

A dark blue vertical bar runs down the left side of the page. A blue arrow points to the right from this bar, containing the text '2016/2017'.

2016/2017

Proyecto OLAP

Tráfico

Several thin, curved lines in dark blue and light grey originate from the bottom left corner and sweep upwards and to the right.

Héctor Rodríguez Salgado & Marta Lorient Nieves

Índice de contenido

Índice de ilustraciones.....	1
Introducción	2
Obtención de datos.....	2
Almacén de datos.....	3
Diseño del esquema OLAP	3
Extracción, transformación y carga de datos.....	4
Cubo OLAP.....	4
Resultados.....	5
Ejemplos de sentencias MDX	5

Índice de ilustraciones

Ilustración 1 - Listado de carreteras.....	3
Ilustración 2 - Listado de vehículos	3
Ilustración 3 - Esquema OLAP	4
Ilustración 4 - Diseño del cubo OLAP	5

Introducción

Se ha desarrollado un sistema de almacenamiento de datos OLAP para ayudar en el futuro análisis de datos de las infracciones de tráfico producidas en España para aumentar los métodos de prevención y evitar las muertes que se producen en la carretera año tras año.

Durante la última década, se ha disminuido la tasa de muertos por accidentes de tráfico en las carreteras españolas de forma considerable. Esto se ha producido, en gran medida, a la mayor conciencia de los conductores por este problema social y a la actuación de las fuerzas de seguridad, con medidas de prevención y de seguridad. Para que todo esto sea posible es necesario analizar una gran cantidad de datos para proteger las zonas más propensas a que se produzcan accidentes, realizar campañas en puntos estratégicos, motivar a los conductores a realizar una conducción respetuosa y a no saltarse las normas, etc.

El objetivo de este proyecto es ayudar a la disminución de los accidentes de tráfico mediante el procesado de datos y el posterior análisis, que sirve de ayuda para las campañas de prevención y la adopción de medidas de seguridad en puntos negros, así como una mayor concienciación para los conductores a la hora de coger su vehículo.

Obtención de datos

El análisis a realizar comprende el período desde el año 2000 hasta el año 2016. Por tanto, se ha realizado una búsqueda de los accidentes de tráfico en España durante ese período.

El objetivo de esta búsqueda es la obtención de las infracciones producidas en España durante este período, necesitando como datos los siguientes:

- Tipo, descripción y penalización de la infracción.
- Lugar dónde se ha producido.
- Fecha en la que se ha producido.
- Conductor que ha cometido la infracción.
- Vehículo que el conductor estaba conduciendo en ese momento.

Tras realizar una larga búsqueda, no se han podido obtener datos reales acerca de las infracciones producidas durante este período, por lo que algunos datos han sido generados aleatoriamente.

Los datos reales que se han obtenido han sido los siguientes:

- Listado de las carreteras españolas (autovías, autopistas y nacionales), proporcionado por el Ministerio de Fomento. Dicho listado incluye el nombre de la carretera, el tipo y el km inicial y el final.

Carretera	Tipo	PK Ini	PK Fin
A-1	AUTOPISTAS LIBRE \ AUTOVÍA	12+00690	336+01030
A-11	AUTOPISTAS LIBRE \ AUTOVÍA	105+00996	467+00000
A-12	AUTOPISTAS LIBRE \ AUTOVÍA	76+00200	131+00059
A-13	AUTOPISTAS LIBRE \ AUTOVÍA	0+00000	5+00150
A-14	AUTOPISTAS LIBRE \ AUTOVÍA	6+00000	16+00300

Ilustración 1 - Listado de carreteras

- Listado con los tipos de vehículos en los que se ha cometido la infracción.

```
makeid,makename,vehicletypeid,vehicletypename
440,Aston Martin,2,Passenger Car
441,Tesla,2,Passenger Car
442,Jaguar,2,Passenger Car
443,Maserati,2,Passenger Car
445,Rolls Royce,2,Passenger Car
448,Toyota,2,Passenger Car
449,Mercedes-Benz,2,Passenger Car
452,BMW,2,Passenger Car
454,Bugatti,2,Passenger Car
456,MINI,2,Passenger Car
```

Ilustración 2 - Listado de vehículos

Los demás datos han sido generados.

Almacén de datos

Diseño del esquema OLAP

Con el objetivo de agrupar y almacenar todos los datos que se necesitan para soportar el tipo de análisis a realizar, se ha diseñado un esquema en estrella con una tabla de hechos central y 4 dimensiones (*Ilustración 3*). El diseño y la implementación del modelo se ha realizado con la herramienta MySQL Workbench.

La dimensión *DateTime* permite hacer un análisis, por hora, día, mes, estación y año. Cada uno de ellos tiene un valor identificativo. Además, para el día se desea conocer si es festivo y si pertenece a un período vacacional; para la hora, se desea conocer el tiempo atmosférico.

La dimensión *KmPoint* permite conocer el lugar en dónde se ha producido la infracción. Se desea conocer el punto kilométrico, identificado con un inicio y un fin, el nombre de la carretera de ese punto kilométrico y el tipo de la misma, si se trata de un punto negro, si tiene algún tipo de señalización y si tiene radar.

La dimensión *Driver* está formada por el conductor que ha cometido la infracción. Se necesita conocer la edad del mismo, el sexo, la experiencia que tiene como conductor, las infracciones que haya podido cometer anteriormente y si tiene algún tipo de enfermedad.

La dimensión *Vehicle* permite conocer el vehículo en el que se ha cometido la infracción. Se desea conocer el tipo de vehículo, la marca, el número de asientos disponibles, los pasajeros que iban en el vehículo, la antigüedad del mismo, si se necesita un permiso de conducir para poder transitar con él y si es eléctrico.

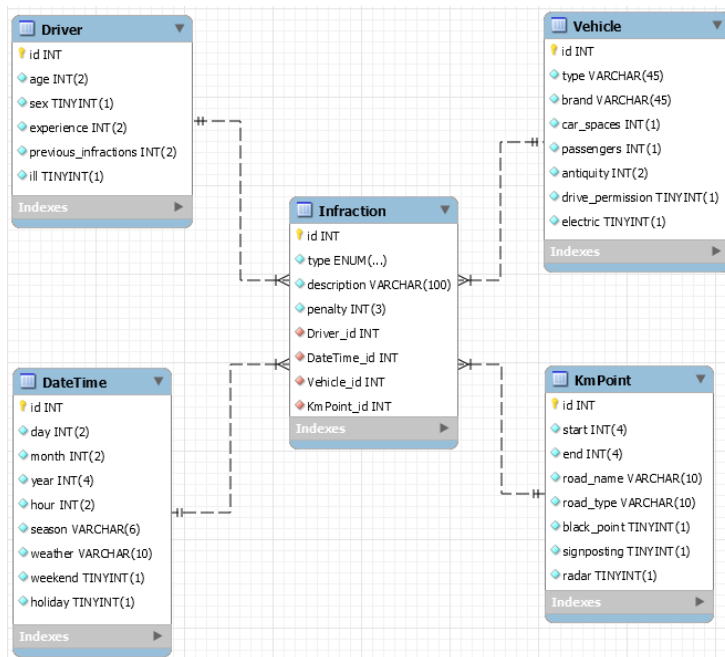


Ilustración 3 - Esquema OLAP

Extracción, transformación y carga de datos

Para rellenar el almacén de datos, se ha desarrollado un programa en el lenguaje de programación Java utilizando el IDE Eclipse, junto con la librería para poder crear y modificar archivos con formato csv y la librería que permite conectarse a MySQL.

En primer lugar, se han tratado los datos que se han extraído de las fuentes oficiales y se han creado los correspondientes archivos. En segundo lugar, se han generado los datos restantes de forma aleatoria y se han volcado a los archivos correspondientes. Una vez realizado esto, las fuentes de datos necesarios estaban completas.

Posteriormente, se realizó la transferencia de estas fuentes de datos a la base de datos SQL.

Ambas tareas han requerido bastante tiempo de procesamiento debido a la gran cantidad de datos que se han generado.

Cubo OLAP

Después de haber construido la base de datos, se ha diseñado el cubo OLAP que va a dar soporte a las futuras consultas a realizar. Para ello, se ha utilizado la herramienta Pentaho Schema Workbench. El diseño de un cubo OLAP comprende la creación de un cubo con medidas, dimensiones y jerarquías dentro de ellas, para posteriormente soportar diferentes consultas MDX. En la *Ilustración 4* se puede observar el diseño de dicho cubo, donde se pueden observar las distintas dimensiones, algunas de ellas organizadas en jerarquías, permitiendo un manejo más fácil a la hora de realizar las consultas.



Ilustración 4 - Diseño del cubo OLAP

Resultados

Ejemplos de sentencias MDX

1. Importe de las infracciones producidas en los vehículos de tipo coche

SELECT

[Measures].[penalty] ON COLUMNS,

[vehicle.vehicles].[type].[car] ON ROWS

FROM [Traffic]

Axis #0:

{}

Axis #1:

{[Measures].[penalty]}

Axis #2:

{[vehicle.vehicles].[Car]}

Row #0: 15.113.400

2. Importe total de las infracciones

```
SELECT
    [Measures].[penalty] ON COLUMNS
FROM [Traffic]
```

Axis #0:

{}

Axis #1:

{[Measures].[penalty]}

Row #0: 27.788.100

3. Importe de las infracciones producidas en cada tipo de vehículo para los conductores de 18 años durante el año 2010

```
SELECT
    [Measures].[penalty] ON COLUMNS,
    [vehicle.vehicles].[type].MEMBERS ON ROWS
FROM [Traffic]
WHERE ([Driver.Features].[age].[18])
```

Axis #0:

{[Driver.Features].[18], [DateTime.datetimes].[2010]}

Axis #1:

{[Measures].[penalty]}

Axis #2:

{[vehicle.vehicles].[Car]}

{[vehicle.vehicles].[Lorry]}

{[vehicle.vehicles].[Motorbyke]}

{[vehicle.vehicles].[Tractor]}

{[vehicle.vehicles].[Van]}

{[vehicle.vehicles].[Without_carnet]}

Row #0: 2.600

Row #1:

Row #2: 500

Row #3: 900

Row #4:

Row #5:

4. Importe de las infracciones por estación en la AP-9

```
SELECT
    {[DateTime].[season].MEMBERS} ON COLUMNS,
    {[kmpoint].[road_name].[AP-9]} ON ROWS
FROM [Traffic]
```

Axis #0:

{}

Axis #1:

```
{[DateTime.datetimes].[2000].[autumn]}
{[DateTime.datetimes].[2000].[spring]}
{[DateTime.datetimes].[2000].[summer]}
{[DateTime.datetimes].[2000].[winter]}
{[DateTime.datetimes].[2001].[autumn]}
{[DateTime.datetimes].[2001].[spring]}
{[DateTime.datetimes].[2001].[summer]}
{[DateTime.datetimes].[2001].[winter]}
...
{[DateTime.datetimes].[2016].[autumn]}
{[DateTime.datetimes].[2016].[spring]}
{[DateTime.datetimes].[2016].[summer]}
{[DateTime.datetimes].[2016].[winter]}
```

Axis #2:

{[kmpoint.road].[AP-9]}

Row #0: 300

Row #0:

Row #0:

Row #0:

Row #0: 500

Row #0:

Row #0:

Row #0:

...

Row #0: 100

Row #0:

Row #0:

Row #0:

5. Penalizaciones de los conductores de 20 años durante el año 2016

```
SELECT
    Measures.[penalty] ON COLUMNS,
    {[Driver.Features].[age].[20]} ON ROWS
FROM [Traffic]
WHERE [Datetime.datetimes].[year].[2016]
```

Axis #0:

{[DateTime.datetimes].[2016]}

Axis #1:

{[Measures].[penalty]}

Axis #2:

{[Driver.Features].[20]}

Row #0: 10.300

6. Penalizaciones de los conductores de 20 años en la carretera A-1

```
SELECT
    [kmpoint.road].[road_name].[A-1] ON COLUMNS,
    Measures.[penalty] ON ROWS
FROM [Traffic]
WHERE [Driver.Features].[age].[20]
```

Axis #0:

{[Driver.Features].[20]}

Axis #1:

{[kmpoint.road].[A-1]}

Axis #2:

{[Measures].[penalty]}

Row #0: 5.400