

**E**scola **S**uperior de **E**nxeñaría **I**nformática

|  |
| --- |
| PRÁCTICA INTELIGENCIA DE NEGOCIO Y ALMACENES DE DATOS  (curso 2020-2021)  **Código Grupo: G2\_1**  **Título: HOTEL** |

Alumna/o: Rubén Gómez Dopazo

Alumna/o: Antonio Cebreiro Bernárdez

Alumna/o: Manuel Certal Puga

Alumna/o: César Gabriel Márquez Rodríguez

ÍNDICE

[1. Descripción del proceso 3](#_Toc66904921)

[1.1 Descripción textual de la organización 3](#_Toc66904922)

[1.2 Modelo E/R original del sistema OLTP preexistente 3](#_Toc66904923)

[1.3 Modelo E/R con las entidades que almacenan información de interés para el mercado de datos 5](#_Toc66904924)

[1.4 Descripción de las fuentes externas 10](#_Toc66904925)

[1.5 Descripción textual de la actividad a modelar 11](#_Toc66904926)

[2. Selección de la granularidad 12](#_Toc66904927)

[2.1 Descripción del gránulo de la actividad a modelar y de cada una de las dimensiones 12](#_Toc66904928)

[2.2 Razonar por qué no se ha optado por una granularidad mayor o menor 12](#_Toc66904929)

[3. Identificación de las dimensiones 14](#_Toc66904930)

[3.1 Descripción textual de las dimensiones (a nivel general) 14](#_Toc66904931)

[3.2 Diagrama de las dimensiones 15](#_Toc66904932)

[4. Selección de las medidas asociadas al hecho 16](#_Toc66904933)

[4.1 Descripción textual de las medidas y de dónde se obtienen (a qué atributo(s) se corresponde en la BD OLTP). 16](#_Toc66904934)

[4.2 Diagrama con la incorporación de las medidas 16](#_Toc66904935)

[5. Almacenamiento de valores precalculados en la tabla de hechos 17](#_Toc66904936)

[5.1 Definición de valores precalculados y descripción acerca de cómo se obtienen 17](#_Toc66904937)

[5.2 Diagrama con incorporación de los valores precalculados 17](#_Toc66904938)

[6. Terminación de las tablas de dimensión 18](#_Toc66904939)

[6.1 Descripción detallada de las dimensiones. 18](#_Toc66904940)

[6.2 Diagrama final de diseño deL DWH versión ROLAP (incorporando claves) 28](#_Toc66904941)

[7. Selección de la duración de la base de datos 29](#_Toc66904942)

[7.1 Indicación de la duración elegida y justificación 29](#_Toc66904943)

[8. Control de las dimensiones lentamente cambiantes 29](#_Toc66904944)

[8.1 Determinación de las dimensiones lentamente cambiantes y cuáles son los atributos afectados 29](#_Toc66904945)

[8.2 Tratamiento de atributos cambiantes indicar cómo será tratado (Tipo 1, 2 o 3) 29](#_Toc66904946)

[9. Modos de consulta y prioridades de ejecución 30](#_Toc66904947)

[9.1 Descripción textual de las consultas OLAP 30](#_Toc66904948)

# Descripción del proceso

# Descripción textual de la organización

Elite Hotels es una compañía hotelera que cuenta con numerosos hoteles de lujo. Cada uno cuenta con uno o varios edificios, distribuidos en diferentes plantas. Los hoteles de la cadena cuentan con todo tipo de instalaciones para el disfrute de los huéspedes, desde restaurantes y gimnasios hasta campos de golf o tenis.

En sus hoteles, la compañía ofrece dos tipos de hospedaje: apartamentos y habitaciones. Además, existen espacios de trabajo únicamente accesibles para los trabajadores del hotel.

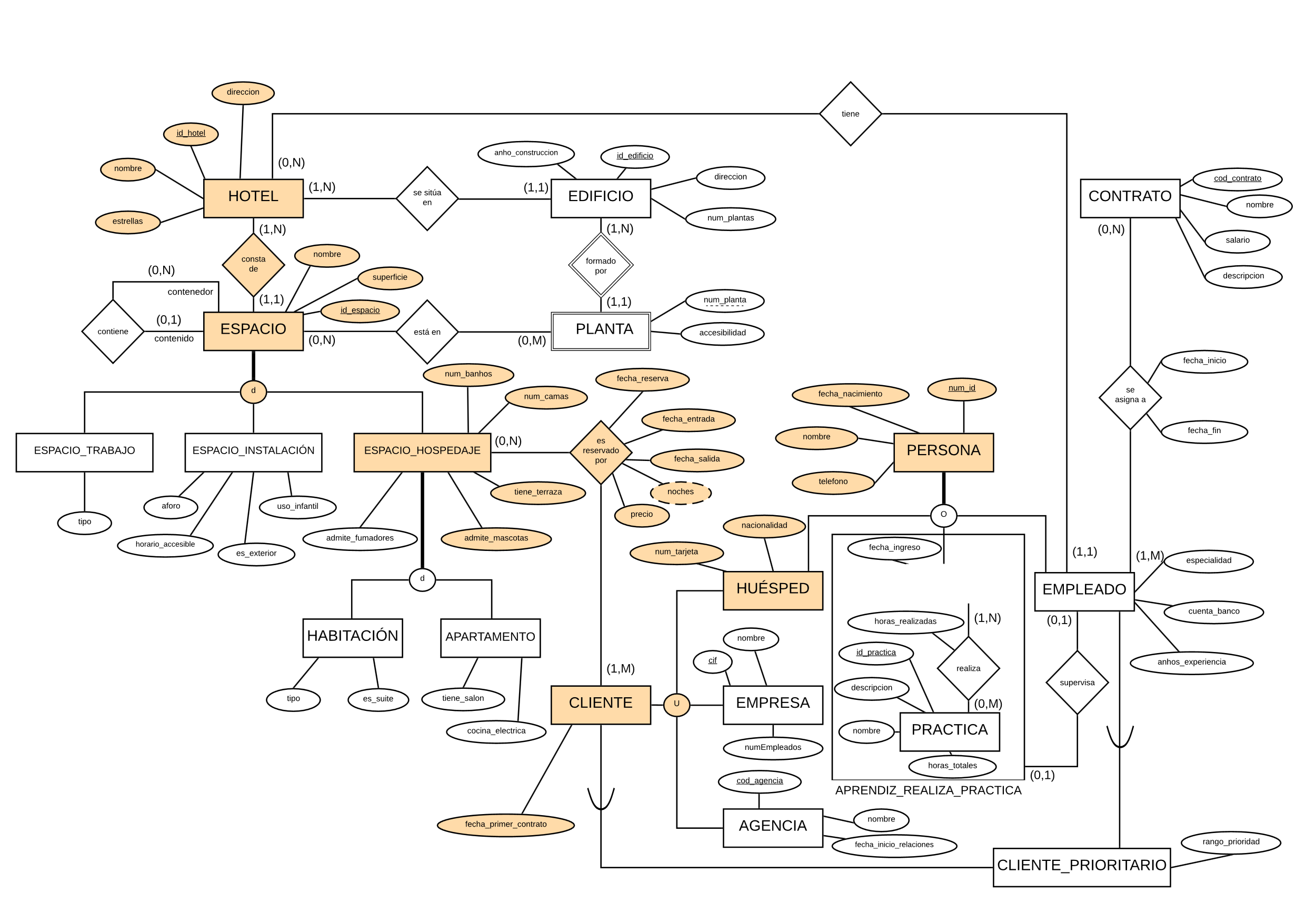
La cadena hotelera maneja información de distintos tipos de personas: huéspedes, empleados con contrato y aprendices. Estos últimos realizan prácticas que son supervisadas por empleados de la cadena. Además, en el caso de que alguien sea empleado y a la vez cliente, se le trata como cliente prioritario, por lo que tiene ciertas ventajas al utilizar los servicios de la cadena.

Los huéspedes no son el único tipo de clientes con los que la compañía mantiene relaciones, sino que esta también trabaja con agencias de viajes y otras empresas de todo tipo. De esta manera, cuando una empresa, agencia o huésped quiere contratar los servicios de la cadena, ésta los trata a todos por igual como clientes. Estos realizan reservas de los espacios de hospedaje que ofrece la cadena dentro de sus hoteles para una fecha de entrada y salida concretas.

# Modelo E/R original del sistema OLTP preexistente

En un primer momento, la compañía hotelera diseñó e implementó un sistema OLTP para registrar sus operaciones diarias. Este sistema queda descrito por el siguiente modelo E/R, en el que se indican las entidades, atributos y relaciones que modelan la actividad de la organización.

Sobre el diagrama mostrado se ha coloreado el área que será de interés para la creación del nuevo sistema.



***Figura 1.*** *Diagrama del modelo E/R correspondiente al sistema OLTP preexistente.*

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

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

A continuación, se muestran los nuevos atributos que han sido añadidos a las relaciones señaladas como interesantes en la Figura 1. Además, se da una justificación de por qué se considera necesario para el nuevo sistema de mercado de datos.

Para la relación **HOTEL**:

|  |  |  |
| --- | --- | --- |
| **Atributo(s)** | **Tipo** | **Justificación** |
| PAIS y CIUDAD | varchar(32) | Interesa conocer cómo se desarrolla la actividad de los hoteles de diferentes ciudades y países. |
| PARKING | enum('ABIERTO',  'CERRADO') | Se podrá ver cómo influye el tipo de aparcamiento sobre el rendimiento de los hoteles. |
| TRASLADO\_  AEROPUERTO,  TIENE\_ASCENSOR,  TIENE\_SH | enum('Sí', 'No') | Permitirá comprobar si contar con servicios como traslado al aeropuerto, presencia de ascensores o servicio de habitaciones influye sobre la actividad de los hoteles. |
| VALORACION | decimal(2,1) | Podrá mostrar la relación entre las características del hotel y las opiniones de los clientes. |
| TIPO | enum('HISTORICO',  'CONCEPTUAL',  'MODERNO',  'TRADICIONAL') | Se podrá ver cómo influye el tipo de hotel sobre el rendimiento de los hoteles con relación a factores como qué tipos de cliente los visitan, en qué épocas... |
| PRESUPUESTO, PRES\_MARKETING e INGRESOS | decimal(11,2) | Se podrá ver cómo influyen el presupuesto total, el presupuesto dedicado a marketing y los ingresos generados de los hoteles dadas sus características, las características de los clientes que los visitan, etc. |
| DISTANCIA\_CENTRO, DISTANCIA\_PLAYA | enum('< 500m', '< 1Km', '< 2Km', '< 3Km', '< 5Km') | Permitirá comprobar cómo influyen las distancias a lugares de interés como playas o el centro de la ciudad sobre la actividad de los hoteles. |

Para la relación **ESPACIO**:

|  |  |  |
| --- | --- | --- |
| **Atributo(s)** | **Tipo** | **Justificación** |
| TIENE\_INTERNET, TIENE\_VISTAS, ESTA\_ACLIMATIZADO | enum('Sí', 'No') | Interesa saber cómo influyen características del espacio sobre su rendimiento de cara a reservas, uso por parte de los clientes… |
| TIPO\_ESPACIO | enum('ESPACIO\_  HOSPEDAJE', 'ESPACIO  \_INSTALACION', 'ESPACIO  \_TRABAJO’) | Permitirá facilitar el trabajo para la carga de información en el datamart cargando ´la información sólo del tipo que nos interese (ESPACIO\_HOSPEDAJE). |

Para la relación **ESPACIO\_HOSPEDAJE**:

|  |  |  |
| --- | --- | --- |
| **Atributo(s)** | **Tipo** | **Justificación** |
| ADMITE\_FUMADORES,  CANCELACION\_GRATIS,  MOVILIDAD\_ADAPTADA | enum('Sí', 'No') | Interesa saber cómo influyen características del espacio sobre su rendimiento de cara a reservas, uso por parte de los clientes… |
| TIPO\_HOSPEDAJE | enum('APARTAMENTO', 'HABITACION') | Se podrá ver la relación entre el tipo de hospedaje y los perfiles de clientes que los reservan. |
| PRECIO\_NOCHE | decimal(5,2) | Permitirá relacionar categorizar a los alojamientos por precio y relacionarlos con los perfiles de cliente que los reservan. |
| LUMINOSIDAD | enum('BAJA', 'MEDIA', 'ALTA') | Se podrá ver cómo afecta el tipo de luminosidad al a las reservas de los espacios de hospedaje. |

Para la relación **PERSONA**:

|  |  |  |
| --- | --- | --- |
| **Atributo(s)** | **Tipo** | **Justificación** |
| GENERO | enum('Masculino', 'Femenino', 'Otro') | Se podrán realizar análisis clasificando a la gente por género. |

Para la relación **HUESPED**:

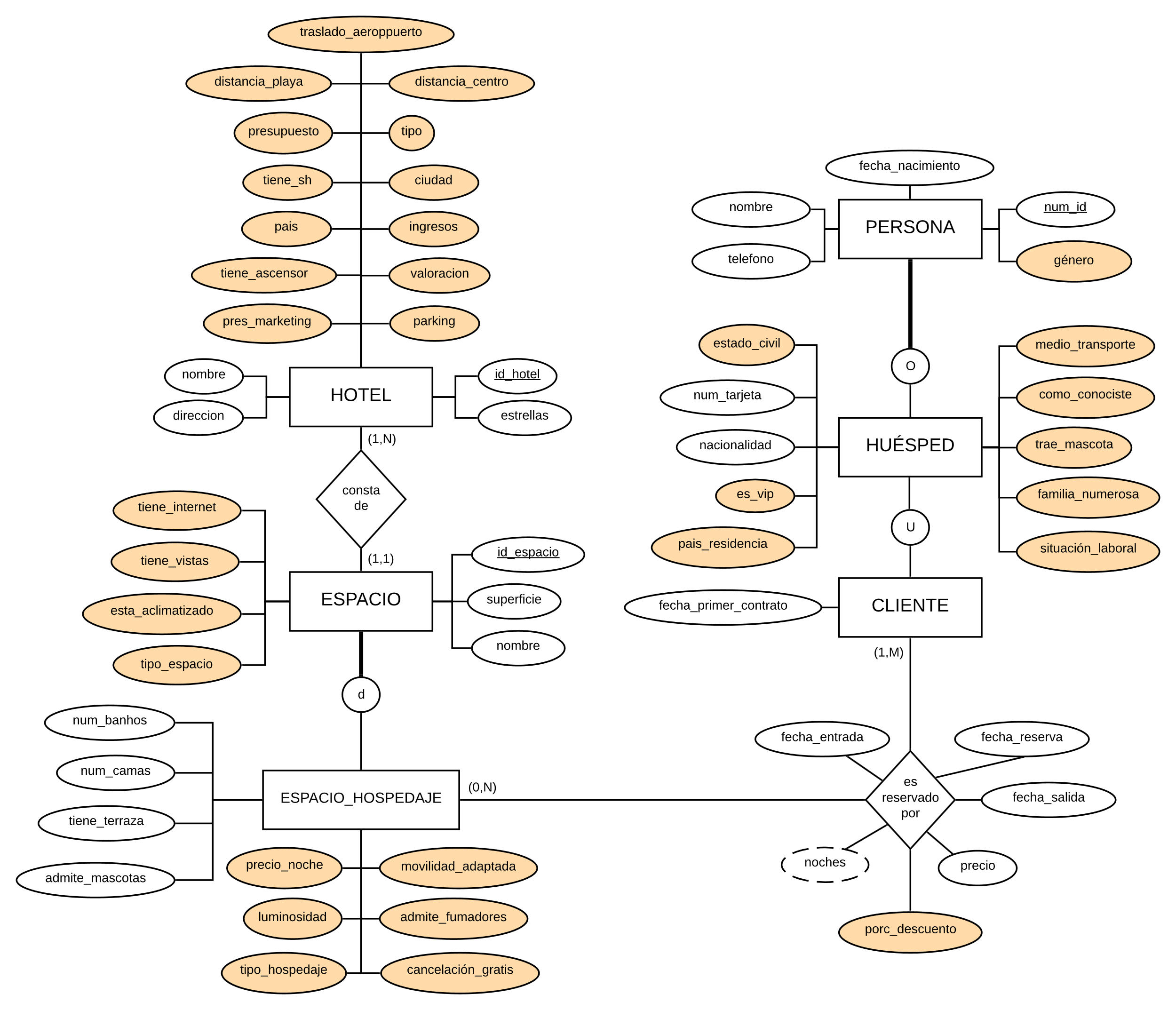
|  |  |  |
| --- | --- | --- |
| **Atributo(s)** | **Tipo** | **Justificación** |
| FAMILIA\_NUMEROSA | varchar(20) | Se podrán realizar análisis clasificando a la gente por género. |
| ESTADO\_CIVIL | |  |  | | --- | --- | |  | enum('Soltero', 'Casado', 'Viúdo', 'Divorciado') | | Se podrá ver la relación entre el estado civil de los huéspedes y sus comportamientos a la hora de hacer reservas. |
| SITUACION\_LABORAL | enum('Empresario','Autónomo',  'Asalariado', 'Cooperativista',  'En paro', 'Jubilado') | Información como la situación laboral o el medio de transporte en el que viajaron nos permitirán categorizar a los clientes, |
| MEDIO\_TRANSPORTE | enum('Coche','Avión','Tren',  'Moto','Bus','Camión') |
| COMO\_CONOCISTE | enum('Internet','Publicidad',  'Conocidos','Otros') | Permitirá ver qué medios de difusión son más efectivos en función de los tipos de cliente. |
| FAMILIA\_NUMEROSA, TRAE\_MASCOTA,  ES\_VIP | enum('Sí', 'No') | Se podrá ver cómo características como el tener familia numerosa, traer mascotas o ser vip influyen sobre las reservas de los clientes. |
| PAIS\_RESIDENCIA | varchar(40) | Se podrá ver cómo influye el país de residencia de los huéspedes sobre cómo hacen reservas. |

Para la relación **CLIENTE\_RESERVA\_ESP\_HOSPEDAJE**:

|  |  |  |
| --- | --- | --- |
| **Atributo(s)** | **Tipo** | **Justificación** |
| PORC\_DESCUENTO | decimal(2,0) | Permitirá saber cómo afectan los porcentajes de descuento a las reservas y relacionar esto con los clientes que las hacen |

Además de todos los atributos descritos anteriormente, se ha incorporado a cada tabla de entidad un nuevo campo ‘ULTIMA\_ACTUALIZACION’. Este es una marca de tiempo que nos permitirá llevar la información nueva (o actualizada) al nuevo sistema.

***1.3.2. Diagrama del modelo E/R con las entidades de interés para el mercado de datos y los nuevos atributos***

En el siguiente diagrama E/R se muestran solamente las entidades y relaciones de interés señaladas en la Figura 1, incluyendo los nuevos atributos descritos en el apartado anterior. Los nuevos atributos han sido coloreados para distinguirlos de los que ya formaban parte del E/R original.

***Figura 2.*** *Diagrama del modelo E/R ampliado.*

# Descripción de las fuentes externas

En las siguientes tablas, se presentan las fuentes externas que le serán incorporadas al nuevo sistema de mercado de datos. De cada una, se hace una breve descripción y se señalan los atributos específicos que se manejan.

|  |  |  |
| --- | --- | --- |
| **FUENTE** | **World Bank: Total Population** | |
| **Descripción** | World Bank ofrece un dataset que contiene la evolución poblacional de los diferentes países del mundo desde el año 1960 hasta la actualidad. El formato de los datos es XLS. | |
| **ATRIBUTO** | **TIPO** | **DESCRIPCIÓN** |
| País | String | El país al que corresponden los datos. |
| Año | Integer | Año al que corresponden los datos del país. |
| Población | Integer | La población del país para el año y país |

|  |  |  |
| --- | --- | --- |
| **FUENTE** | **World Bank: Countries classified by income group** | |
| **Descripción** | World Bank ofrece un dataset en el que agrupa a los diferentes países del mundo en función de su nivel de ingresos. Además, los agrupa a su vez según la región en la que se localicen. El formato de los datos es CSV. | |
| **ATRIBUTO** | **TIPO** | **DESCRIPCIÓN** |
| País | String | El país al que corresponden los datos. |
| Nivel de ingresos | Enum: bajo, medio-bajo, medio-alto, alto | Muestra el grupo de nivel de ingresos del país. |

|  |  |  |
| --- | --- | --- |
| **FUENTE** | **Kaggle: Average Temperature per country per year** | |
| **Descripción** | Kaggle ofrece un dataset en el que se muestran las temperaturas medias de los países del mundo en los últimos años. El formato de los datos es CSV. | |
| **ATRIBUTO** | **TIPO** | **DESCRIPCIÓN** |
| País | String | El país al que corresponden los datos. |
| Año | Integer | Año al que corresponde la temperatura media del país. |
| Temperatura | Float | Temperatura media del año para el país. |

# Descripción textual de la actividad a modelar

La actividad de la organización a modelar en el nuevo sistema de mercado de datos es la reserva de espacios de hospedaje en los diferentes hoteles realizadas por los huéspedes de la cadena.

# Selección de la granularidad

# Descripción del gránulo de la actividad a modelar y de cada una de las dimensiones

Se almacenan las **reservas diarias** que un **huésped** realiza para un **espacio de hospedaje** de un hotel. En la siguiente tabla se muestran los gránulos para cada una de las dimensiones del nuevo sistema de mercado de datos.

|  |  |  |
| --- | --- | --- |
| **Dimensión** | **Gránulo Menor** | **Gránulo Mayor** |
| Quién | Huésped | País de residencia |
| Qué | Espacio hospedaje | Espacio hospedaje |
| Cuándo | Día | Año |
| Dónde | Hotel | País |

# Razonar por qué no se ha optado por una granularidad mayor o menor

**¿Quién? Huésped y país de residencia**

El estudio con mayor nivel de detalle se centrará en analizar los datos personales de los huéspedes (edad, familia numerosa, género, etc.). De esta manera, se podrán hacer consultas para perfiles concretos de clientes.

El estudio más general se hará a nivel de país de residencia de los huéspedes, de forma que los podamos agrupar según su origen. Se analizarán características como el nivel de ingresos medio del país.

**¿Qué? Espacio hospedaje**

Se ha considerado que la dimensión ‘qué’ es ‘plana’, en el sentido de que interesa realizar estudios siempre al mismo nivel de detalle, esto es, **a nivel de espacio de hospedaje**. Agrupar o diluir la información de esta dimensión no proporciona un valor añadido para los futuros análisis. De este modo, se trabajará siempre con **características concretas de los alojamientos** como su tipo, su número de baños o camas o su superficie.

**¿Cuándo? Día y Año**

Puesto que para la toma de decisiones no es relevante conocer las horas exactas o momentos del día en los que se hacen las reservas, se ha llegado a la conclusión de que el mayor nivel de detalle deseado debe ser **‘día’**.

Siguiendo la misma filosofía, se ha elegido **‘año’** como gránulo mayor, puesto que el período en el que la empresa manejará los datos será de 10 años. Así, no tendría sentido escoger un gránulo como ‘lustro’, puesto que sólo habría dos intervalos de estudio.

**¿Dónde? Hotel y País**

El estudio con mayor nivel de detalle se centrará en analizar los datos de los **‘hoteles’** de la cadena (distancia a la playa/centro, estrellas, valoración, etc.). En estos, los huéspedes reservan espacios de hospedaje.

A nivel superior, se utilizará **‘país’** como gránulo mayor de manera que podamos agrupar a los hoteles a nivel internacional, permitiendo el estudio de tendencias en este ámbito. Se analizarán características como la población o la temperatura media del país.

# Identificación de las dimensiones

# Descripción textual de las dimensiones (a nivel general, sin detalle de los atributos)

**Dimensión ‘Dónde’**

Esta dimensión responde a la pregunta *‘¿Dónde se realizan las reservas?’*,englobando a las características de los hoteles de la cadena. Permitirá hacer estudios utilizando dichas características como filtros.

**Dimensión ‘Qué’**

La dimensión ‘Qué’ responde a la cuestión *‘¿Qué se reserva?’*,agrupando a las características de los espacios de hospedaje de los hoteles de la cadena. De esta manera, se podrán realizar análisis según los rasgos de los alojamientos reservados.

**Dimensión ‘Cuándo’**

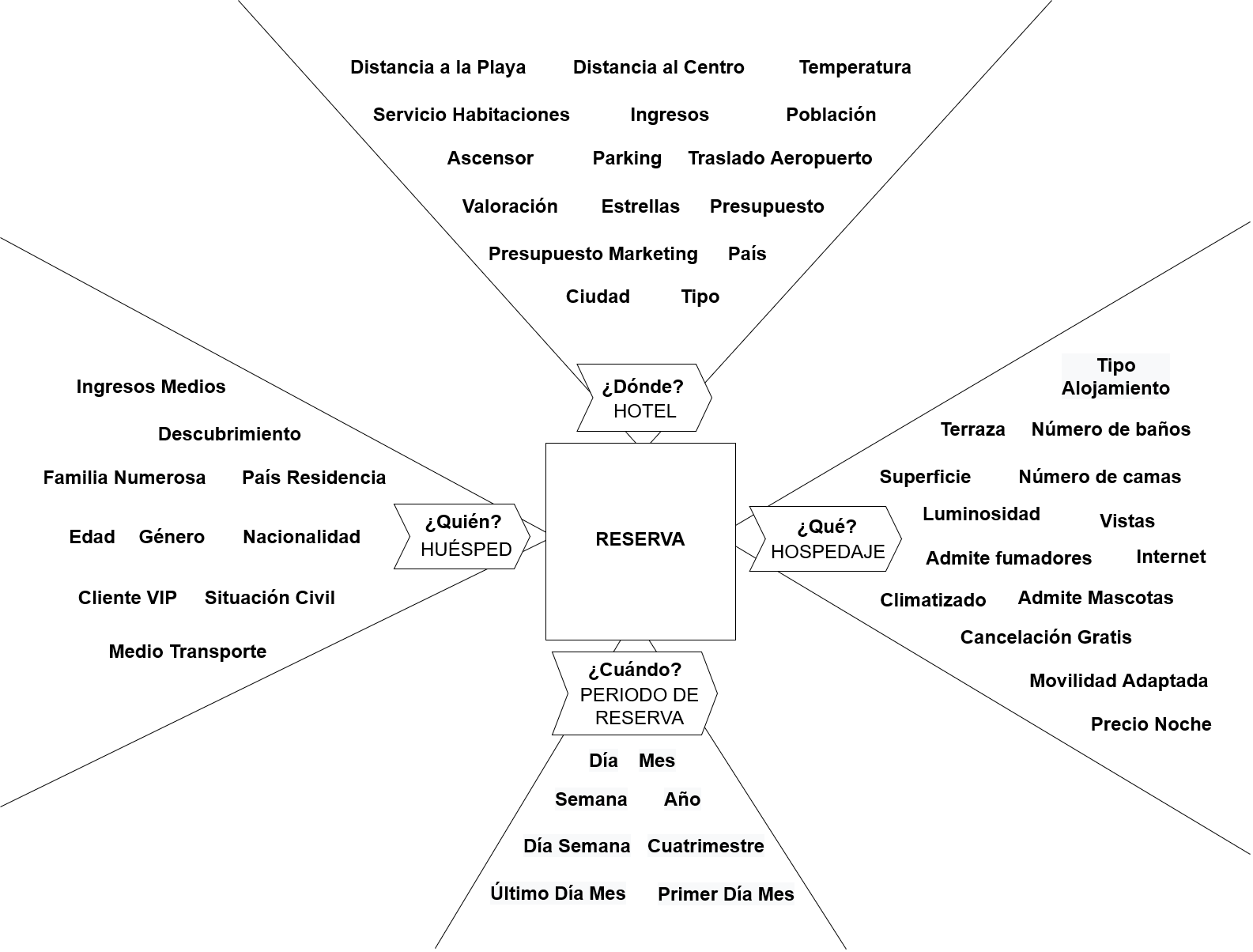
Esta dimensión responde a la pregunta *‘¿Qué período comprende la reserva?’*, agrupando las características de dicho periodo y representando dicha información con multitud de niveles de detalle.

**Dimensión ‘Quién’**

La dimensión ‘Quién’ responde a la cuestión *‘¿Quién realiza las reservas?’*, englobando las características de los huéspedes. De esta manera, se podrán hacer consultas en función de dichos rasgos.

# Diagrama de las dimensiones

A continuación, se muestra un diagrama en el que se muestran las dimensiones del datamart, tal y como fueron descritas en las secciones anteriores.

Además, se indican los atributos que han sido identificados para cada una de las dimensiones. Estos serán tomados tanto de los atributos del sistema OLTP preexistente (viejos y nuevos), como de las fuentes externas descritas anteriormente.

***Figura 3.*** *Diagrama de las dimensiones del datamart sin las métricas.*

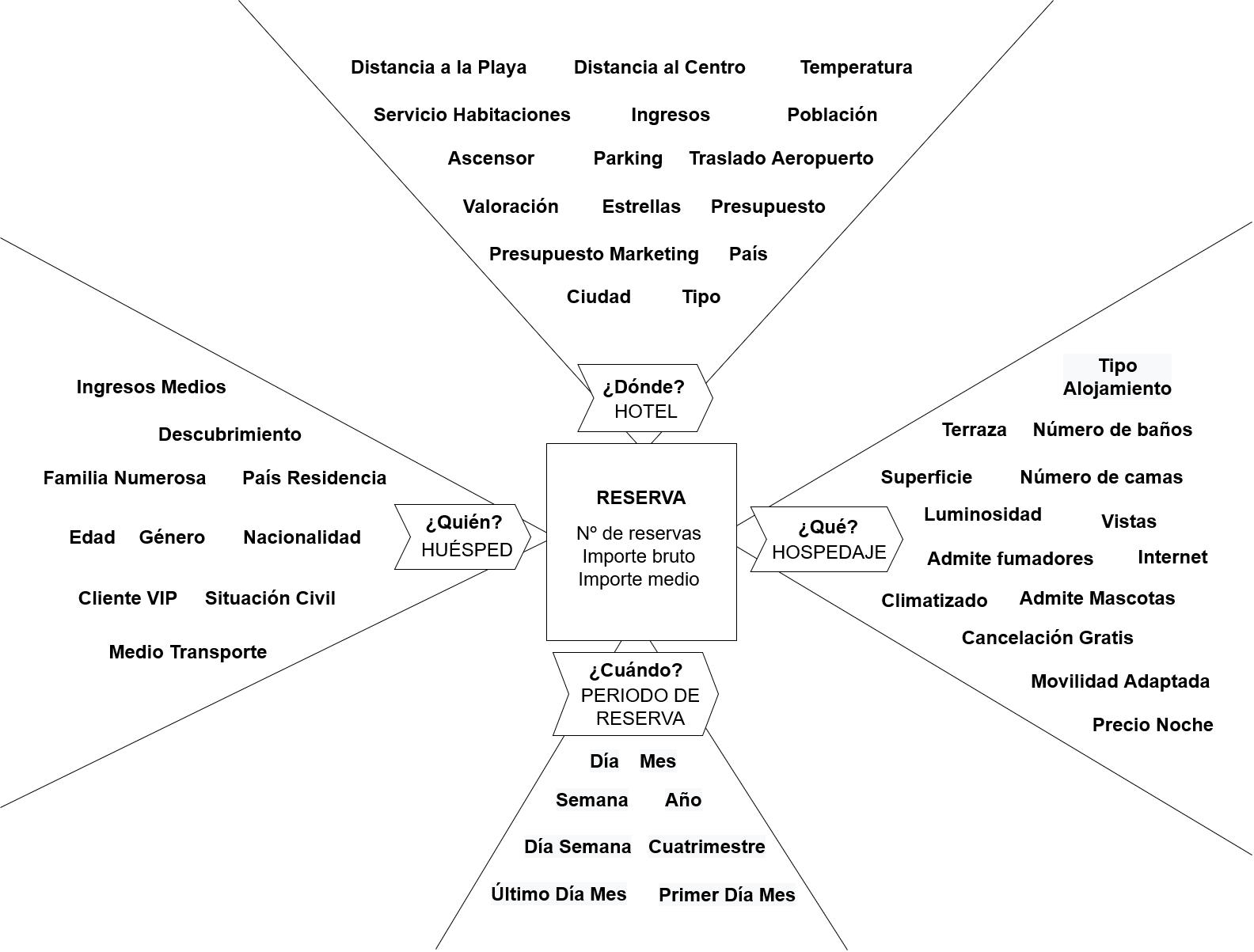
# Selección de las medidas asociadas al hecho

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

En la siguiente tabla se presentan las medidas que se emplearán para evaluar el rendimiento de la actividad modelada en el datamart.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MEDIDA** | **TIPO** | **DESCRIPCIÓN** | **TABLA/FUENTE**  **ORIGEN** | **ATRIBUTO/CAMPO**  **ORIGEN** |
| Número de reservas | *int* | Nº de reservas realizadas | CLIENTE\_ RESERVA\_  ESPACIO\_HOSPEDAJE | ID\_RESERVA |
| Importe bruto | *float* | Sumatorio de los importes de las reservas | CLIENTE\_ RESERVA\_  ESPACIO\_HOSPEDAJE | PRECIO |
| Importe medio | *float* | Media de los importes de las reservas | CLIENTE\_ RESERVA\_  ESPACIO\_HOSPEDAJE | PRECIO |

# Diagrama con la incorporación de las medidas



***Figura 4.*** *Diagrama de las dimensiones del datamart con las métricas.*

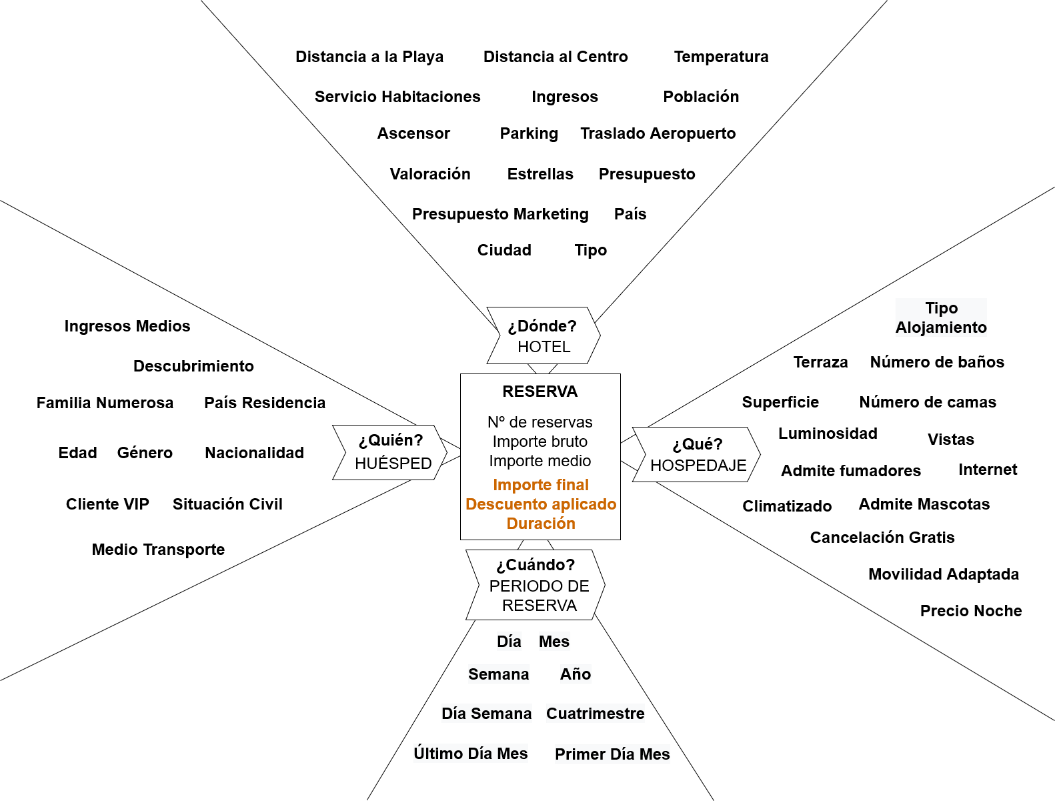
# Almacenamiento de valores precalculados en la tabla de hechos

# Definición de valores precalculados y descripción acerca de cómo se obtienen

A continuación, se presentan las medidas identificadas en el estudio realizado que emplean valores precalculados.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MEDIDA** | **TIPO** | **DESCRIPCIÓN** | **TABLA/FUENTE**  **ORIGEN** | **ATRIBUTO/CAMPO**  **ORIGEN** |
| Importe final | *float* | Sumatorio de los importes con descuento aplicado | CLIENTE\_ RESERVA\_  ESPACIO\_HOSPEDAJE | PRECIO \* (1 - PORC\_DESCUENTO \* 0.01) |
| Descuento Aplicado | *float* | Sumatorio de los descuentos aplicados | CLIENTE\_ RESERVA\_  ESPACIO\_HOSPEDAJE | PORC\_DESCUENTO  \* PRECIO \* 0.01 |
| Duración | *float* | Número medio de noches de las reservas | CLIENTE\_ RESERVA\_  ESPACIO\_HOSPEDAJE | NOCHES = FECHA\_ENTRADA – FECHA\_SALIDA |

# Diagrama con incorporación de los valores precalculados



***Figura 5.*** *Diagrama de las dimensiones del datamart con las métricas.*

# Terminación de las tablas de dimensión

# Descripción detallada de las dimensiones.

***6.1.1. Descripción de los atributos de cada dimensión***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DIMENSIÓN** | **QUÉ** | | | |
| **Descripción** | Responde a la cuestión *‘¿Qué se reserva?’*, agrupando a las características de los espacios de hospedaje de los hoteles de la cadena. De esta manera, se podrán realizar análisis según los rasgos de los alojamientos reservados. | | | |
| **ATRIBUTO** | **TIPO** | **DESCRIPCIÓN** | **TABLA/**  **FUENTE**  **ORIGEN** | **ATRIBUTO/**  **CAMPO**  **ORIGEN** |
| Nº de baños | *decimal(1,0)* | El nº de baños del alojamiento | Tabla: espacio\_  hospedaje | NUM\_  BANHOS |
| Nº de camas | *decimal(1,0)* | El nº de camas del alojamiento | Tabla: espacio\_  hospedaje | NUM\_  CAMAS |
| Tiene terraza | *enum('Sí', 'No')* | Describe si el alojamiento tiene terraza o no | Tabla: espacio\_  hospedaje | TIENE\_  TERRAZA |
| Admite fumadores | *enum('Sí', 'No')* | Describe si el alojamiento permite fumadores o no | Tabla: espacio\_  hospedaje | ADMITE\_  FUMADORES |
| Admite mascotas | *enum('Sí', 'No')* | Describe si el alojamiento permite mascotas o no | Tabla: espacio\_  hospedaje | ADMITE\_  MASCOTAS |
| Cancelación gratis | *enum('Sí', 'No')* | Describe si el alojamiento permite cancelación gratis o no | Tabla: espacio\_  hospedaje | CANCELACION\_  GRATIS |
| Movilidad adaptada | *enum('Sí', 'No')* | Describe si el alojamiento tiene movilidad adaptada o no | Tabla: espacio\_  hospedaje | MOVILIDAD\_  ADAPTADA |
| Precio Noche | *enum(‘X<100’,*  *’100<=X<200’,*  *’200<=X<300’,*  *’300<=X<400’,*  *’400<=X<500’,*  *‘X>=500’)* | Precio por noche del espacio de hospedaje | Tabla: espacio\_  hospedaje | PRECIO\_  NOCHE |
| Luminosidad | *enum('BAJA', 'MEDIA', 'ALTA')* | Índice de luminosidad del espacio | Tabla: espacio\_  hospedaje | LUMINOSIDAD |
| Tipo de alojamiento | *enum('APARTAMENTO', 'HABITACION')* | Distintivo de espacio entre Apartamento y Habitación | Tabla: espacio\_  hospedaje | TIPO\_  HOSPEDAJE |
| Superficie | *enum(‘Pequeño’,*  *’Mediano’,*  *‘Grande’)* | Superficie del espacio de la reserva | Tabla: espacio | SUPERFICIE |
| Tiene Internet | *enum('Sí', 'No')* | Describe si el espacio tiene internet o no | Tabla: espacio | TIENE\_  INTERNET |
| Vistas | *enum('Sí', 'No')* | Vistas paisajísticas del lugar | Tabla: espacio | TIENE\_  VISTAS |
| Climatizado | *enum('Sí', 'No')* | Indica si el espacio está climatizado | Tabla: espacio | ESTA\_  CLIMATIZADO |

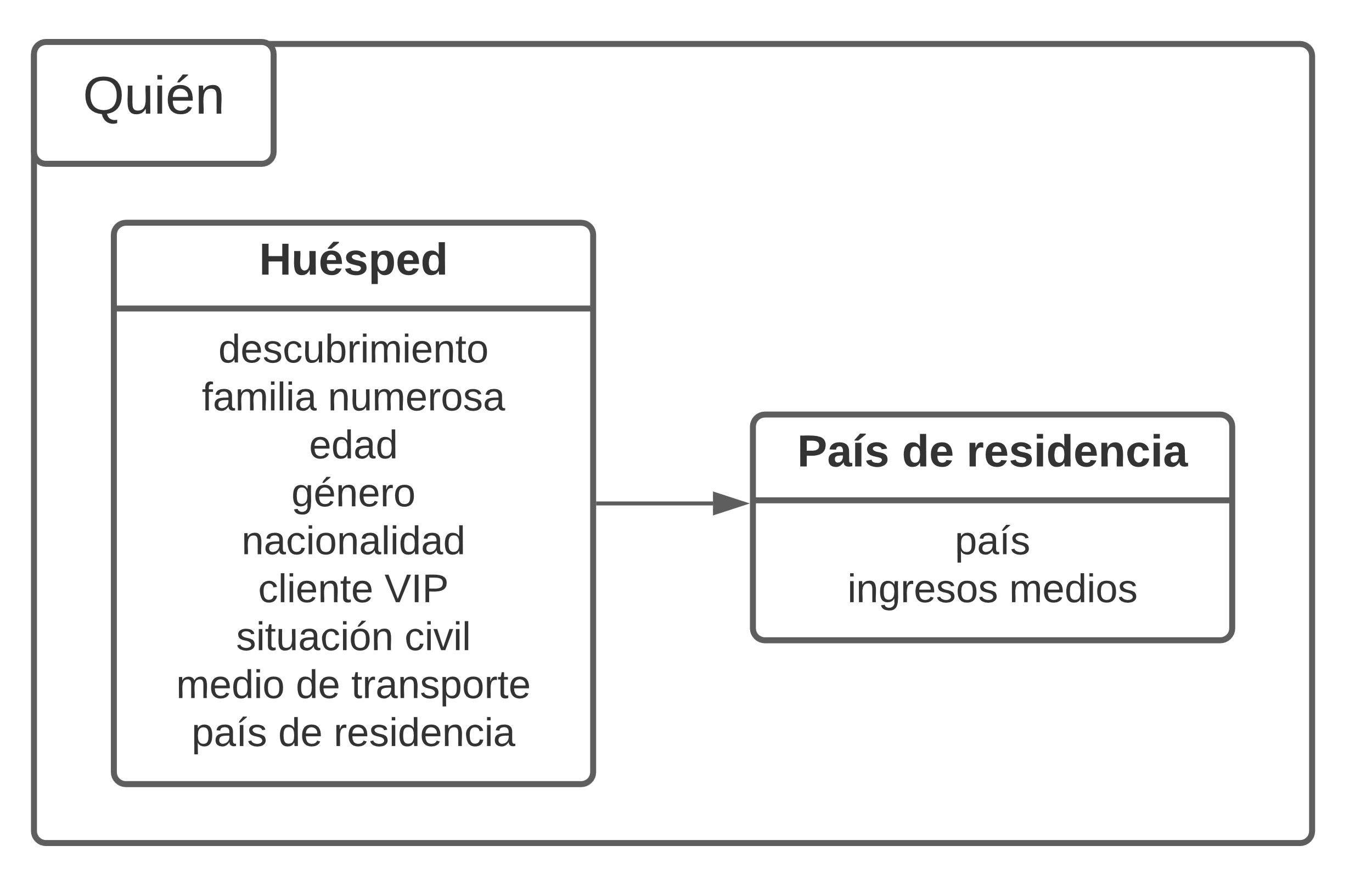
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DIMENSIÓN** | **CUÁNDO** | | | |
| **Descripción** | Responde a la pregunta *‘¿Qué período comprende la reserva?’*, agrupando las características de dicho periodo y representando dicha información con multitud de niveles de detalle. | | | |
| **ATRIBUTO** | **TIPO** | **DESCRIPCIÓN** | **TABLA/**  **FUENTE**  **ORIGEN** | **ATRIBUTO/**  **CAMPO**  **ORIGEN** |
| Día del mes | *TINYINT* | El día del mes de la fecha | - | - |
| Nombre  del día | *CHAR(12)* | El nombre del día de la semana | - | - |
| Primer día del mes | *CHAR(10)* | Si es el primer día del mes | - | - |
| Último día del mes | *CHAR(10)* | Si es el último día del mes | - | - |
| Mes | *TINYINT* | El mes del año del la fecha | - | - |
| Semana | *TINYINT* | La semana del año de la fecha | - | - |
| Cuatrimestre | *TINYINT* | El cuatrimestre del año de la fecha | - | - |
| Año | *SMALLINT* | El año de la fecha | - | - |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DIMENSIÓN** | **DÓNDE** | | | |
| **Descripción** | Esta dimensión responde a la pregunta *‘¿Dónde se realizan las reservas?’*, englobando a las características de los hoteles de la cadena. Permitirá hacer estudios utilizando dichas características como filtros. | | | |
| **ATRIBUTO** | **TIPO** | **DESCRIPCIÓN** | **TABLA/**  **FUENTE**  **ORIGEN** | **ATRIBUTO/**  **CAMPO**  **ORIGEN** |
| Distancia a la playa | *enum(‘N/A’, '< 500m', '< 1Km', '< 2Km', '< 3Km', '< 5Km')* | Distancia del hotel a la playa más cercana | Tabla: hotel | DISTANCIA\_  PLAYA |
| Distancia al centro | *enum(N/A’, '< 500m', '< 1Km', '< 2Km', '< 3Km', '< 5Km')* | Distancia del hotel al centro de la ciudad más cercana | Tabla: hotel | DISTANCIA\_  CENTRO |
| Temperatura | *enum(‘Clima Frío’,’Clima Templado’, ‘Clima Cálido’)* | Tipo de clima según temperatura media del país donde se encuentra el hotel | Fuente externa:  Temperatures.csv | [País, Año] |
| Población | *int* | Número de habitantes del país en el que se encuentre el hotel | Fuente externa:  Population.csv | [Año, País] |
| Servicio Habitaciones | *enum(‘Sí’, ‘No’)* | Indica si el hotel ofrece servicio de habitaciones | Tabla: hotel | TIENE\_SH |
| Ingresos | *decimal(11,2)* | Ingresos brutos del hotel en el último año fiscal | Tabla: hotel | INGRESOS |
| Valoración | *decimal(2,1)* | Valoración media de todos los clientes del hotel | Tabla: hotel | VALORACIÓN |
| Ascensor | *enum(‘Sí’, ‘No’)* | Indica si el hotel dispone de ascensor | Tabla: hotel | TIENE\_  ASCENSOR |
| Traslado Aeropuerto | *enum(‘Sí’, ‘No’)* | Indica el hotel posee un servicio de traslado entre el hotel y el aeropuerto | Tabla: hotel | TRASLADO\_  AEROPUERTO |
| Parking | *enum('ABIERTO’,*  *'CERRADO')* | Muestra si el aparcamiento del hotel es abierto o cerrado, o si, en su defecto, carece de aparcamiento propio | Tabla: hotel | PÁRKING |
| Presupuesto | *decimal(11,2)* | Presupuesto total utilizado por el hotel en el último año | Tabla: hotel | PRESUPUESTO |
| Presupuesto Marketing | *decimal(11,2)* | Presupuesto del hotel utilizado en sus campañas publicitarias | Tabla: hotel | PRES\_  MARKETING |
| País | *varchar(32)* | País en el cual se encuentra el hotel | Tabla: hotel | PAÍS |
| Estrellas | *decimal(1,0)* | Estrellas de las que dispone el hotel | Tabla: hotel | ESTRELLAS |
| Tipo de Hotel | *decimal(1,0)* | Tipo de hotel según estilo arquitectónico | Tabla: hotel | TIPO |
| Ciudad | *varchar(58)* | La ciudad en la que se encuentra el hotel | Tabla: hotel | CIUDAD |
| Activo | *enum(‘Sí’, ‘No’)* | Indica si el hotel sigue abierto al público | Tabla: hotel | ACTIVO |

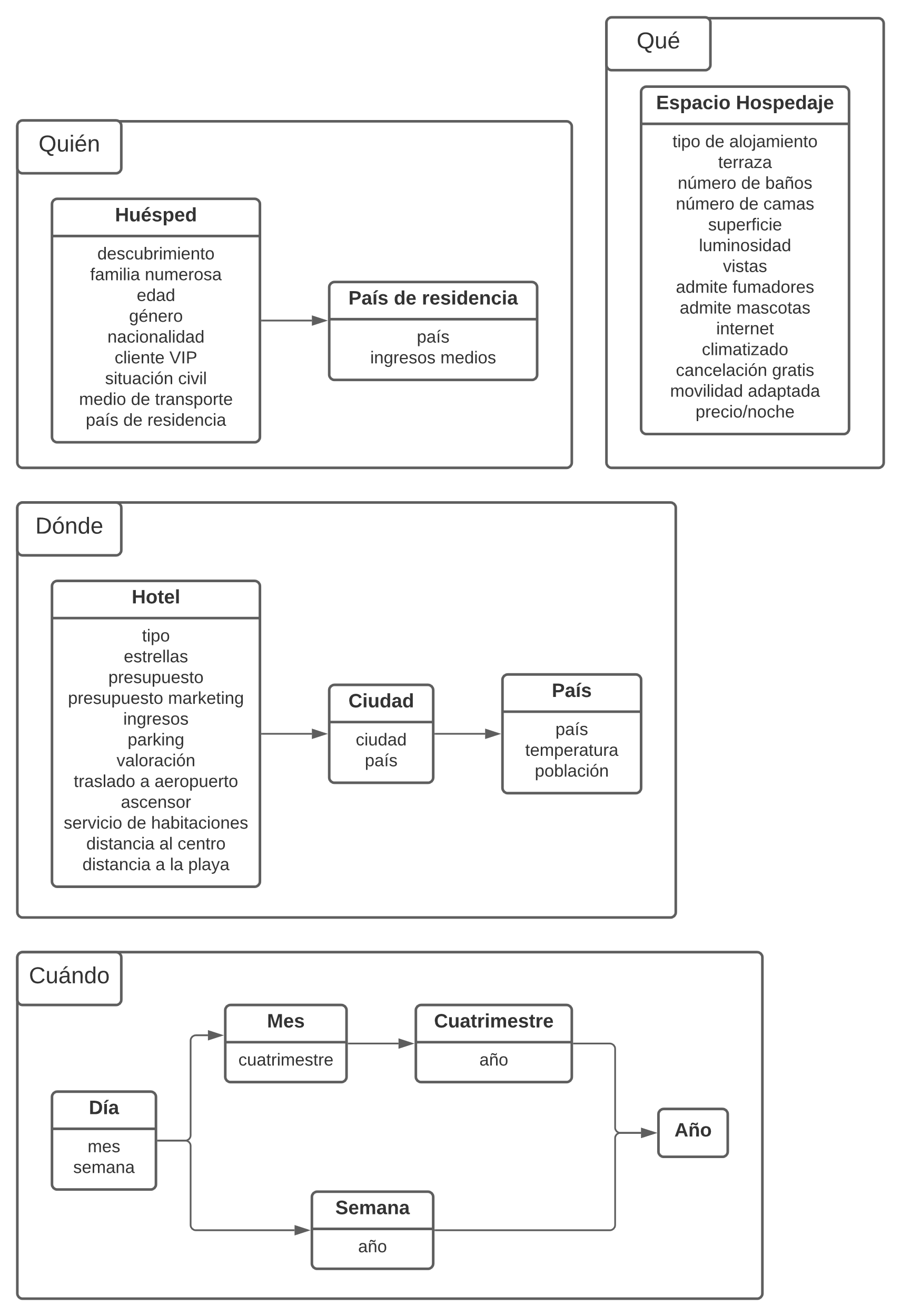
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DIMENSIÓN** | **QUIÉN** | | | |
| **Descripción** | La dimensión ‘Quién’ responde a la cuestión *‘¿Quién realiza las reservas?’*, englobando las características de los huéspedes. De esta manera, se podrán hacer consultas en función de dichos rasgos. | | | |
| **ATRIBUTO** | **TIPO** | **DESCRIPCIÓN** | **TABLA/**  **FUENTE**  **ORIGEN** | **ATRIBUTO/**  **CAMPO**  **ORIGEN** |
| Ingresos Medios | *Enum: bajo, medio-bajo, medio-alto, alto* | Renta per cápita en el país de residencia del cliente | Fuente externa:  Countries by income.csv | [Nivel de ingresos, País] |
| Descubrimiento | *enum('Internet','Publicidad','Conocidos','Otros')* | Medio a través del cual el cliente ha conocido esta empresa | Tabla: huésped | COMO\_  CONOCISTE |
| Familia Numerosa | *enum(‘Sí’,’No’)* | Indica si el cliente se encuentra en situación de familia numerosa | Tabla: huésped | FAMILIA\_  NUMEROSA |
| País residencia | *varchar(40)* | País en el cual reside el cliente | Tabla: huésped | PAIS\_  RESIDENCIA |
| Edad | *enum(‘< 25’,*  *‘25 <= X < 40 ‘,*  *‘40 <= X < 65 ‘,*  *‘>= 65’)* | Rango de edad del cliente | Tabla: persona | FECHA\_  NACIMIENTO |
| Género | *enum(‘Masculino’,’Femenino’,’Otro’)* | Género del cliente | Tabla: persona | GENERO |
| Nacionalidad | *varchar(20)* | País de origen del cliente | Tabla: huésped | NACIONALIDAD |
| Cliente VIP | *enum('Sí', 'No')* | Indica si el cliente disfruta de ventajas VIP | Tabla: huésped | ES\_VIP |
| Situación Civil | *decimal(1,0)* | Indica el estado civil del cliente | Tabla: huésped | ESTADO\_  CIVIL |
| Medio Transporte | *enum('Coche','Avión','Tren','Moto','Bus','Camión')* | Medio de transporte principal utilizado por el cliente para iniciar su estancia | Tabla: huésped | MEDIO\_  TRANSPORTE |
| Válido Desde | *Date* | Atributos empleados en el proceso de carga de la dimensión lentamente cambiante de tipo 2. | Estos atributos son generados por Kettle de forma automática durante el proceso de carga. | |
| Válido Hasta | *Date* |
| Número de Versión | *Smallint(5)* |

***6.1.2. Establecimiento de la jerarquía entre los atributos de cada dimensión***

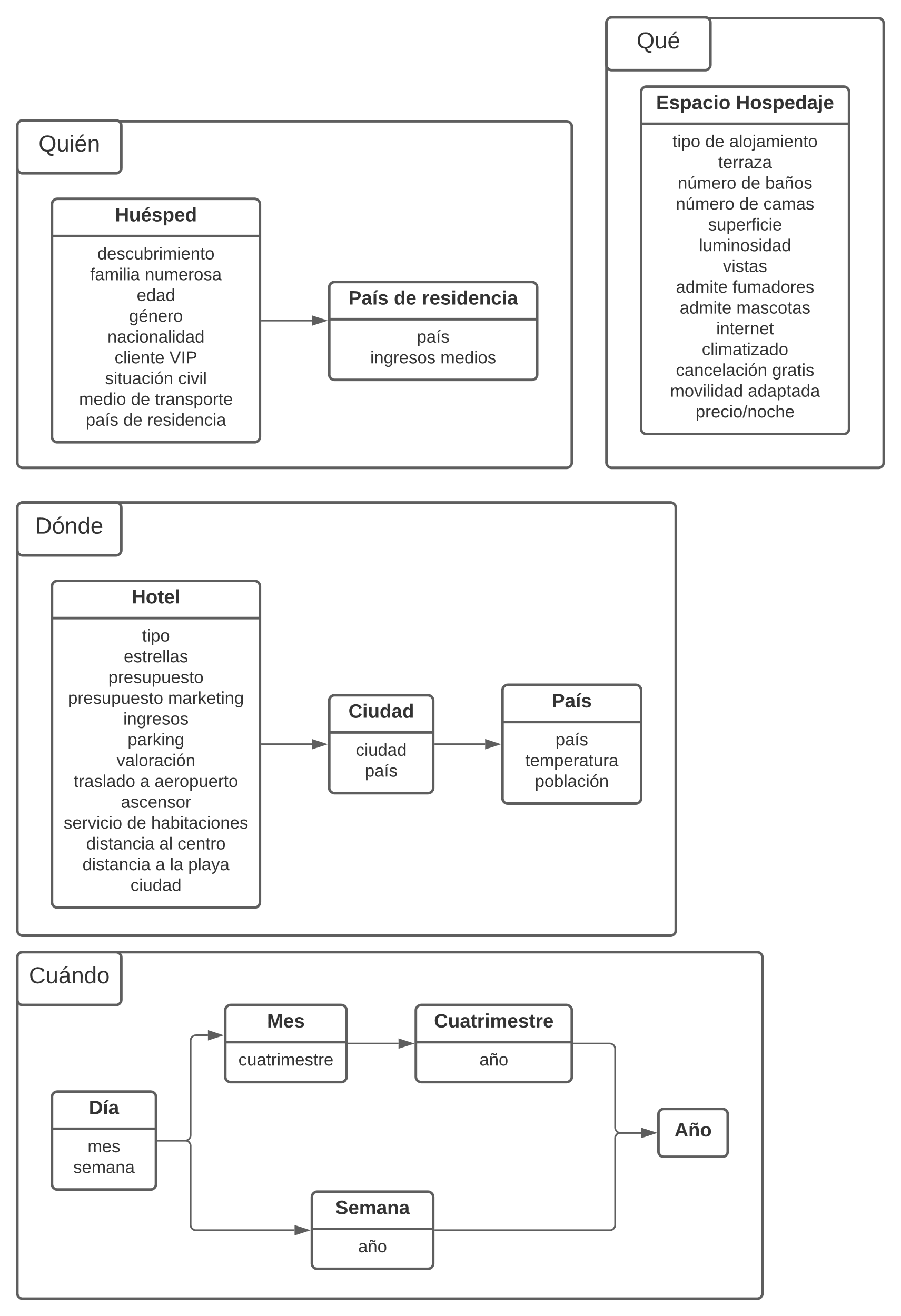
En las próximas figuras se describe se describe la jerarquía que seguirán los atributos de cada dimensión, tomando en cuenta las granularidades descritas en el paso 2 y los atributos explicados en la sección anterior.



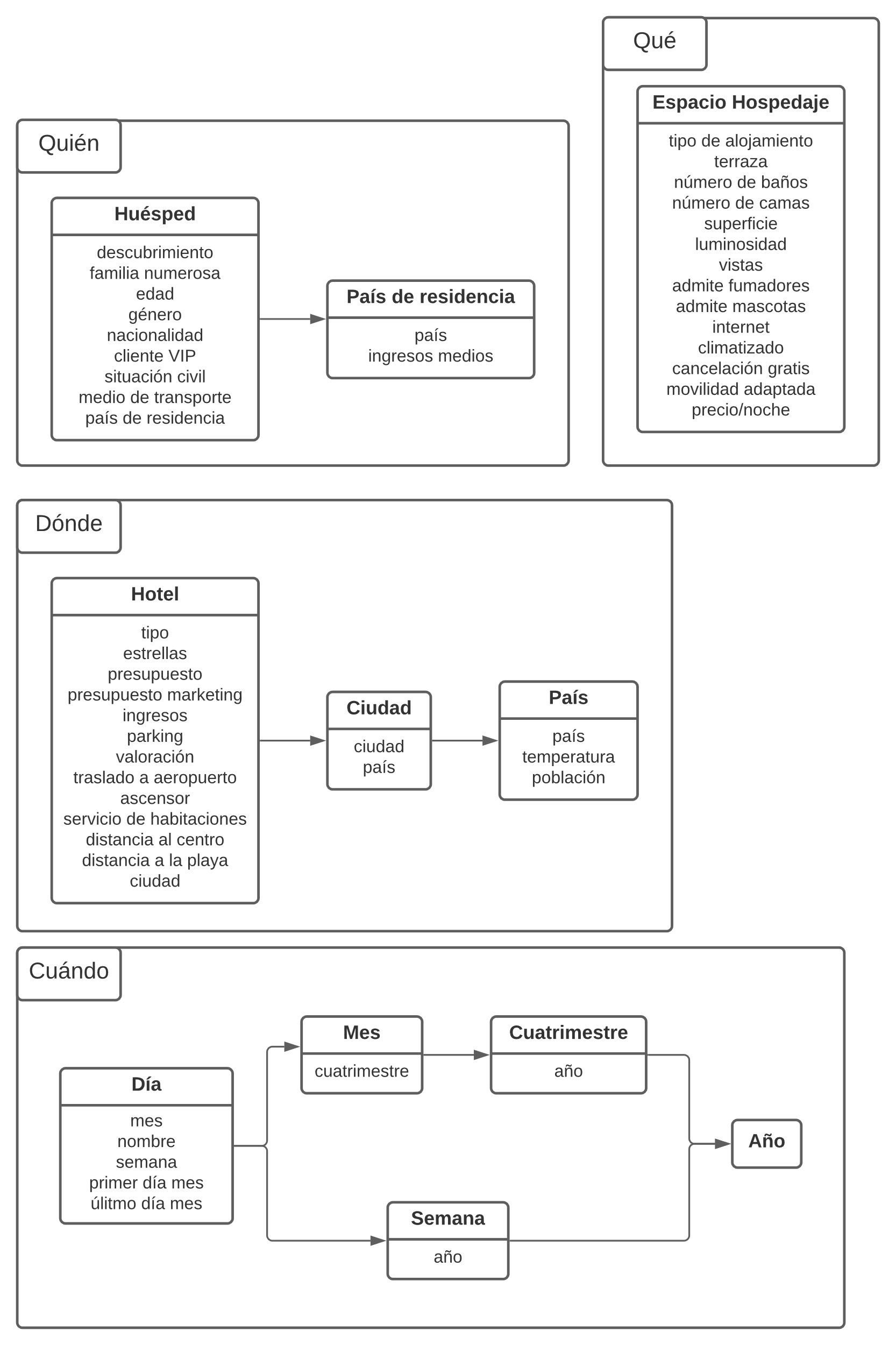
***Figura 6.1.*** *Diagrama descriptivo de la jerarquía de la dimensión ‘Quién’.*



***Figura 6.2.*** *Diagrama descriptivo de la jerarquía de la dimensión ‘Qué’.*



***Figura 6.3.*** *Diagrama descriptivo de la jerarquía de la dimensión ‘Dónde’.*



***Figura 6.4.*** *Diagrama descriptivo de la jerarquía de la dimensión ‘Cuándo’.*

***6.1.3. Conversión de datos: numerización/etiquetado, discretización***

En las siguientes tablas se hará una descripción de los procesos de conversión de datos para cada dimensión que se llevarán a cabo para su posterior carga en el nuevo sistema de datamart.

|  |  |  |
| --- | --- | --- |
| **DIMENSIÓN** | **QUÉ** | |
| **ATRIBUTO** | **TIPO DE CONVERSIÓN** | **DESCRIPCIÓN** |
| Precio Noche | Discretización | Se trabajará con los siguientes intervalos:   |  |  | | --- | --- | | * Menor que 100 * Entre 100 y 200 * Entre 200 y 300 | * Entre 300 y 400 * Entre 400 y 500 * Más de 500 | |
| Superficie | Discretización | Se trabajará con la siguiente clasificación:   * ‘Pequeño’, menos de 50 * ’Mediano’, de 50 a 100 * ’Grande’, más de 100 |

|  |  |  |
| --- | --- | --- |
| **DIMENSIÓN** | **DÓNDE** | |
| **ATRIBUTO** | **TIPO DE CONVERSIÓN** | **DESCRIPCIÓN** |
| Temperatura | Discretización | Se trabajará con la siguiente clasificación:   * ‘Clima frío’, menor de 10ºC * ‘Clima templado’, de 10ºC a 20ºC * ‘Clima cálido’, más de 20ºC |
| Tipo | Numerización | Se asignará a cada tipo los siguientes valores:   |  |  | | --- | --- | | * 0: Histórico * 1: Conceptual | * 2: Moderno * 3: Tradicional | |

|  |  |  |
| --- | --- | --- |
| **DIMENSIÓN** | **QUIÉN** | |
| **ATRIBUTO** | **TIPO DE CONVERSIÓN** | **DESCRIPCIÓN** |
| Edad | Discretización | Se trabajará con los intervalos siguientes:   |  |  | | --- | --- | | * Menos de 25 años * De 25 a 40 años | * De 40 a 65 años * Más de 65 años | |
| Situación civil | Numerización | Se asignarán los siguientes valores:   |  |  | | --- | --- | | * 0: Soltero * 1: Casado | * 2: Divorciado * 3: Viudo * 4: NS/NC | |

***6.1.4. Manejo de datos erróneos***

En las siguientes tablas se describirá cómo se manejarán aquellas situaciones en las que puedan aparecer datos erróneos.

|  |  |
| --- | --- |
| **DIMENSIÓN** | **QUÉ** |
| **ATRIBUTO** | **DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO** |
| Precio Noche | El precio por noche de los alojamientos debe ser superior a 10. En caso contrario, **se reemplazará por el valor mínimo permitido de 10**. |

|  |  |
| --- | --- |
| **DIMENSIÓN** | **DÓNDE** |
| **ATRIBUTO** | **DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO** |
| Ingresos | Los ingresos no pueden ser una cantidad negativa. De ser así, **se remplazará el valor por la media de los ingresos** en el sistema OLTP. |

|  |  |
| --- | --- |
| **DIMENSIÓN** | **QUIÉN** |
| **ATRIBUTO** | **DESCRIPCIÓN DEL TIPO DE ERROR Y CÓMO RESOLVERLO** |
| Edad | La edad no debe estar fuera del rango [18,120]. Si está fuera de ese rango**, se agrupará en el intervalo inferior (<25) y superior (>65)**, respectivamente. |
| Nacionalidad | Los caracteres del atributo nacionalidad debe de ser alfabéticos. En caso contrario, **se remplazará por el valor ‘NS/NC’.** |

***6.1.5. Manejo de datos faltantes***

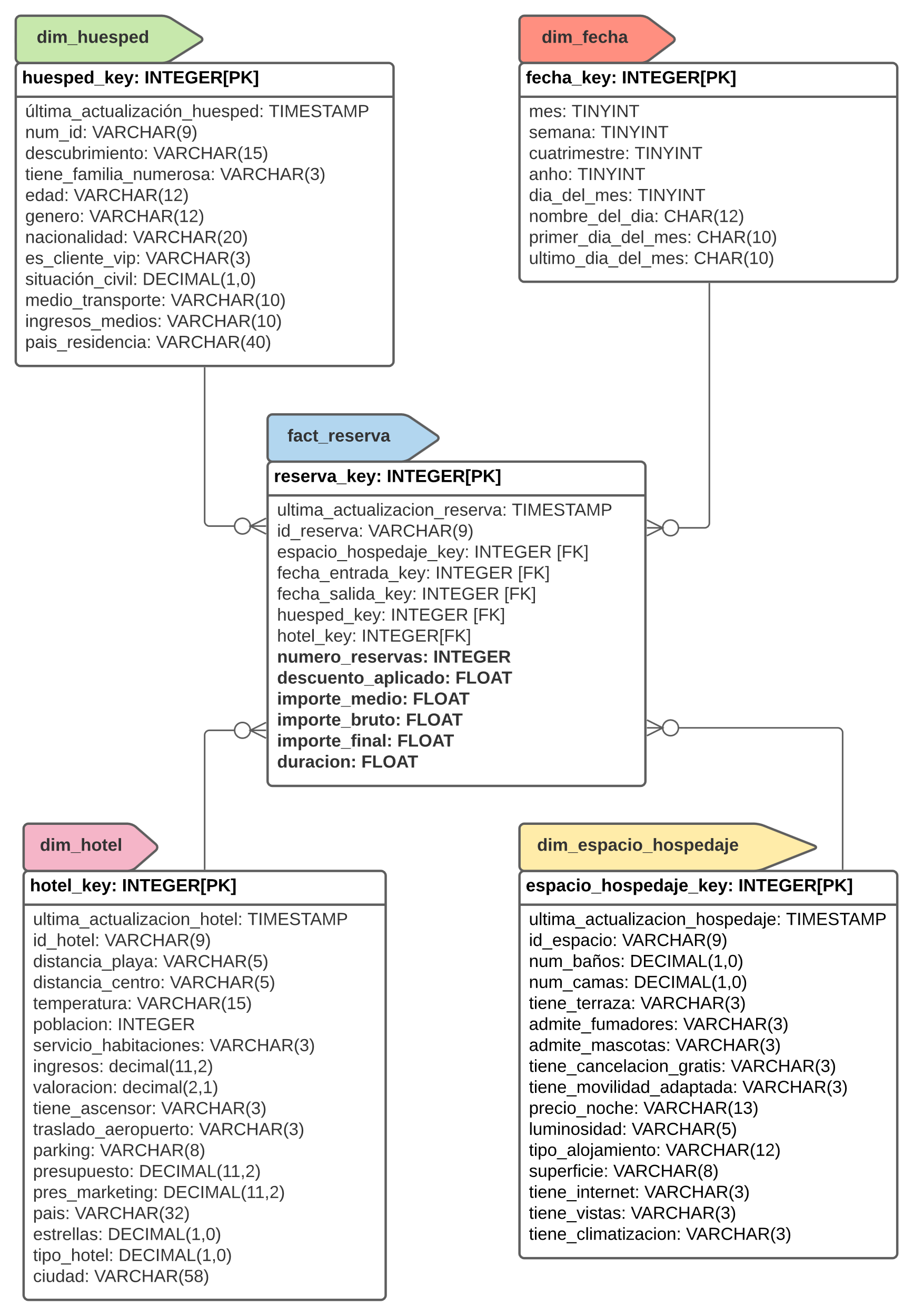
En las siguientes tablas se describirá cómo se manejarán aquellas situaciones en las que puedan aparecer datos faltantes.

|  |  |
| --- | --- |
| **DIMENSIÓN** | **QUÉ** |
| **ATRIBUTO** | **DESCRIPCIÓN DEL TIPO DE PROBLEMA**  **Y CÓMO RESOLVERLO** |
| Número de baños | El nº de baños puede ser NULL. En tal caso, **se reemplazará el valor por el valor medio** de este atributo en el sistema OLTP. |

|  |  |
| --- | --- |
| **DIMENSIÓN** | **DÓNDE** |
| **ATRIBUTO** | **DESCRIPCIÓN DEL TIPO DE PROBLEMA**  **Y CÓMO RESOLVERLO** |
| Distancia a la playa | La distancia a la playa puede ser NULL. En tal caso, **se sustituirá por el valor ‘N/A’**. |
| Distancia al centro | La distancia al centro puede ser NULL. En tal caso, **se sustituirá por el valor ‘N/A’**. |

|  |  |
| --- | --- |
| **DIMENSIÓN** | **QUIÉN** |
| **ATRIBUTO** | **DESCRIPCIÓN DEL TIPO DE PROBLEMA**  **Y CÓMO RESOLVERLO** |
| Situación civil | La situación civil del huésped puede faltar. En tal caso, se considerará como **un valor “NS/NC”**. |
| Ingresos medios | Si el dato no está disponible en la fuente externa, **se reemplazará por el valor por defecto “Medio-Bajo”** en la fuente externa. |

# Diagrama final de diseño de la BD Datawarehouse versión ROLAP (incorporando claves)



***Figura 7.*** *Diagrama ROLAP con las tablas de dimensión y hechos.*

# Selección de la duración de la base de datos

# Indicación de la duración elegida y justificación

La duración de los datos en nuestra base de datos se extenderá hasta un máximo de 10 años. Esto se debe a que, a partir de 10 años, la información de nuestras fuentes puede quedarse obsoleta, puesto que el turismo es un sector muy cambiante y sensible a tendencias que varían rápidamente en el tiempo.

Estas tendencias pueden incluso dejar de ser relevantes en periodos de dos o tres años, de modo que las tendencias de más de 10 años no afectan a nuestro estudio de mercado.

# Control de las dimensiones lentamente cambiantes

# Determinación de las dimensiones lentamente cambiantes y cuáles son los atributos afectados

Se ha llegado a la conclusión de que las siguientes dimensiones son cambiantes:

|  |  |
| --- | --- |
| **DIMENSIÓN** | **ATRIBUTOS AFECTADOS** |
| Quién | Edad  Situación civil |
| Qué | Internet |
| Dónde | Ascensor |

# Para cada uno de los atributos cambiantes indicar cómo será tratado (Tipo 1, 2 o 3)

A continuación, se describe de forma justificada cómo serán tratados los atributos indicados anteriormente.

|  |  |  |  |
| --- | --- | --- | --- |
| **DIMENSIÓN** | **ATRIBUTOS AFECTADOS** | **TIPO** | **JUSTIFICACIÓN** |
| Quién | Edad  Situación civil | Tipo 2 | Se quiere mantener la temporalidad de forma trazable para analizar los cambios de comportamiento. |
| Qué | Internet | Tipo 1 | Son atributos que pueden cambiar, pero cuyo cambio no es interesante. |
| Dónde | Ascensor |

# Modos de consulta y prioridades de ejecución

# Descripción textual de las consultas OLAP

|  |  |
| --- | --- |
| **MEDIDAS - Cuál fue…** | |
| * El número de reservas * El importe total * El importe final | * El importe medio, * El descuento aplicado * La duración de las reservas |

|  |  |  |  |
| --- | --- | --- | --- |
| **QUIÉN** | **DÓNDE** | **QUÉ** | **CUÁNDO** |
| **“Los huéspedes que…”** | **“Visitan un hotel…”** | **“Reservan un alojamiento con…”** | **“Cuyo periodo de reserva…”** |
| Son del género X | Localizado en una ciudad X | Precio noche en el intervalo X | En el día X |
| Conocieron a la organización a través de X | Con X valoración | Cancelación gratis | Con nombre  de día X |
| Proceden de un país X | Localizado en un país con temperatura de clima X | Con climatización incluida | Siendo primer día del mes |
| Tiene una edad de X a Y | Con distancia a la playa menor que X | Con tipo de alojamiento X | En el mes X |
| Proceden de países con ingresos medios X | Con X estrellas | Vistas | En el cuatrimestre X |
| Tienen nacionalidad X | Con tipo X | Posibilidad de admitir fumadores | En el año X |