



ugr

Universidad
de Granada

Sistemas Multidimensionales

*Diseño Multidimensional:
Nivel Conceptual y Lógico*

Práctica 7: Mortalidad en EEUU

Alberto Jesús Durán López

Ingeniería Informática y Matemáticas

2019/2020

Realizamos el diseño Conceptual del siguiente caso: **Mortalidad en EEUU**

Nuestro dataset se ha obtenido del siguiente enlace: [Mortality reporting system](#)

Dividimos el proceso en varias etapas:

-Seleccionar el proceso de negocio a modelar:

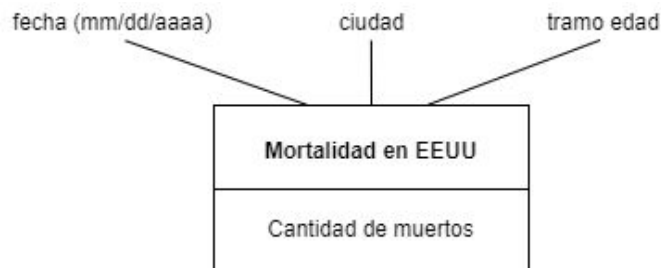
En nuestro caso, sería la mortalidad en EEUU.

Mortalidad en EEUU

Se ha tomado la consideración de realizar dos cubos, ya que la mortalidad venía dada por un motivo (neumonía e influenza y otros) o por un rango de edad, ambos independientes.

Por ello, el proceso sigue así:

-Seleccionar la *granularidad* del proceso de negocio:



<<línea de mortalidad>>

Significado de los hechos: Cantidad de muertos de un tramo de edad determinado en una ciudad y una fecha



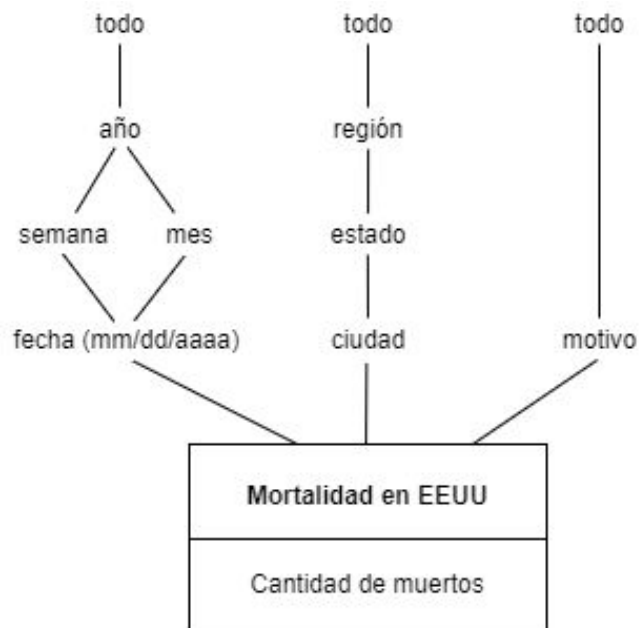
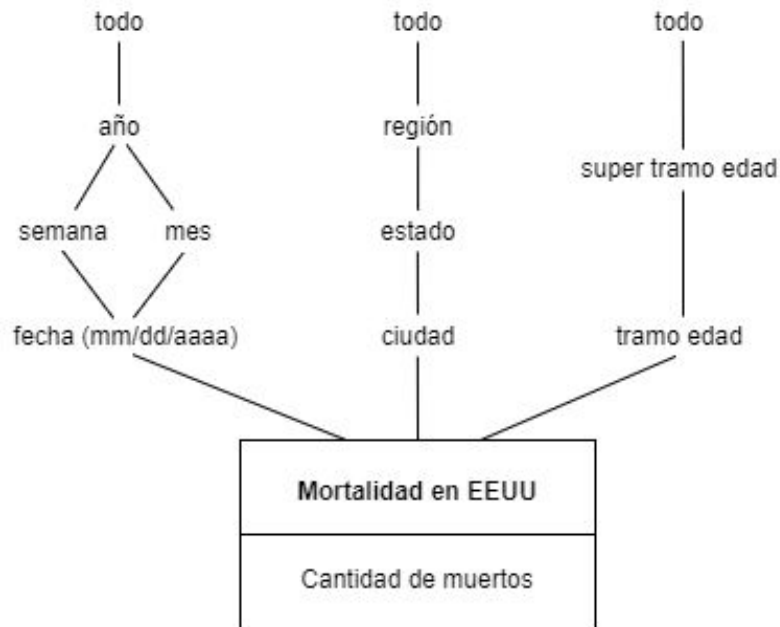
<<línea de mortalidad>>

Significado de los hechos: Cantidad de muertos por un motivo determinado en una ciudad y una fecha.

-Diseñar las *dimensiones* (niveles y jerarquías) y las mediciones

Tramo Edad: <1año / 1-24 años / 25-44 años / 45-64 años / +65 años

Motivo: Neumonía e influenza / otros



Mediciones: Cantidad de Muertos

Aditividad:

Aditiva. Tiene sentido su suma en todas las dimensiones

Estimamos el nº de instancias:

Dimensión Dónde: nº regiones · nº estados/región · nº ciudades/estado = $9 \cdot 4 \cdot 3 - 4 \approx 130$

Hay 9 regiones en total, con unos 4 estados por región (~40 estados) y unas 3 o 4 ciudades por estado. Esto hace un total de 130 instancias.

Dimensión Cuándo): nº años · nºsemanas/año · nº días/semana

Cabe comentar que aunque una semana tenga 7 días o un año tenga 365 días, aquí no se contabilizará así ya que no tenemos la información de todos los días, sino de solo 1 vez a la semana, es decir, en vez de contabilizar día a día se ha contabilizado un día de la semana que recopila toda la información de muertes de ésta. Por ello, las instancias se calcularían:

$$54 \text{ años} \cdot 52 \text{ semanas/año} \cdot 1 \text{ día/semana} \approx 2800 \text{ instancias}$$

Dimensión Tramo Edad: nº tramos de edad. En total hay 5 instancias

Dimensión Motivo : nº de motivos de fallecimiento. En total hay 2 instancias.

Hechos motivo: Se guarda la cantidad de muertos por cada uno de los dos motivos por cada ciudad y por cada fecha, luego:

$$\text{Nº ciudades} \cdot \text{nº fechas} \cdot \text{nº motivos} = 130 \cdot 2808 \cdot 2 \approx 730.000 \text{ instancias}$$

Hechos tramo edad: Se guarda la cantidad de muertos por cada tramo de edad por cada ciudad y por cada fecha, luego:

$$\text{Nº ciudades} \cdot \text{nº fechas} \cdot \text{nº tramos de edad} = 130 \cdot 2808 \cdot 5 \approx 1.825.000 \text{ instancias}$$

Como vemos, el número de instancias de nuestro dataset es muy elevado ya que se han recopilado datos de todas las regiones durante más de 50 años.

A la hora de hacer las dimensiones en Spoon, los pasos/nodos tardaban bastante tiempo en hacer modificaciones y la máquina proporcionada se quedaba bloqueada. Por ello, he decidido reducir las instancias del dataset original y quedarme con unas regiones y un rango de años determinado.

Como consideración adicional, he nutrido la dimensión **Cuándo** añadiendo un nuevo campo llamado *Mes* y la dimensión **Tramo Edad** donde he añadido el campo *super tramo de edad*.

Diseño Lógico