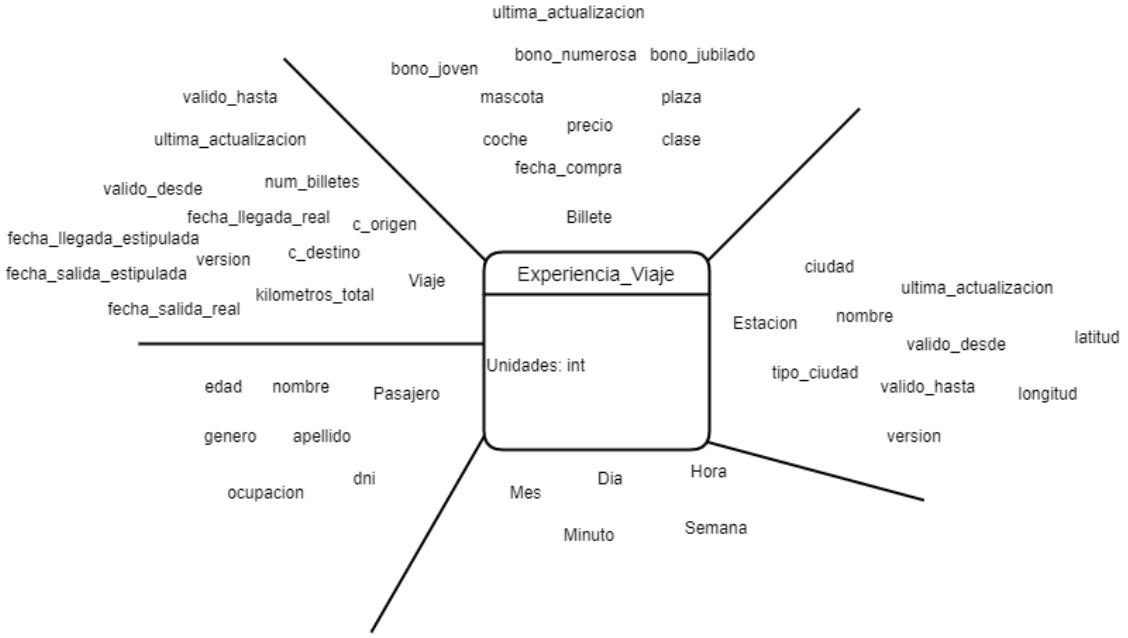


4. Selección de las medidas asociadas al hecho

4.1. Descripción textual de las medidas y de dónde se obtienen (a qué atributo(s) se corresponde en la BD OLTP).

MEDIDA	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGEN	ATRIBUTO/CAMPO ORIGEN
unidades	int	unidades vendidas	por defecto (1)	

4.2. Diagrama con la incorporación de las medidas



5. Almacenamiento de valores precalculados en la tabla de hechos

5.1. Definición de valores precalculados y descripción acerca de cómo se obtienen

precio_final: este valor representa el importe total de la venta tras haber aplicado el descuento por los bonos y la tasa de la mascota.

precio_final = precio + (mascota*precio) - (bono*precio)

diferencia_temporal: la diferencia de tiempo entre la fecha de llegada real y la fecha de llegada estipulada..

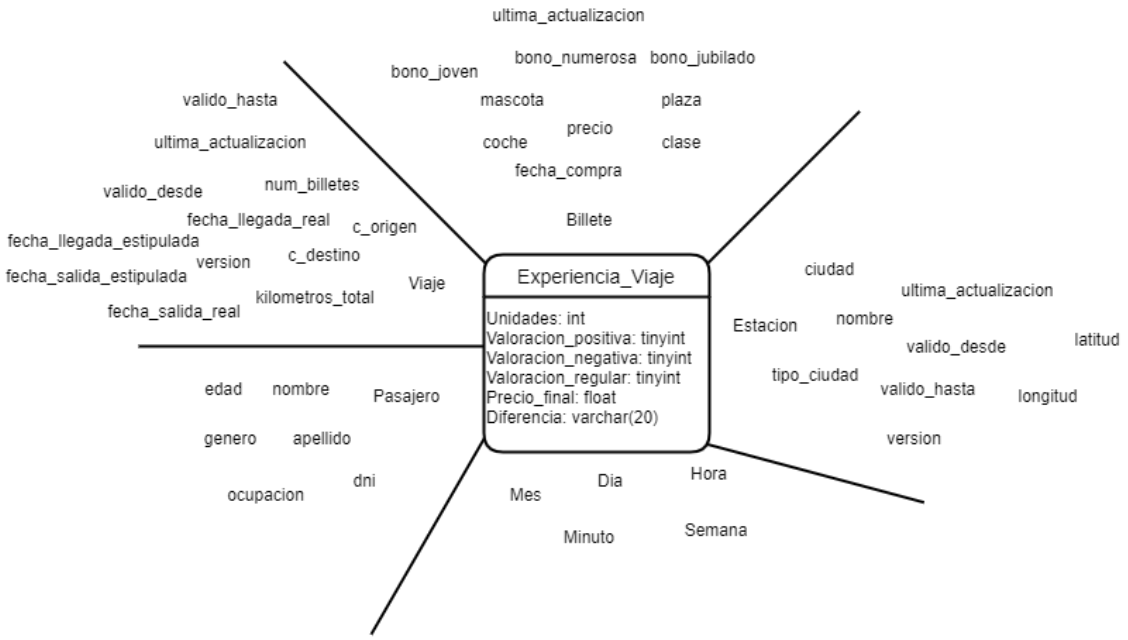
diferencia_temporal = fecha_llegada_real - fecha_llegada_estipulada

valoracion_positiva = diferencia_temporal <= 0

valoracion_regular = diferencia_temporal > 0 && diferencia_temporal <= 10

valoracion_negativa = diferencia_temporal > 10

5.2. Diagrama con incorporación de los valores precalculados



6. Terminación de las tablas de dimensión

6.1. Descripción detallada de las dimensiones. Para cada uno de sus atributos indicar:

6.1.1. Cómo se obtiene (atributo(s) específico(s) en OLTP, campo calculado, fuente externa X, etc.)

DIMENSIÓN	dim_viaje			
Descripción de la dimensión	Dimensión donde se guardan los datos relevantes al viaje			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGEN	ATRIBUTO/CAMPO ORIGEN
viaje_key	integer	clave primaria de viaje	dim_viaje	viaje_key
id_viaje	bigint	id de viaje	viaje	id
kilometros_total	smallint	cantidad de kilometros	viaje	kms
num_billtes	smallint	número de billetes	viaje	n_billetes
c_origen	varchar(20)	ciudad de origen	trayecto	c_origen
c_destino	varchar(20)	ciudad de destino	trayecto	c_destino
fecha_salida_estipulada	datetime	fecha de salida estipulada para el viaje	viaje	f.salidaEstipulada
fecha_llegada_estipulada	datetime	fecha de llegada estipulada para el viaje	viaje	f.llegadaEstipulada
fecha_salida_real	datetime	fecha de salida real del viaje	viaje	f.salidaReal
fecha_llegada_real		fecha de llegada real del viaje	viaje	f.salidaReal

DIMENSIÓN	dim_estacion			
Descripción de la dimensión	Dimensión donde se guardan los datos relevantes a la estacion			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGEN	ATRIBUTO/CAMPO ORIGEN
estacion_key	integer	clave primaria de estacion	dim_estacion	estacion_key
id_estacion	decimal(20)	id de la estacion	estacion	id
ciudad	varchar(100)	nombre de la ciudad	estacion	ciudad
nombre	varchar(100)	nombre de la estacion	estacion	nombre
tipo_ciudad	varchar(20)	tipo de ciudad	estacion	tipo
latitud	varchar(20)	latitud de la estacion	Ciudades listado completo	latitud
longitud	varchar(20)	longitud de la estacion	Ciudades listado completo	longitud

DIMENSIÓN	dim_pasajero

MODELADO MULTIDIMENSIONAL

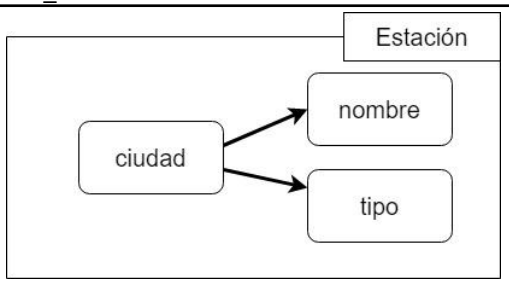
Descripción de la dimensión	Dimensión donde se guardan los datos relevantes al viaje			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGEN	ATRIBUTO/CAMPO ORIGEN
pasajero_key	integer	clave primaria de pasajero	dim_pasajero	pasajero_key
nombre	varchar(20)	nombre del pasajero	pasajero	nombre
apellido	varchar(25)	apellido del pasajero	pasajero	apellido
dni	varchar(9)	dni del pasajero	pasajero	dni
ocupacion	varchar(50)	a qué se dedica el pasajero	pasajero	ocupacion
genero	varchar(20)	género del pasajero	pasajero	genero
edad	smallint	edad del pasajero	dim_pasajero	edad

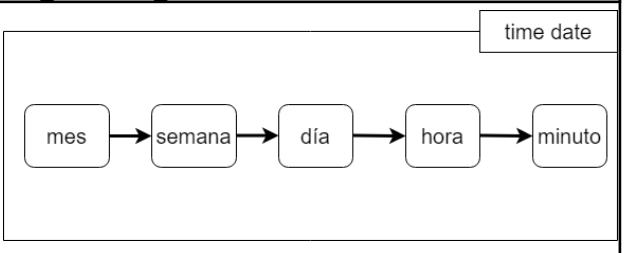
DIMENSIÓN	dim_tiempo			
Descripción de la dimensión	Momento en horas de realización de una venta			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGEN	ATRIBUTO/CAMPO ORIGEN
time_key	integer	clave primaria de time		
time_value	time	hora		
hours	tinyint	hora		
minutes	tinyint	minuto		

DIMENSIÓN	dim_date			
Descripción de la dimensión	Fecha de la realización de la venta			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGEN	ATRIBUTO/CAMPO ORIGEN
date_key	integer	clave primaria de date		
date_value	date	fecha		
day_name	char(12)	día de la semana de la venta		
week_in_year	tinyint	semana del año de la venta		
month_name	char(12)	nombre del mes de la venta		

DIMENSIÓN	dim_billete			
Descripción de la dimensión	Dimensión donde se guardan los datos relevantes al billete			
ATRIBUTO	TIPO	DESCRIPCIÓN	TABLA/FUENTE ORIGEN	ATRIBUTO/CAMPO ORIGEN
billete_key	integer	clave primaria del billete	dim_billete	billete_key
fecha_compra	date	fecha de compra del billete	billete	fecha_compra
precio	float	precio del billete	billete	precio
coche	tinyint	coche a que da acceso el billete	billete	coche
plaza	varchar(3)	asiento al que da acceso el billete	billete	plaza
clase	varchar(15)	clase de pasajero	billete	clase
mascota	boolean	lleva o no mascota	billete	mascota
bono_joven	tinyint(1)	tipo de bono aplicado al precio	billete	bono
bono_numerosa	tinyint(1)	tipo de bono aplicado al precio	billete	bono
bono_jubilado	tinyint(1)	tipo de bono aplicado al precio	billete	bono

6.1.2. Establecimiento de la jerarquía entre los atributos de cada dimensión (p.ej: día, mes, año)

DIMENSIÓN	dim_estacion
Jerarquía	

DIMENSIÓN	dim_date/dim_time
Jerarquía	

6.1.3. Conversión de datos: numerización/etiquetado, discretización (hacer al menos 2)

DIMENSIÓN	dim_pasajero	
ATRIBUTO	TIPO DE CONVERSIÓN	DESCRIPCIÓN
edad	discretización	[niño < 13] [joven > 13 & < 23] [adulto > 23 & < 65] [anciano > 65]

DIMENSIÓN	dim_billete	
ATRIBUTO	TIPO DE CONVERSIÓN	DESCRIPCIÓN
bono_joven	numerización	[C_JOVEN->15%]
bono_numerosa	numerización	[FAMILIA_NUMEROSA -> 15%]
bono_jubilado	numerización	[JUBILADO->65%]
mascota	numerización	[MASCOTA=1->25%]

6.1.4. Cómo se debe tratar en caso de datos erróneos

DIMENSIÓN	dim_billete
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO
precio	Precio menor que 0. precio = -1

DIMENSIÓN	dim_viaje
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO
fecha_salida	Si fecha_salida mayor que fecha_llegada: fecha_salida ”1900-01-01 00:00:00”

DIMENSIÓN	dim_viaje
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO
kilometros	Distancia incorrecta. kilometros = -1

DIMENSIÓN	dim_pasajero
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO
edad	Edad mayor de 80. edad = 80

DIMENSIÓN	dim_estacion
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO
nombre	Estacion de otra ciudad. nombre = “no corresponde”

6.1.5. Cómo se debe tratar en caso de datos faltantes

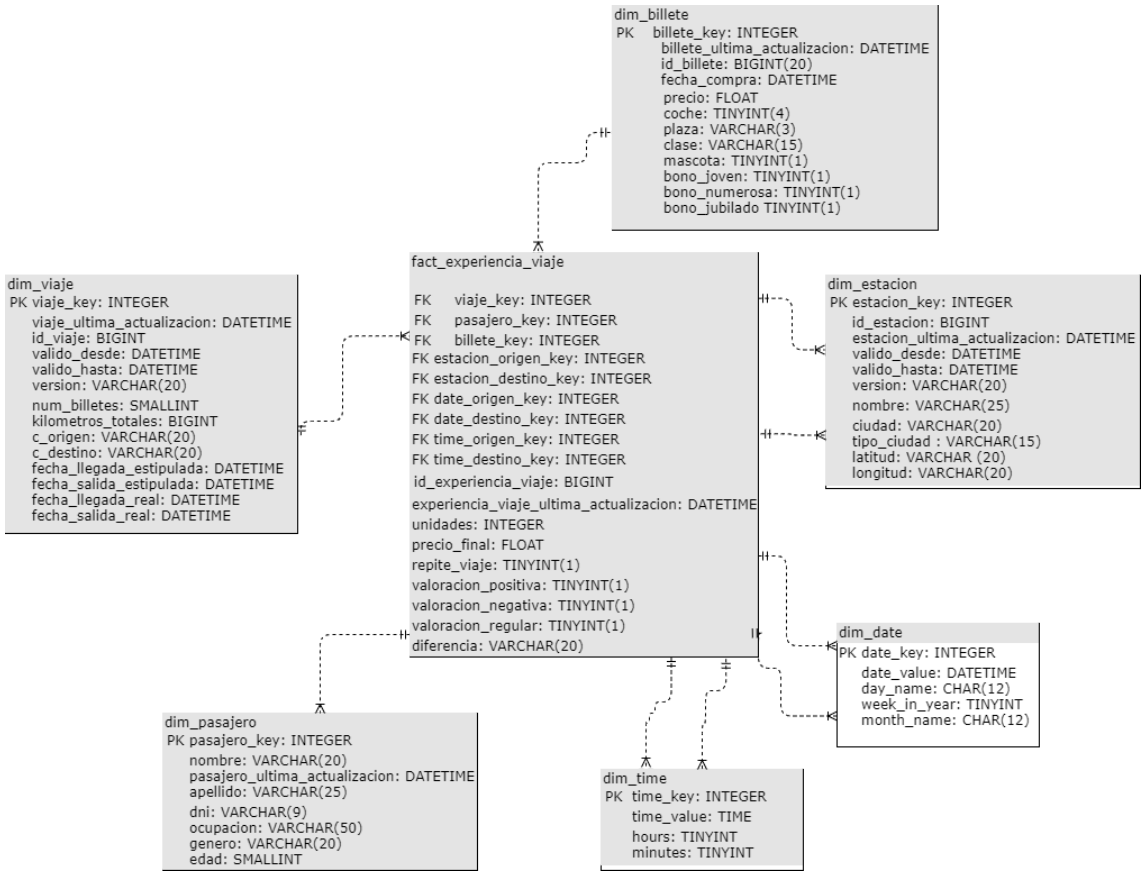
DIMENSIÓN	dim_pasajero
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE PROBLEMA Y CÓMO RESOLVERLO
genero	Falta el atributo. Se escribe “NA”
edad	Falta el atributo. Se escribe 0.

DIMENSIÓN	dim_viaje
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE PROBLEMA Y CÓMO RESOLVERLO
kilometros	Falta el atributo. Se escribe 0.

DIMENSIÓN	dim_billete
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE PROBLEMA Y CÓMO RESOLVERLO
mascota	Falta el atributo. Se escribe false.

DIMENSIÓN	dim_estacion
ATRIBUTO	DESCRIPCIÓN DEL TIPO DE PROBLEMA Y CÓMO RESOLVERLO
tipo_ciudad	Falta el atributo. Se escribe “falta”.

6.2. Diagrama final de diseño de la BDDatawarehouse versión ROLAP (incorporando claves)



7. Selección de la duración de la base de datos

7.1. Indicación de la duración elegida y justificación

Dos años.

Debemos conocer los viajes, fechas y precios con los que se ha estado trabajando hasta ahora. Esto puede dar una primera idea de las temporadas de viajes más comunes.

También es importante conocer los precios para discurrir las variaciones que hayan podido sufrir.

8. Control de las dimensiones lentamente cambiantes

8.1. Determinación de las dimensiones lentamente cambiantes y cuáles son los atributos afectados.
dim_viaje: kilimetros
dim_estacion: nombre

8.2. Para cada uno de los atributos cambiantes indicar cómo será tratado (Tipo 1, 2 o 3).*Es necesario incluir al menos un atributo cambiante Tipo 2.*
kilometros: tipo 2
nombre: tipo 2

9. Modos de consulta

9.1. Descripción textual de las consultas OLAP

MEDIDAS: ***"Mostrar la evolución de…"***	
- *precio_final*	- *unidades*

PASAJERO	BILLETE	TIEMPO	VIAJE	ESTACION
"de pasajeros…"	**"de billetes…"**	**"adquiridos…"**	**"de viajes…"**	**"de estaciones…"**
de edad X	con/sin mascota	en temporada alta	con muchos billetes	en la ciudad X
de género Y o Z	con bono X, Y	en días sueltos	con pocos billetes	en ciudades tipo Y
con ocupación K	de clase Z	…	en fechas festivas	…