



# UNIVERSIDAD DE GRANADA

## **Sistemas Multidimensionales**

### **Práctica 8 – Diseño multidimensional**

---

Arturo Alonso Carbonero

## 1. Descripción de los datos

Los datos obtenidos se corresponden con el gasto nacional en protección ambiental por sector institucional, estructura porcentual y tasas de variación interanuales, obtenidos en el siguiente enlace:

<https://datos.gob.es/es/catalogo/ea0010587-gasto-nacional-en-proteccion-ambiental-por-sector-institucional-estructura-porcentual-y-tasas-de-variacion-interanuales-identificador-api-t26-p088-serie-l0-01002-px>

### Dimensiones

- Qué: Tipo de indicador.
  - Niveles: Diferentes tipos de indicadores.
- Dónde: Sector institucional.
  - Niveles: Diferentes tipos de sectores.
- Cuándo: Período.
  - Jerarquía (de abajo a arriba): Año – Decenio.

### Hechos

Un hecho (cuánto) significa el gasto total del tipo concreto de gasto que se ha llevado a cabo en determinado sector institucional en un año concreto.

## 2. Procesos de extracción y actualización de datos

### Extracción

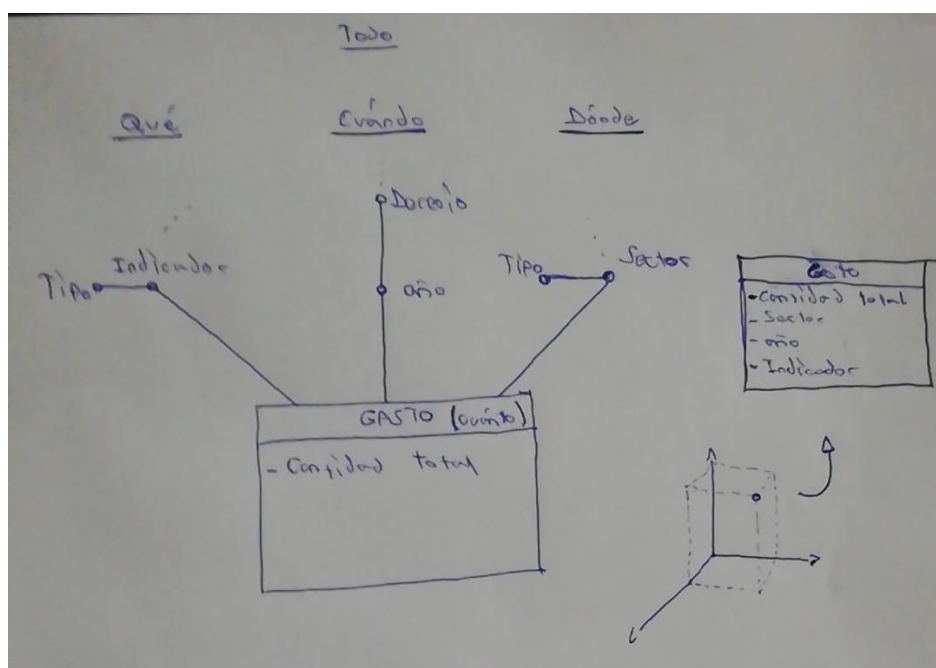
Esta parte se corresponde con la labor del componente ETL, esto es, extraer, transformar y cargar los datos. En primer lugar, los datos se extraen de la base de datos OLTP pertinente. En este caso, proceden de una base de datos del gobierno de España, que contiene conjuntos de datos públicos, la cual se puede acceder desde el siguiente enlace: <https://datos.gob.es/es>

### Transformación y carga

Una vez disponemos de los datos en el sistema ETL, como por ejemplo SSIS, aplicamos a estos una serie de transformaciones para adaptarlos al resultado que deseamos obtener y los enviamos a la base de datos OLAP. Para este caso concreto, he considerado las siguientes transformaciones:

- Eliminación de la columna final → En el fichero con extensión ‘csv’ que contiene los datos seleccionados aparece una columna al final que no tiene nombre y que, aparentemente, no representa nada relevante, ya que no tiene nombre y solo contiene valores numéricos, por lo que se eliminaría dicha columna.
- Modificación de la columna ‘periodo’ → En dicha columna, la cual contiene los años en los que se ha llevado a cabo el gasto, aparecen algunas filas con el año y la cadena de caracteres ‘(avance)’. Habría que adaptar dicha columna de forma que únicamente aparezcan los años.
- Modificación de la columna ‘Total’ → En algunas de las filas de la columna que contiene el valor total en euros de cada gasto aparece la cadena de caracteres ‘..’ en lugar de un valor numérico, por lo que habría que sustituir dichas celdas por el valor numérico ‘0’.
- Modificaciones generales → Para este caso concreto, habría que adaptar los tipos de las columnas al tipo deseado. Concretamente habría que colocar el tipo numérico a las columnas con valores numéricos.

### 3. Diseño conceptual



#### Aditividad de los hechos

El gasto tiene propiedad semi-aditiva ya que no siempre tiene sentido sumar dicha cantidad. Es posible que no sea útil o carezca de sentido el hecho de sumar el gasto de diferentes sectores, pero sí tiene sentido sumar el coste a lo largo de los años.

#### Medidas calculadas

Como medidas calculadas, se podrían calcular en algún caso alguna de las siguientes entre otras: La suma de sectores de un tipo o la sumatoria total de sectores. En cualquier caso, nunca tendría sentido calcular la sumatoria de la dimensión 'cuándo'.

#### Número de instancias

Tipo de indicador (2) x Años (10) x Tipo de sector (4) = 120 instancias. Sin embargo, no habrá un único gasto para cada tipo e indicador en cada año. Supongamos que el número de indicadores pues pasa a ser de 2 y el número de sectores y años se mantiene, es decir, para un mismo sector habrá de media un gasto por año para 2 tipos de indicador. Esto hace un total de 80 instancias.

#### 4. Diseño lógico

El sistema se actualiza una única vez por año. Se trata de un sistema multidimensional dedicado a consultas, por lo que lo lógico es emplear un diseño ROLAP en estrella como el de la imagen. Al tratarse de un caso sencillo, no existe prácticamente ninguna excepción lógica por lo que no necesitamos dimensiones de tipo cajón de sastre, degeneradas o desdoblamientos.

Aun así, podemos considerar la dimensión de los sectores como una dimensión lentamente cambiante. Si añadimos un gasto en un sector que que entre en conflicto con otro, por coincidir el indicador, pero de otro año, habría que añadir un nuevo registro para esa entrada por lo que se trata de una SCD de tipo 2.

