

# Caso práctico: almacén de datos para el análisis del impacto conductual de la COVID-19 sobre la población

## PRA1- Análisis y diseño del *data warehouse*

### Presentación

La PRA1 consiste en el análisis y el diseño multidimensional de un almacén de datos para examinar el impacto conductual de la COVID-19 sobre la población.

Para llevar a cabo su desarrollo, se ha publicado la siguiente información:

- **Caso práctico** (Caso\_Práctico\_Enunciado.pdf). Documento con la información del caso de uso para desarrollar un almacén de datos que permita analizar la información del impacto conductual de la COVID-19 sobre la población. En el documento se describen el contexto, los usuarios potenciales y las fuentes de datos. También se realiza una descripción del enunciado de la práctica con cada una de las fases para construir el almacén de datos, así como los programas que se utilizarán y la bibliografía.
- **Fuentes de datos** (fuentes.zip). Fichero comprimido con las fuentes de datos disponibles para desarrollar el caso práctico.

### Descripción

A partir del análisis del contexto del caso y de las fuentes de datos disponibles, el estudiante deberá diseñar y proponer un almacén de datos que ofrezca soporte al análisis del impacto conductual de la COVID-19 sobre la población.

Para ello **se pide** llevar a cabo los siguientes puntos:

- 1) **Análisis de los requerimientos** que incluya las preguntas a las que el sistema debe dar respuesta. Se deben definir un mínimo de cinco preguntas y un máximo de diez.
- 2) **Análisis de las fuentes de datos** proporcionadas.
- 3) **Análisis funcional** que proponga el tipo de arquitectura para la factoría de información que mejor se adecue al proyecto.
- 4) **Diseño del modelo conceptual, lógico y físico** del almacén de datos. Se deben identificar y diseñar las tablas de hechos (*facts*), las dimensiones del análisis (*dimensions*) y los atributos que permitan tener el nivel de granularidad suficiente para implementar los requerimientos.

El documento de la solución de esta actividad (PRA1) debe incluir una descripción y una justificación de todas las acciones realizadas en los puntos anteriormente enumerados.

## Guía de muestra

Con el fin de ayudar a alcanzar los objetivos planteados de la PRA1 se desarrolla esta guía. Esta muestra, con un ejemplo, cómo debe de llevarse a cabo y cuáles son los pasos que seguir:

- 1) análisis de los requerimientos,
- 2) análisis de las fuentes de datos,
- 3) análisis funcional, y
- 4) diseño completo del modelo.

### 1. Análisis de los requerimientos

El análisis de los requerimientos se basa en identificar las necesidades que tiene una organización particular respecto al análisis de la información. La necesidad principal es disponer de la información integrada para su análisis y difusión. Esta información debe ayudar a los usuarios potenciales a tomar decisiones y a alcanzar los objetivos deseados.

Según el contexto del caso práctico, una de las necesidades identificadas que nuestro sistema deberá cubrir es la siguiente:

- Analizar la gestión de las llamadas al teléfono de urgencias 112 de Cataluña desde diferentes perspectivas.

Otra forma de definir los requerimientos es plantear las preguntas a las que nuestro sistema deberá responder, como, por ejemplo, la siguiente:

- ¿En qué fechas se ha detectado un mayor número de llamadas al teléfono de urgencias 112?

El estudiante, en este punto, deberá completar la definición de los requerimientos. Para ello tendrá que identificar otras necesidades y plantear otras preguntas a las que el sistema deberá responder.

## 2. Análisis de las fuentes de datos

En este apartado se deben revisar las fuentes de datos proporcionadas e identificar el tipo de información que contienen, su formato y la volumetría de los datos que deben ser cargados.

A continuación, se muestra un ejemplo:

- **poblacion\_9687bsc.csv.** Cifras de la población española por provincia a 1 de enero de 2020.

Este fichero plano desde el origen viene con las siguientes características:

- formato: CSV
- primera línea con etiquetas de los campos
- separador de campos: punto y coma (;)

| Nombre del campo | Tipo     | Ejemplo            |
|------------------|----------|--------------------|
| Edad simple      | Texto    | Total              |
| Provincias       | Texto    | 02 Albacete        |
| Sexo             | Texto    | Ambos sexos        |
| Periodo          | Texto    | 1 de enero de 2020 |
| Total            | Numérico | 389.830            |

Observaciones:

- El campo «Edad simple» siempre tiene el mismo valor («Total»).
- El campo «Sexo» siempre tiene el mismo valor («Ambos sexos»).
- El campo «Periodo» siempre tiene el mismo valor («1 de enero de 2020»).

Total: 53 registros.

### Estimación de la volumetría

La estimación del volumen de datos en el almacén de datos para cargar el fichero anteriormente analizado sería la siguiente:

| Fuente de datos       | Datos                                      |
|-----------------------|--|
| poblacion_9687bsc.csv | 53 registros<br>× 5 valores<br>= 265 datos |
| ..                    |  |
| Total de datos        | ..   |

El estudiante en este punto deberá completar la definición del resto de fuentes proporcionadas.

### 3. Análisis funcional

En el análisis funcional se debe proponer el tipo de arquitectura para la factoría de información que mejor se adecue al caso de estudio. Para ello se consideran los requisitos funcionales y se establece la prioridad entre exigible (E) o deseable (D). En el contexto de esta actividad, los requerimientos exigibles son aquellos que demanda el enunciado, y los deseables, los que complementan la actividad.

Por otro lado, en términos de la escala de prioridades asignamos una prioridad de 1 a 3, siendo 1 completamente prioritario para la actividad y 3 no prioritario.

En la tabla se describen algunos de los requerimientos funcionales teniendo en cuenta las consideraciones del caso práctico que estamos desarrollando:

| # | Requerimiento  | Prioridad | Exigible/deseable |
|---|--|-----------|-------------------|
| 1 | Extraer de forma adecuada la información de las fuentes de datos (considerando solo la información relevante). | 1         | E                 |
| 2 | Crear un almacén de datos.   | 1         | E                 |

El estudiante en este punto deberá completar la tabla de los requerimientos funcionales, asignar las prioridades y elegir la arquitectura funcional que considere más adecuada para el caso de estudio de todas las posibles y estudiadas en los módulos teóricos de la asignatura.

### 4. Diseño del modelo conceptual, lógico y físico del almacén de datos

Para el correcto desarrollo del almacén de datos es preciso definir los hechos (*facts*), las dimensiones del análisis (*dimensions*), las métricas y los atributos que nos permitan dar respuesta a las preguntas que se han definido en el análisis de los requerimientos.

#### 4.1. Diseño conceptual

Del análisis de las fuentes de datos se determina que uno de los hechos que se deben analizar son las llamadas al 112, es decir, la tipología de las llamadas al teléfono de urgencias 112 de Cataluña.

El análisis realizado determina el diseño de la siguiente tabla de hechos:

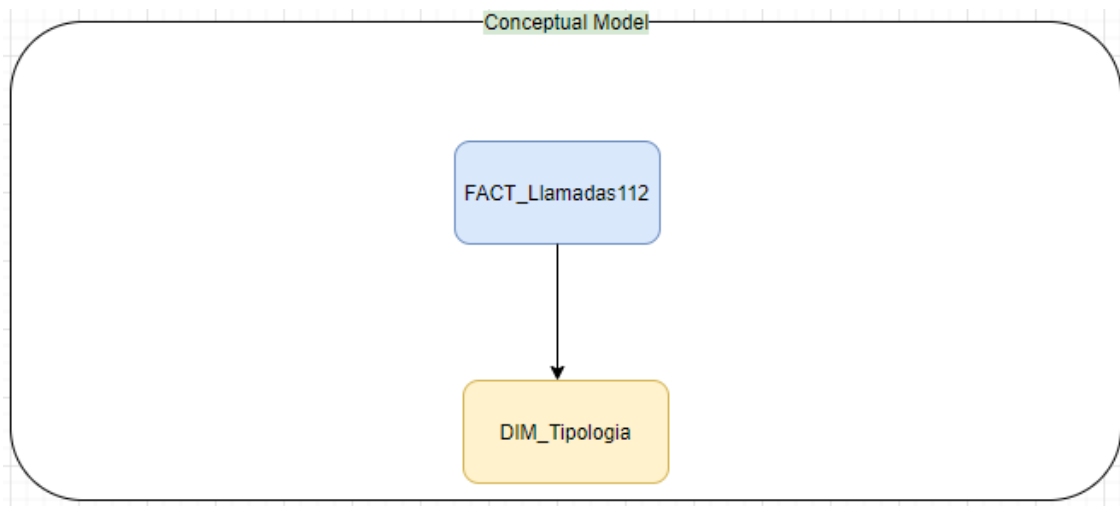
| Tabla de hechos  | Descripción  |
|------------------|--|
| FACT_Llamadas112 | Gestión de las llamadas al teléfono de urgencias 112 de Cataluña |

Las métricas que se identifiquen para la tabla de hechos podrán ser analizadas desde diferentes perspectivas a través de las dimensiones. Por ejemplo, una de estas dimensiones será la tipología, como vemos a continuación:

| Dimensiones | Descripción                                   |
|-------------|---|
| Tipología   | Tipología de las llamadas sobre la incidencia |

A partir de las dimensiones y de la tabla de hechos identificados, se construye el modelo conceptual, siendo tanto las dimensiones como los hechos entidades independientes que forman parte de nuestro modelo de estrella.

El **diseño conceptual parcial** para la tabla de hechos y la dimensión identificada de ejemplo sería la siguiente:



## 4.2. Diseño lógico

Una vez obtenido el modelo conceptual del almacén de datos, se detallan las métricas de cada una de las tablas de hechos y sus atributos para diseñar el modelo lógico.

A continuación, se muestra, a modo de ejemplo, una tabla con alguna de las métricas identificadas en el diseño conceptual de la tabla de hechos FACT\_Llamadas112:

| Tabla de hechos  | Métricas       | Descripción   |
|------------------|----------------|---|
| FACT_Llamadas112 | Llamada<br>... | Número de llamadas operativas de incidencias<br>... |

Seguidamente, se detallarán los atributos que contiene cada tabla de dimensiones. Los atributos, junto con las métricas, permitirán realizar los diferentes análisis de los requerimientos planteados.

En esta guía se muestran algunos atributos descriptores, a modo de ejemplo, de la dimensión identificada para la tabla de hechos:

| Dimensiones   | Atributos descriptores    |
|---------------|---------------------------|
| DIM_Tipologia | Nombre de la tipología... |

### 4.3. Diseño físico

Una vez determinadas las tablas de hechos, las dimensiones, las métricas y los atributos que existen en nuestro modelo, podemos decidir las claves primarias de las dimensiones y las claves foráneas que deben definirse en el modelo físico. En este paso también es necesario tener en cuenta el tamaño adecuado de los atributos de las tablas del modelo, como, por ejemplo, qué longitud tiene una cadena o si los valores numéricos contienen decimales.

En este punto se debería reflexionar acerca de las consideraciones previas que hay que tener en cuenta para el diseño físico, como en el siguiente ejemplo:

- El diseño físico debe estar orientado a generar un buen rendimiento en el procesamiento de las consultas.

Como es lógico, primero se crean las tablas de dimensiones y, posteriormente, las de hechos, ya que contienen atributos referenciales a las de dimensiones. De esta forma, creamos cada una de las tablas de nuestro almacén de datos.

A continuación, se muestra, a modo de ejemplo, una posible solución del diseño físico de una de las dimensiones identificadas para la FACT\_Llamadas112.

#### Dimensiones

- DIM\_Tipologia.** Tipología de llamada.

| Nombre del campo  | Tipo     | Tamaño |
|-------------------|----------|--------|
| pk_tipologia (PK) | Numérico | 4      |
| ...               | ...      | ...    |

#### Tablas de hechos

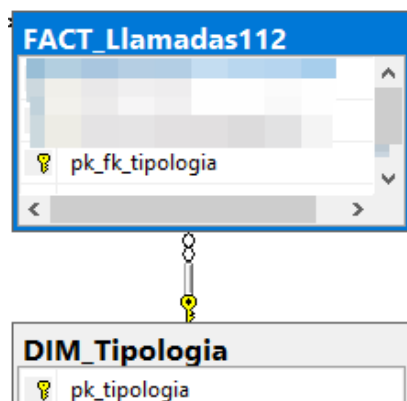
El modelo físico de las tablas de hechos consistirá en crear tablas cuyos campos serán las métricas, los atributos y los atributos referenciales definidos en el modelo conceptual y en el lógico. Para crear los atributos referenciales en las tablas de hechos, se definen como *claves foráneas* las primarias de las dimensiones con las que están relacionadas, siguiendo el diagrama de estrella definido.

- FACT\_Llamadas112.** Es la tabla física que contendrá la información que permitirá realizar el análisis de la gestión de las llamadas al teléfono de urgencias 112 de Cataluña. Tendrá que estar presente el siguiente campo, entre otros:

| Nombre del campo | Tipo     | Tamaño |
|------------------|----------|--------|
| pk_fk_tipologia  | Numérico | 4      |
| ...              | ...      | ...    |

Según lo definido en la guía de muestra, en la siguiente imagen se indica el diseño del modelo físico para la tabla de hechos.

**Observación:** en este modelo se muestra un ejemplo de una **solución parcial** que el estudiante deberá completar.



El estudiante, para realizar la entrega de la PRA1, y tomando como modelo la guía de muestra, deberá completar el análisis y el diseño de la forma que permita construir el almacén de datos para el análisis del impacto conductual de la COVID-19 sobre la población. Este va a cubrir las necesidades de los usuarios potenciales en el contexto del caso de estudio.

## Formato y fecha de entrega

La entrega de esta actividad debe realizarse enviando un único mensaje al buzón «Registro de AC» del apartado «Evaluación» del aula. Junto con el mensaje se enviará un único archivo en formato Word o PDF con la solución de la práctica. El nombre del archivo debe ser la composición del nombre de usuario con el sufijo «\_DUBDA\_PRA1.doc». Por ejemplo, si el nombre de usuario es «jmgarcia», entonces el archivo será «jmgarcia\_DUBDA\_PRA1.doc».

**La fecha máxima de entrega es el 14/04/2021 a las 23:59 horas.**