

Adquisición y preparación de datos

Ejercicio práctico – Compañía aérea.



Este ejercicio práctico de la asignatura se centrará en el diseño y la construcción de un almacén de datos a partir de diferentes fuentes con el objetivo de comprender como se estructuran y preparan los datos para este fin. Para ello nos basaremos en el caso de estudio de una compañía aérea. La práctica se realizará utilizando herramientas como Microsoft Visio o Draw.io para el diseño conceptual y MySQL Workbench para el diseño lógico y físico del almacén de datos,

La práctica está dividida en cinco partes con el objetivo el objetivo de explorar los diferentes enfoques para la creación de un almacén de datos, tomando como caso de estudio el diseño de un almacén de datos para un aeropuerto.

Parte 1: Diseño Conceptual a partir de la Especificación

En esta primera parte, el objetivo es crear un modelo conceptual del almacén de datos del aeropuerto basándose en una especificación proporcionada. Se tratará de identificar el hecho, su granularidad, las dimensiones (y niveles de esta), medidas y relaciones para poder representar correctamente los datos necesarios para el análisis y la toma de decisiones en el aeropuerto. En concreto el departamento de marketing nos ha entregado la siguiente especificación:

El departamento de marketing de una compañía aérea quiere analizar la actividad de los vuelos de cada miembro de su programa de viajeros frecuentes. Este departamento está interesado en analizar qué vuelos eligen los pasajeros frecuentes de la compañía, con qué aviones viajan, qué tarifa pagan, cómo ganan y canjean sus puntos, cuál es la duración de sus estancias en destino, etc.

Solo nos interesa la gestión de vuelos realizados por pasajeros frecuentes de una compañía aérea. No nos centramos en la reserva de vuelos que no resulte en el embarque de un pasajero frecuente en el avión.

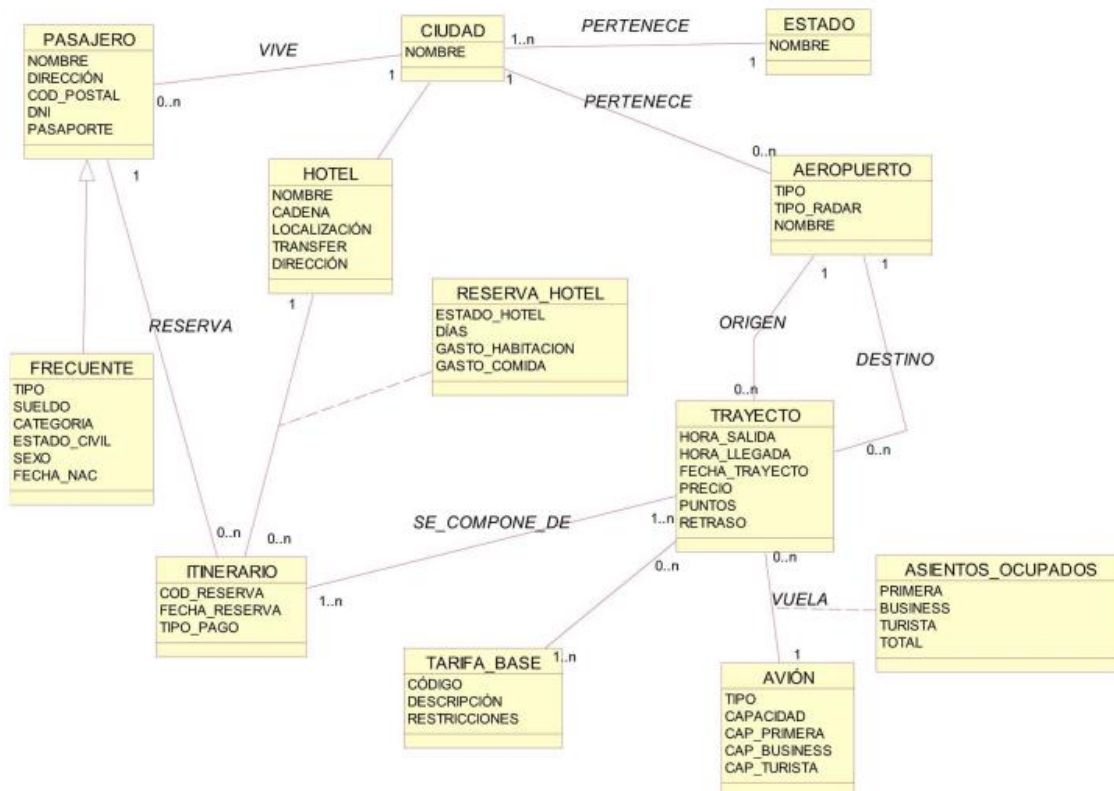
Para ello, la compañía aérea captura datos a nivel de trayectos. Un trayecto representa un avión que despegue en un aeropuerto y que aterriza en otro aeropuerto sin paradas intermedias. De cada trayecto se conoce información acerca de la fecha de realización, la hora de salida, la hora de

llegada, el aeropuerto origen y destino, el precio base del trayecto, los puntos que se obtienen (o se gastan) al realizar el trayecto y el retraso acumulado por el trayecto. También se conoce el avión que realiza el trayecto y la tarifa base, así como los asientos disponibles de cada tipo en cada vuelo.

Realiza el diseño conceptual del almacén de datos basado en la especificación entregada por el departamento de marketing.

Parte 2: Diseño Conceptual a partir de Fuentes Transaccionales

En esta parte, el departamento de marketing nos ha enviado el diseño de sus fuentes transaccionales que contienen datos operativos y detallados que ha almacenado el departamento de marketing, el cual se presenta a continuación:



A partir del diseño de las fuentes transaccionales, realiza el modelo conceptual para el almacén de datos.

NOTA 1: Debes de considerar que el objetivo del departamento de Marketing es el mismo “Este departamento está interesado en analizar qué vuelos eligen los pasajeros frecuentes de la compañía, con qué aviones viajan...”.

NOTA 2: A pesar de que el objetivo es el mismo, debes de realizar este ejercicio teniendo en cuenta **UNICAMENTE** el diseño de las fuentes transaccionales proporcionado en esta parte de la práctica para el diseño del almacén de datos. Por lo que pueden aparecer o desaparecer dimensiones, niveles o medidas con respecto a la parte 1 de la práctica.

Parte 3: Diseño Conceptual a partir de las Fuentes de Datos

El departamento de marketing de la compañía aérea finalmente nos ha enviado unos conjuntos de datos extraídos de sus fuentes transaccionales que utilizan y tienen acceso. Para la realización de esta parte de la práctica, se utilizarán los ficheros proporcionados para identificar y modelar las fuentes de datos que pueden ser importadas al almacén. El desafío es comprender la estructura de los datos proporcionados y traducir esta información en un modelo conceptual que se pueda integrar con otras fuentes de datos.

Realiza el diseño conceptual del almacén de datos a partir de la información contenida en el fichero “datos aeropuerto.zip”.

NOTA 1: A pesar de la similitud con otras partes de la práctica, debes de utilizar **UNICAMENTE** la información proporcionada en el fichero “datos aeropuerto.zip” para el diseño del almacén.

Parte 4: Análisis de Inconsistencias en los Datos y Propuestas de Solución

El objetivo de esta parte es identificar al menos 3 inconsistencias y/o problemas en los datos proporcionados para la construcción del almacén y proponer cómo podrían resolverse (o que haría falta para poder resolverse). Aunque no es necesario resolver las inconsistencias explícitamente, tan solo debéis de analizar de manera crítica los datos y presentar posibles soluciones para mejorar su calidad y coherencia.

Parte 5: Diseño Lógico y Físico del Almacén de Datos

En esta ultima parte de la práctica, debéis de realizar el diseño lógico y físico del almacén de datos utilizando MySQL Workbench. Para ello, debes de utilizar el diseño conceptual **basado en las fuentes de datos (parte 3)**. Ten en cuenta, que a pesar de que partimos del conceptual, en este diseño se presenta información diferente (¡y también de manera diferente!). Una vez realizado el diseño lógico, debes de crear un esquema físico que incluya las tablas, atributos, relaciones, y restricciones necesarias para almacenar de manera eficiente los datos en una base de datos.

