



**ESCUELA SUPERIOR DE INGENIERÍA  
Tecnologías de Inteligencia de Negocios  
PROYECTO FINAL**

**Alumna: Comaleras María Manuela  
Fecha de entrega: 2-07-2020**



## 1. Objetivo del proyecto

El proyecto tiene como objetivo el análisis de personas que inmigraron a España. Para esto, se utilizan bases de datos que contienen visados efectuados en las Oficinas Consulares españolas.

Dicho análisis se realiza mediante la herramienta *Pentaho de Business Intelligence* destinada tanto para el desarrollador como para el usuario final. El mismo permite obtener una visión integrada y de calidad de la cantidad de inmigrantes discriminados por su sexo, edad, lugar de origen, año y dependiendo fundamentalmente del tipo de visado que se ha realizado durante su estancia.

## 2. Métrica y dimensiones

La métrica principal que es objeto de interés es el número de visados. Dicha métrica se analiza desde distintos puntos de vistas o dimensiones. Las dimensiones principales son las siguientes:

### Tiempo

- Visados expedidos en 2016 en oficinas consulares.
- Visados expedidos en 2017 en oficinas consulares.

### Tipo

- Visados de corta duración
- Visados de larga duración

### Edad

- Visados de personas comprendidas entre cero y quince años.
- Visados de personas comprendidas entre dieciséis y cuarenta y cuatro años.
- Visados de personas comprendidas entre cuarenta y cinco años y sesenta y cuatro.
- Visados de personas de sesenta y cinco años o más.

### Sexo

- Visados de mujeres
- Visados de hombres

### Región

- América del Norte
- América Central y del Sur
- África
- Asia

- Oceanía
- Resto de Europa

### 3. Características de los sistemas de origen

Para realizar el estudio, se extrajeron dos bases de datos del Portal de Inmigraciones. El mismo ha utilizado información estadística de EUROSTAT, OCDE y European Migration Network.

Dichas bases de datos poseen formato EXCEL.

## 4. Arquitectura del proyecto de BI

### 4.1 Representación visual de BI

En la figura 1 se puede observar el funcionamiento del conjunto de estrategias y herramientas enfocadas al análisis de datos utilizadas en este proyecto.

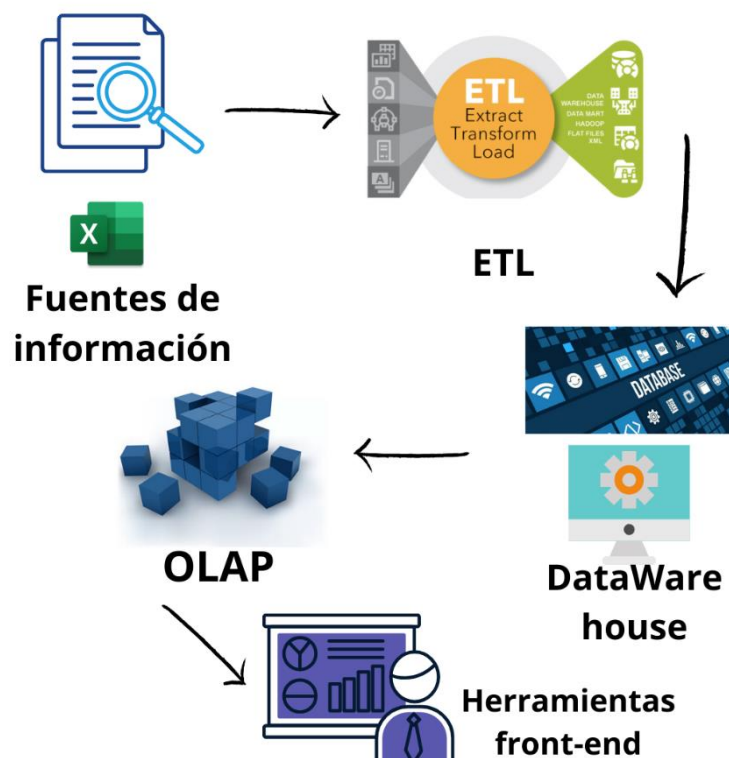


Figura 1. Representación visual BI

El proyecto parte de las fuentes de información que se utilizan para alimentar el DataWarehouse. A partir de esta, mediante un proceso ETL se realiza la extracción, transformación y carga de datos en el DataWarehouse.

El mismo es un almacén de datos con un enfoque dimensional.

Luego, en el entorno OLAP la información se conceptualiza en estructuras de datos de modelado dimensional para realizar tareas de análisis.

Posteriormente se generan cuadros de mando e informes (herramientas Front-end) para que los usuarios finales puedan observar la información de forma clara y sencilla.

#### 4.2 Modelo de datos diseñado para el DW

Los usuarios solicitarán información a la base de datos en términos de hechos y dimensiones. En este proyecto se emplea un modelo de datos de “esquema en estrella” para el DW. La tabla central representa el hecho (número de visados) y las demás tablas representan dimensiones o puntos de vista (temporal, regional, de tipo, de sexo y edad) del hecho.

En la siguiente imagen se puede observar un esquema representativo.

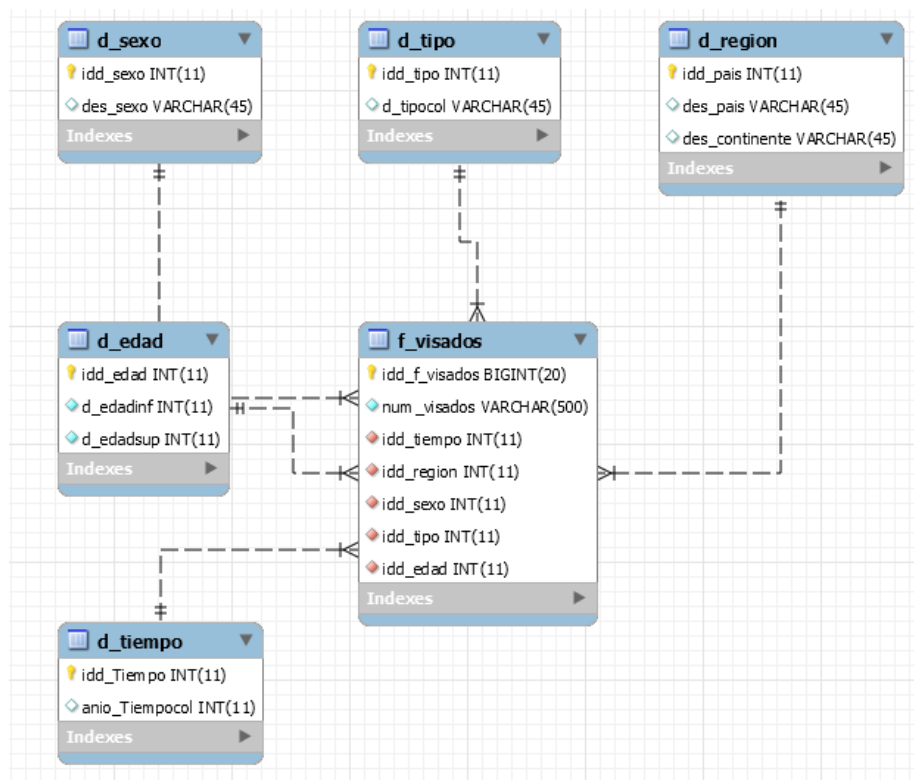


Figura 2. Esquema en estrella

#### 4.3 Proceso ETL

Para la integración de datos se realiza el proceso ETL en *Data Integration*, mediante la herramienta *Spoon* para diseñar y ejecutar los procesos y poblar el DW. El mismo realiza consultas offline de datos con una base MySql.

Las actividades que son llevadas a cabo mediante el ETL y la herramienta son las siguientes:

- Extracción. Donde identifica el subconjunto de los datos fuente que deben ser suministrados al flujo posterior.
- Transformación. Con tareas de validación, limpiado, decodificación y agregación de los datos.
- Carga de las tablas de hecho y dimensiones.

En el trabajo se realizan dos transformaciones (para los años 2016 y 2017 respectivamente) por donde fluyen los datos a través de diferentes *steps*. Cada *step* realiza distintas actividades de extracción, transformación y carga.

Los valores están asociados con un *field*, el cual contiene metadatos (sexo, edad, región, tipo y tiempo). A continuación, se ilustra el proceso que contiene cada una de dichas transformaciones.

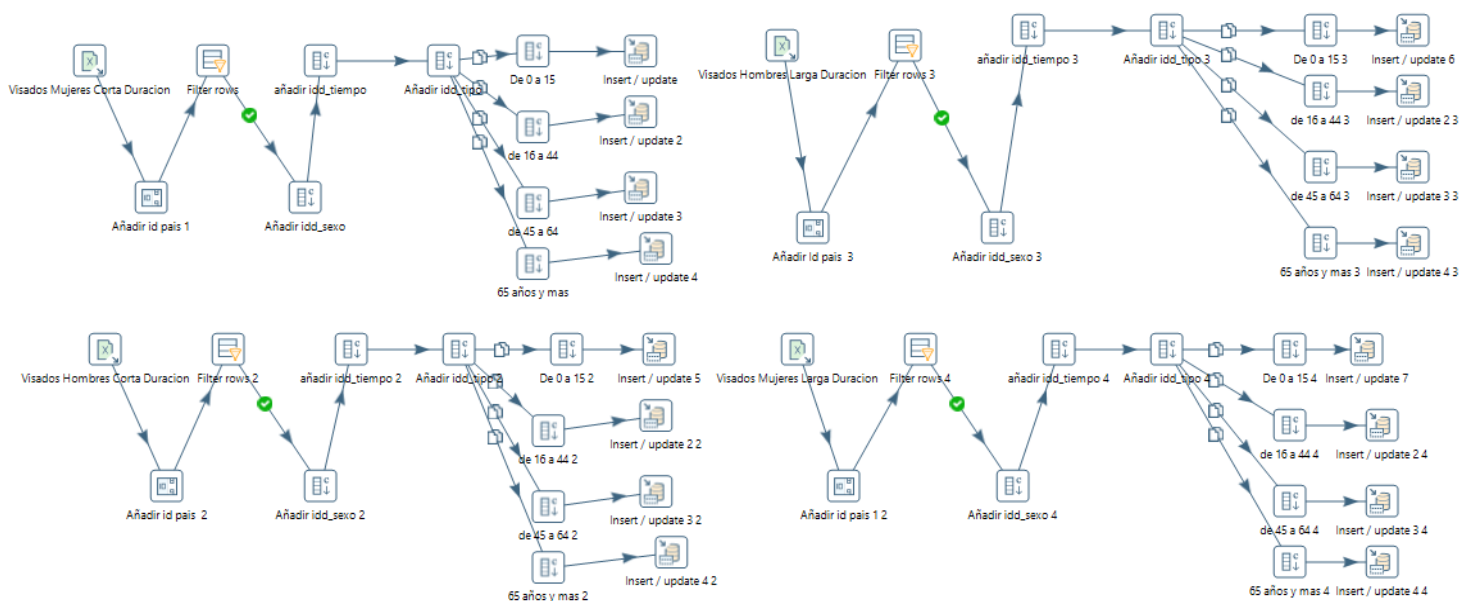


Figura 3. Transformación en herramienta Spoon

#### 4.4 Informe

Para realizar los informes se utiliza *Pentaho Report Designer*.

Se crean dos reportes con la herramienta. Para esto se conecta la misma con la fuente de datos MySQL. Posteriormente se utiliza el conector JDBC de Java. Luego, se crea una query para escribir la consulta SQL para extraer los subconjuntos de datos requeridos para cada informe.

Posteriormente se extraen como PDF y adjuntados en el repositorio.

El primer informe está formado por tablas que contienen los rangos de edad y dentro de cada tabla un listado con el numero de visados según el continente.

julio 01, 2020 @ 07:21

**Visados por edad y continente**

Rango de edad: 0 - 15	
continente	visados
África	310.800
Asia	220.582
América del Norte	1.588
Oceanía	105
Resto de Europa	603.572
América Central y del Sur	48.334
Australia	55
Rango de edad: 16 - 44	
continente	visados
Resto de Europa	1.983.018
América Central y del Sur	333.652
Australia	494
África	1.065.698
Asia	1.578.268
América del Norte	51.154
Oceanía	1.058
Rango de edad: 45 - 64	
continente	visados

Figura 4. Reporte Visados 1

El segundo informe discrimina dos tablas, del 2016 y 2017 respectivamente determinando el numero de visados según la duración de la estancia. Se puede deducir a simple vista el crecimiento de inmigrantes al país.

julio 01, 2020 @ 07:22

**Visados por año y tipo**

Año 2016	
tipo	visados
corta duracion	2.871.170
larga duracion	1.599.186
Año 2017	
tipo	visados
corta duracion	2.917.956
larga duracion	1.652.058

Report Footer

Figura 4. Reporte Visados 2

#### 4.5 Cubo OLAP

OLAP permite a los usuarios finales obtener conocimiento de los datos mediante el acceso rápido, consistente e interactivo a una gran variedad de posibles vistas de la información estructurada.

Para el proyecto se utiliza la herramienta Manage Datasource. Para visualizar el cubo *Pivot4J View* de *Pentaho User Console*. El mismo soporta consultas multidimensionales MDX sin necesidad de escribir el código.

El cubo está compuesto por la métrica que es el número de visados y sus dimensiones mencionadas anteriormente. Posee una jerarquía de la dimensión región, que es una estructura que permite obtener información en dos niveles: continentes y países. El mismo se puede observar a continuación.

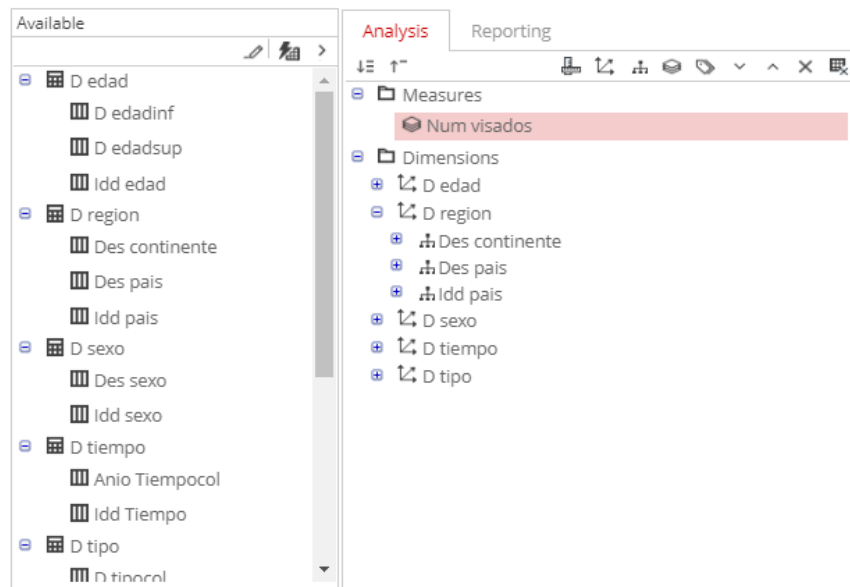


Figura 5. Cubo OLAP

#### 4.6 Cuadros de mando

Se realiza un cuadro de mando para observar de forma gráfica y precisa la información. Para el mismo se utiliza *Pentaho User Console* con tecnología *Community Dashboard Framework* que posee una herramienta visual para crear, editar y visualizar los cuadros de forma sencilla.

Esta herramienta ofrece tres vistas. Primero el *layout* que permite definir la disposición de los elementos en la página, en este caso se utilizan cinco paneles (dos en la parte superior y tres en la inferior).

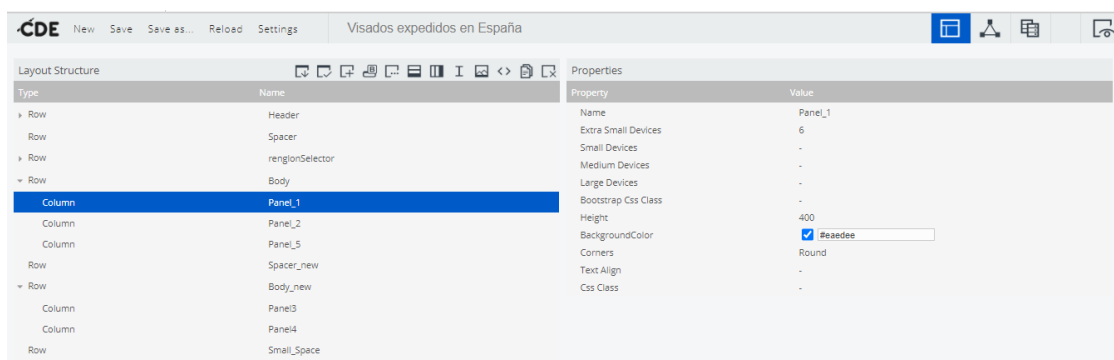


Figura 6. Vista *layout*



La segunda vista son los *components* que permite configurar los elementos visuales, parámetros y scripts.

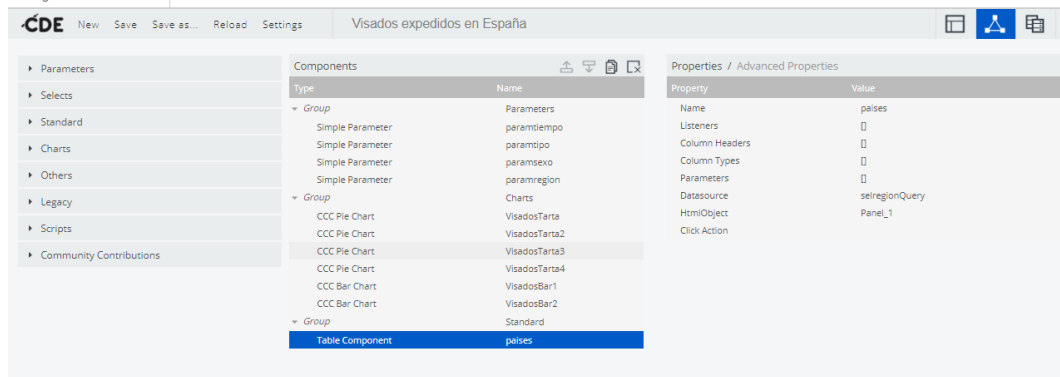


Figura 7. Vista components

Por último, la tercera vista es *Datasources* que permite crear las fuentes de datos.

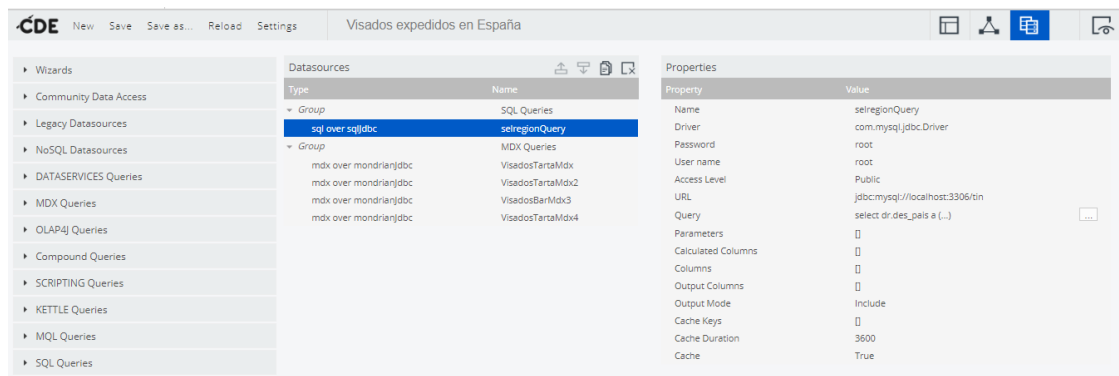


Figura 8. Vista Datasources

La tabla del primer panel en el cuadro de mandos del trabajo brinda información de la cantidad de visados totales realizados por país. En el segundo y tercer panel se visualizan dos gráficos de región y barra respectivamente. De forma iterativa, primero se discrimina por el tipo de visados y luego por continente. Se muestra en la figura 9.

Posteriormente, en el cuarto panel con un gráfico de barras se clasifica por año y luego por sexo en otro de región. Se puede visualizar en la figura 10.

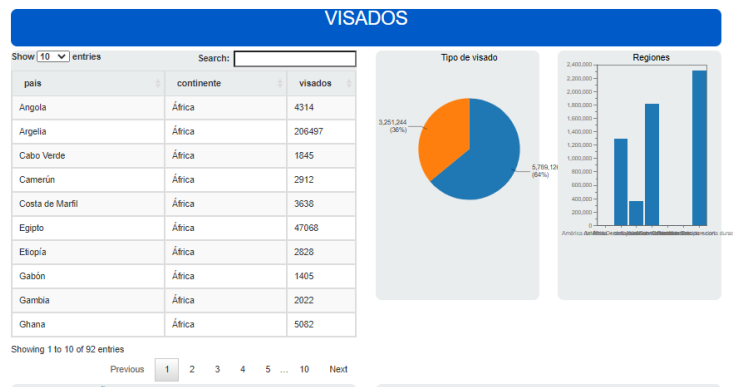


Figura 9. Cuadro de mando

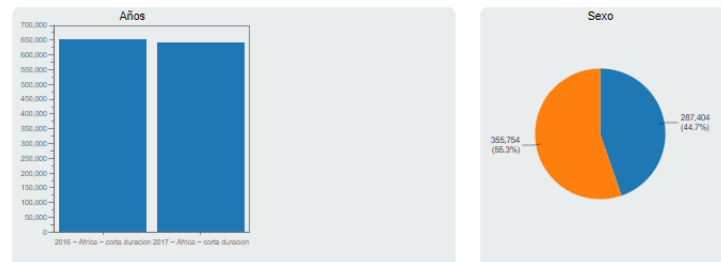


Figura 10. Cuadro de mando

#### 4.7 Principales problemas encontrados

La principal dificultad encontrada fue la falta de conocimiento del uso de base de datos y programación necesaria para llevar a cabo el proyecto. Igualmente, Pentaho ofrece diversas formas gráficas para no tener que recurrir a la necesidad de realizar programación.

## Bibliografía

<http://extranjeros.mitramiss.gob.es/es/estadisticas/operaciones/visados/index.html>

Campus virtual UCA