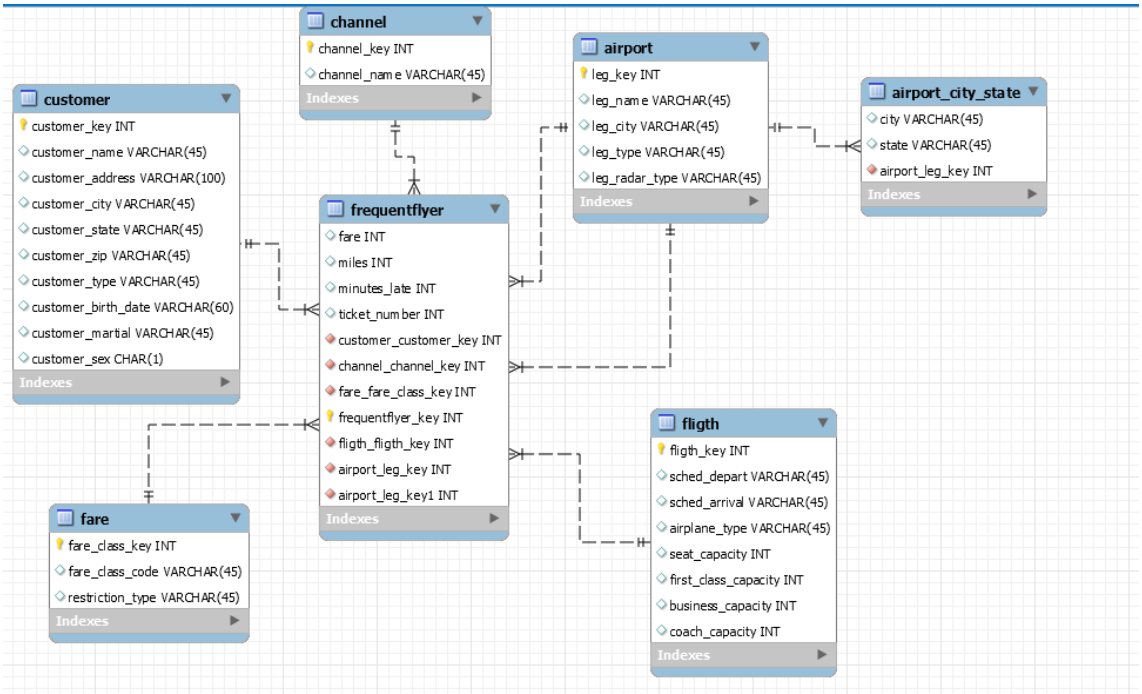
Task 1. Multidimensional design

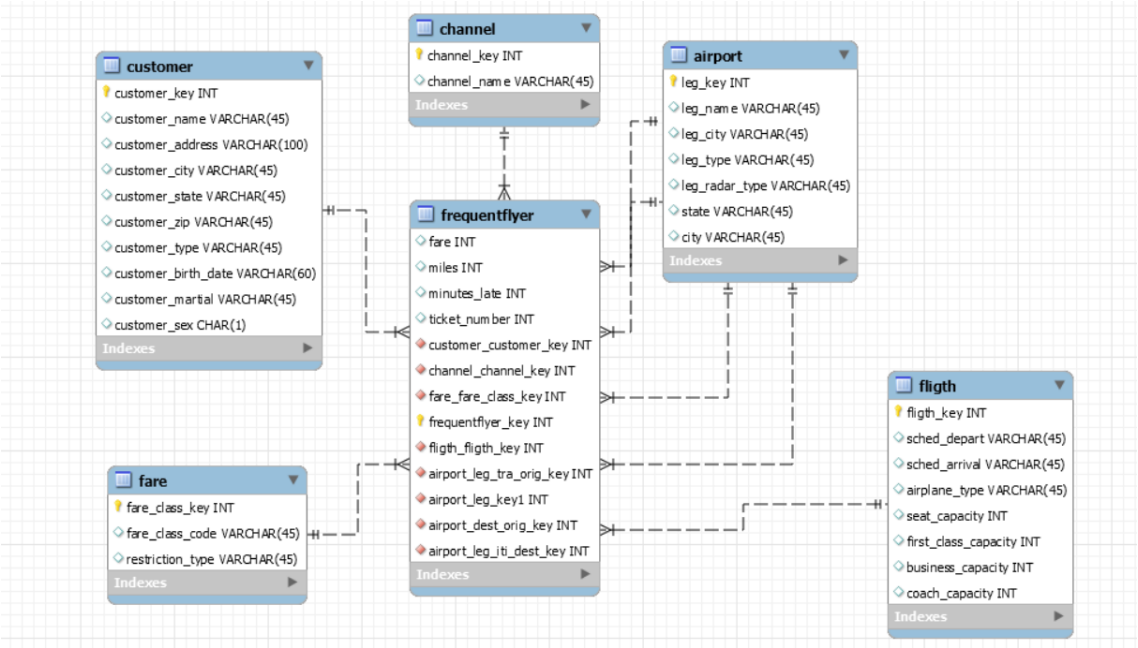
In this task I am going to design the multidimensional model for the specifications given in the first exercise. I am going to do the design based on the client specifications and the CSV available for the analysis.

The client wants to analyze the activity of the flight of each frequent flyer. They are interested in analyzing which flight the frequent passengers of the company choose, what airplanes they use to travel, what rates they pay…

Given this information and the tables we have we could think that the right schema to use could be a fact constellation scheme like this one.



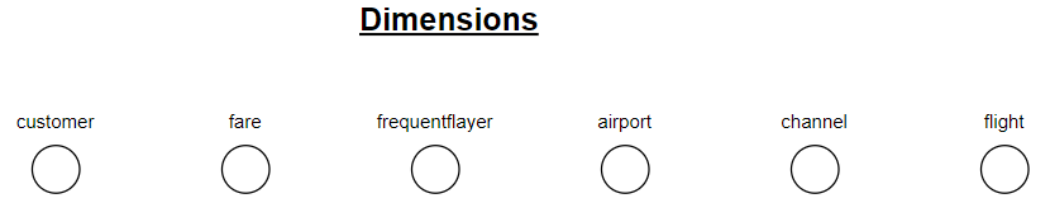
This scheme could be a good scheme to do our analysis but in my point of view I prefer to insert the airpor_city_state table inside the airport table and make a star scheme like the following one.



I am going to use this scheme because I am fonder of the star scheme model and I find it easier to work with.

DIMENSIONS

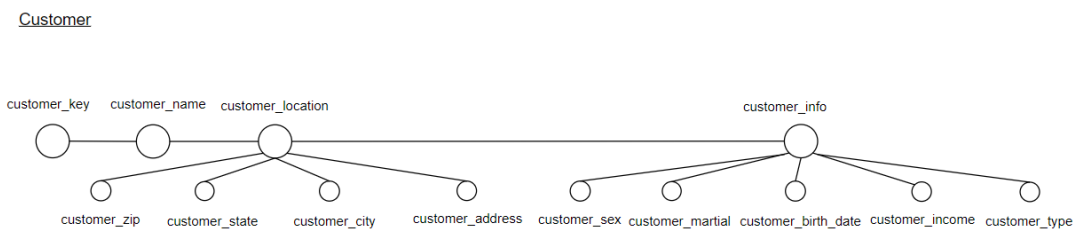
When it comes to the dimensions of the multidimensional scheme, I have chosen to do one dimension for each table in the Airplane dataset. I think this is the most effective and easiest way to embrace the problem.



HIERARCHY LEVELS

-Customer

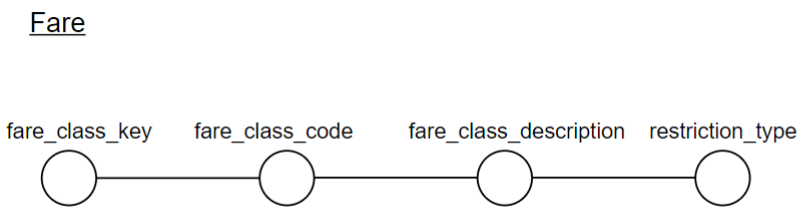
Lest start talking about the Customer hierarchy.



For the customer hierarchy I have chosen to give his own hierarchy level to ‘customer_key’ and ‘customer_name’ and I have put all the location information into the customer information hierarchy and all the personal information inside the ‘customer_info’ hierarchy.

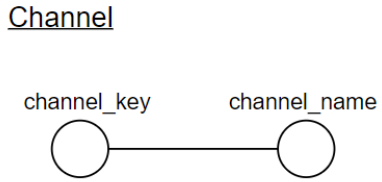
-Fare

For the fare dimension I have chosen to give one hierarchy for each attribute.



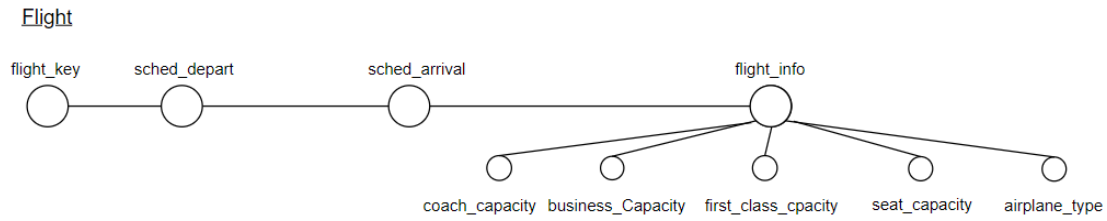
-Channel

For the channel dimension I have chosen to give one hierarchy to each attribute as well.



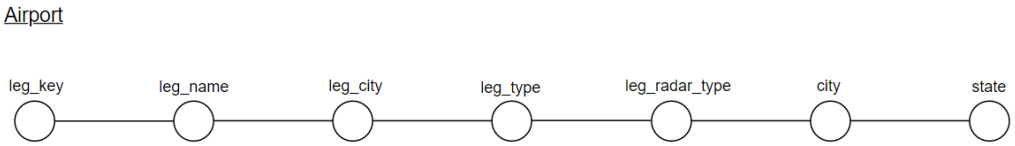
-Flight

Fort the flight dimension I have given one hierarchy level to ‘flight_info’ which contains all the information about the plane such as the coach capacity, the business capacity, the first class capacity the total seat capacity and the airplane type, after that I have given each other attribute its own hierarchy level.



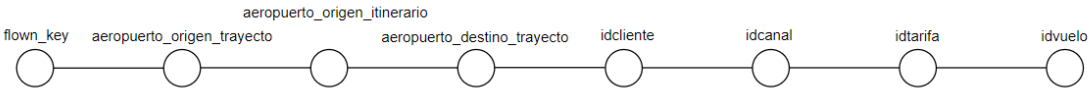
-Airport

For the airport dimension I have given one hierarchy level to each attribute because I think that it is easier to work with.



FACT TABLE AND MEASURES

For the dimensions of the fact table, we have:



Flown_key is the key of one flight inside the client table, if that flight has a stopping point before getting to the destiny it will have the same flown key.

Aeropuerto origen trayecto is where you get the flight in this moment.

Aeropuerto origen itinerario is where you first take the flight, meaning that I you have to go from 17 to 20 but you have to stop in 8 to get there in the second flown_key this attribute will be the first flight you got in.

Aeropuerto destino trayecto is where you are going to stop in case that you are going to stop somewhere in between your destiny.

Aeropuerto destino itinerario is where you are heading in the first place.

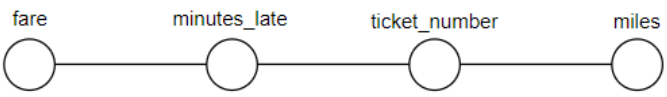
Idcliente is the id for each client.

Idcanal is the id for each channel.

Idtarifa is the id for each fare you got to flight.

Idvuelo is the id for the flight you took.

For the measures we have the ones specified in the case study.



The fare is the price you pay for each flight. With it we can also calculate the total price of the flight.

The minutes late as it names implies is the delay of the flight.

The ticket number is the number of the ticket you got when buying the flight.

And the miles is the distance your plane is flying.

We are also asked to get the points obtained but with the information we have in the tables is not possible to calculate that.