



UNIVERSIDAD DE GRANADA

Sistemas Multidimensionales Práctica 8: Diseño Multidimensional

Grupo A2
Amador Carmona Mendez
amadorcm@correo.ugr.es

Introducción

Vamos a realizar un sistema multidimensional sobre los datos dados por el siguiente enlace que trata sobre el personal de investigación en España. Nuestro sistema se actualizará una vez por año y lo tendremos únicamente para consultas de datos durante el resto del tiempo. Enlace a la pagina de donde se han obtenido los datos:

<https://datos.gob.es/es/catalogo/e05073401-pi-total-por-tipo-de-universidad-tipo-de-centro-tipo-de-personal-y-nacionalidad>

Enlace a la descarga de los datos en formato xml:

https://estadisticas.mecd.gob.es/EducaJaxiPx/files/_px/es/xlsx/Universitaria/Personal/EPU20/PI/I0/PI0203.px

Descripción de los datos en los que se basa.

Nuestro sistema multidimensional tratara con los datos del número de personal de investigación que trabajan en españa, teniendo en cuenta: el año en el que trabajaron, la nacionalidad de estos, la comunidad autónoma donde trabajan, el tipo de centro donde trabajan si es centro es público o privado y si son centros propios o adscritos , el tipo de personal que es, si son funcionarios o contratados, si es personal docente e investigador (PDI), Personal Empleado Investigador (PEI), Personal Técnico de Apoyo a la investigación (PTA)y Personal de Administración y Servicios (PAS).

Por resumir nuestro sistema multidimensional consta de:

Los hechos: El personal de investigación que ha trabajado en un año determinado, en una comunidad autónoma determinada y que es de un tipo determinado trabajador en un centro determinado.

Las bases de nuestros hechos vienen dadas por el año, la comunidad autónoma, el tipo de trabajador y el tipo de centro.

Las dimensiones:

- 1.Cuando: Almacenará el año del número de trabajadores.
- 2.Donde: Almacenará la comunidad autónoma de los trabajadores.
- 3.Tipo-Trabajador: Almacenará el tipo de personal del que se trata, tendría una jerarquía, en el diseño conceptual lo específico mejor.
- 4.Nacionalidad: Almacenará la nacionalidad de los trabajadores.(Español o extranjero)
- 5.Tipo-Centro: Almacenará el tipo de centro en el que trabaja cada personal, tendremos otra jerarquía.

Descripción de los procesos de extracción y actualización de datos.

Para la extracción de los datos utilizaremos una herramienta ETL.

Los datos los obtendremos de las diferentes fuentes de datos especificadas en el enlace de la introducción.

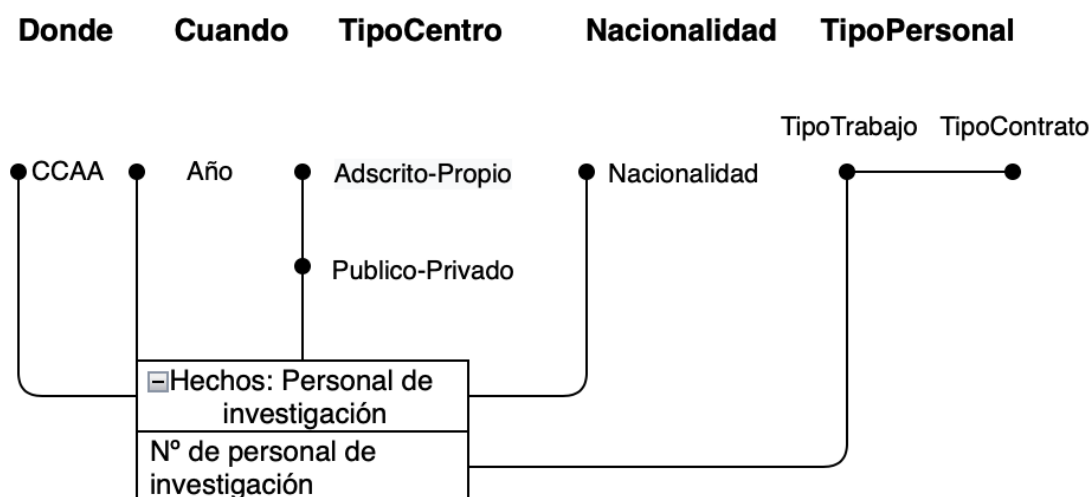
Las transformaciones que sufrirían los datos desde las fuentes, serían las siguientes:

Se borrarán todos los datos nulos, los años y el número de personal se incorporarán en formato entero, los demás en formato alfanumérico, en todas las dimensiones, cada dato iría la dimensión que le pertenezca o a los hechos si pertenece a estos, y si es necesario se crearán nuevas claves donde corresponda para el dato que se inserte, las claves serán de tipo entero.

Y por último se cargarán en nuestro sistema Multidimensional.

Diseño conceptual.

El esquema de nuestro sistema multidimensional es el siguiente:



Donde tenemos las dimensiones anteriormente enumeradas, y donde nos encontramos las siguientes jerarquías:

Jerarquías:

Tipo Personal: Tendríamos una jerarquía en la que se encuentra en el mismo nivel las dos agrupaciones de tipos de personal que podemos tener.

Tipo de contrato — Tipo de trabajo

Tipo Centro: Tendríamos una relación uno a muchos donde un centro público o privado puede tener muchos centros propios o adscritos.

Adscrito-Propio



Público-Privado

También vamos a valorar las medidas acumuladas y la aditividad de los hechos:

Posibles Medidas Acumuladas: Se podría calcular como medida acumulada el total, es decir la sumatoria de los hechos teniendo en cuenta diferentes dimensiones, se puede calcular el total de PI independientemente del tipo, o del centro, o de la comunidad autónoma o de la nacionalidad, la única dimensión que no tendría sentido hacer la sumatoria sería la dimensión tiempo.

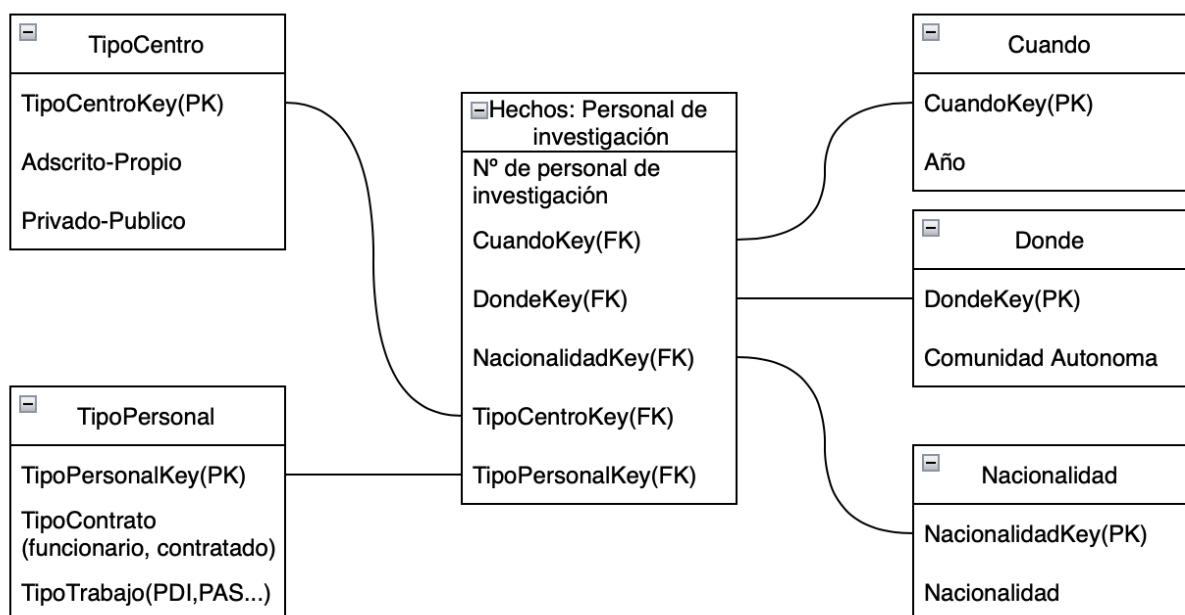
Aditividad de los Hechos: Los hechos tienen propiedad semi-aditiva, ya que no siempre tiene sentido que sean sumados, por ejemplo a lo largo de los años no tiene sentido que sean sumados, pero si tiene sentido sumar los hechos según hemos dicho en el apartado anterior de las medidas acumuladas.

Procedemos a Estimar el número de instancias:

Tenemos 17, comunidades autónomas en España, tenemos 2 tipos de centros (público o privado) x 2 tipos de centro según la jerarquía, 4 tipos de personal (PDI, PEI, PTA, PAS) x 2 tipos de contrato (funcionario y contratado) x 2 tipos de nacionalidad, tiene sentido que se multiplique todo, ya que no va a variar ni los tipos de centros ni los tipos de personal ni el número de comunidades autónomas y cada año se hallará el mismo número de hechos, por lo que tendremos 544 hechos por año sin medidas calculadas, si se añadieran las medidas calculadas (todos los totales, según el tipo de contrato, el total según la nacionalidad, según el tipo de centro, etc) aumentará el número de hechos significativamente, entre x2 y x7 según la cantidad de medidas acumuladas que calculemos.

Diseño lógico.

El esquema lógico de nuestro sistema multidimensional es el siguiente:



Puesto que el sistema multidimensional sólo se actualizará una vez por año, está enfocado a consultas de datos no a inserciones y borrado, por lo que utilizaremos un modelo de datos ROLAP en estrella que es mucho más eficaz para esto. Respecto a las dimensiones, no contamos con ninguna excepción lógica en cuanto a estas por lo que no hay que valorar ningún tipo de solución como desdoblamiento, cajón desastre, dimensión degenerada, etc.