

# **Caso práctico: Almacén de datos para el análisis del impacto ambiental y consumo energético derivados de la actividad económica**

**Autor: David Díaz Arias**  
**Carles Llorach i Rius**  
**Joan Ramon Regolf Queralt**  
**Víctor Ruíz Marqués**  
**María Nerea Sevilla Marchena**

## **Índice**

- Introducción
- Contexto
- Enunciado
- Programas
- Bibliografía

## Introducción

El caso «**Almacén de datos para el análisis del impacto ambiental y consumo energético derivado de la actividad económica**» está creado para practicar el diseño e implementación del almacén de datos como sistema de almacenamiento para el análisis de datos.

El diseño, desarrollo e implantación de un sistema de Data Warehouse (DW) en cualquier organización supone llevar a cabo un proyecto que puede durar meses, o incluso años, en función del alcance del proyecto, de la naturaleza y el grado de madurez de la organización, así como de la participación de equipos multidisciplinares que van implementando diferentes proyectos en un proceso de mejora continua del almacén.

El objetivo de este caso no es desarrollar un almacén de datos que dé respuesta a todas las necesidades, sino entender y utilizar las metodologías para desarrollar este tipo de proyectos en un contexto real, pasando por todas las fases que comprenden los proyectos de esta tipología:

1. **Análisis, diseño e implementación:** consiste en desarrollar e implementar un almacén de datos que permita la gestión de la información disponible.
2. **Carga:** implica diseñar e implementar los procesos de carga de datos necesarios para disponer de información en el almacén de datos implementado en la etapa anterior.
3. **Explotación:** el objetivo es llevar a cabo un análisis multidimensional de dicha información que permita su explotación.

Con el fin de poder desarrollar un proyecto lo más específico posible, el estudiante tendrá que afrontar el reto de desarrollar un almacén de datos que solo describe parte de los servicios que se pueden ofrecer, basándose en datos tratados en el caso y que formarían parte de un sistema real.

A partir del contexto que se describe a continuación, el estudiante deberá adquirir un conocimiento básico del entorno tecnológico, de las necesidades existentes, y definir una propuesta adecuada que responda a ellas.

Mediante el desarrollo del caso, el estudiante se va a encontrar con los problemas, dudas y dificultades que se plantean en un proyecto de estas características.

## Contexto

Nuestro actual estilo de vida se basa en un modelo energético centrado en el consumo de combustibles fósiles (petróleo, gas y carbón). Esto conlleva la emisión a la atmósfera de importantes cantidades de gases de efecto invernadero (GEI), principales causantes del cambio climático.

La famosa «huella de carbono» mide el impacto de este modelo económico que conlleva un uso irracional de la energía.

Desde hace tiempo se impulsan múltiples iniciativas para paliar dicho efecto: protección del aire y el clima; gestión de aguas residuales y residuos; protección de la biodiversidad y los paisajes...

También se abren paso fuentes energéticas alternativas (las renovables) más respetuosas con nuestro planeta, que nos aproximan a los objetivos de desarrollo sostenible (ODS) fijados por la ONU para 2030.



[https://www.un.org/sustainabledevelopment/es/wp-content/uploads/sites/3/2016/08/spanish\\_SDG\\_17goals\\_poster\\_all\\_languages\\_with\\_UN\\_emblem.png](https://www.un.org/sustainabledevelopment/es/wp-content/uploads/sites/3/2016/08/spanish_SDG_17goals_poster_all_languages_with_UN_emblem.png)

Una oportunidad para que los países y sus sociedades emprendan un nuevo camino con el que mejorar la vida de todos, sin dejar a nadie atrás. Destacamos la defensa del medio ambiente, el diseño de nuestras ciudades, una producción y consumo responsables... Se trata de desvincular el crecimiento económico de la degradación medioambiental, aumentar la eficiencia de recursos y promover estilos de vida sostenibles.

## Usuarios potenciales

Como fase inicial del diseño del sistema de análisis de datos ambientales y de consumo energético identificaremos los requerimientos de los usuarios potenciales, para que el sistema los pueda tener en cuenta al dar respuesta a sus necesidades y generar información que les pueda ser útil.

Los usuarios finales que harán uso del sistema son:

- Las **administraciones**. Con la información proporcionada por el sistema integrado, los gobiernos y administraciones dispondrán de información para establecer sus políticas de actuación, controlar el impacto de la actividad económica por zonas geográficas, hacer el seguimiento de las iniciativas para revertir el impacto, procurar la implementación de programas de estímulo para fuentes de energía sostenibles; facilitar el establecimiento de medidas regulatorias que estimen oportunas y mucho más.
- Las **empresas y organizaciones**. El sistema integrado de datos permitirá extraer información útil relativa a los consumos energéticos en su ámbito territorial, además de contribuir a la reducción de emisiones contaminantes, mejorando su imagen y la calidad de sus servicios. Así podrán realizar comparativas y tomar decisiones comerciales mejor orientadas.
- Las **organizaciones ecologistas**. Con el sistema integrado podrían disponer de información de calidad para orientar sus campañas de sensibilización a aquellos sectores menos sensibles, a la vez que podrían poner en marcha campañas específicas maximizando su impacto.
- La **población** en general, ya que puede consultar los datos y valorar la eficacia de las políticas aplicadas, el acierto de las iniciativas comerciales, la constatación de ciertos comportamientos colectivos y, finalmente, les permite tomar consciencia y actuar en beneficio propio y del planeta.

## Fuentes de datos

Uno de los objetivos de este caso de estudio es integrar las diversas fuentes de datos (y formatos) proporcionadas para poder realizar diferentes tipos de análisis. En concreto, disponemos de información detallada de consumos energéticos, emisiones, inversión en protección ambiental y datos económicos.

La relación de ficheros *Open Data* que utilizaremos para la carga inicial son los siguientes:

| Nombre fichero  | Descripción   | Fuente  |
|---|---|---|
| 02002.xls   | Evolución de la inversión en protección ambiental por tipo de equipo e instalación, ámbito medioambiental y sector de actividad económica (en euros).   | www.ine.es<br><a href="https://www.ine.es/jaxi/Tabla.htm?path=/t26/p070/p01/serie/I0/&amp;file=02002.px&amp;L=0">https://www.ine.es/jaxi/Tabla.htm?path=/t26/p070/p01/serie/I0/&amp;file=02002.px&amp;L=0</a>                           |
| Countries.json  | Contiene los nombres de los países (nombres cortos oficiales en español, inglés y francés) en orden alfabético (ISO 3166-1) y los elementos de código ISO 3166-1-alpha-2 y alpha-3 en formato JSON. | Github Gist<br><a href="https://gist.github.com/brenes/1095110/raw/f8eeb4a7efb257921e6236ef5ce2dbc13c50c059/paises.csv">https://gist.github.com/brenes/1095110/raw/f8eeb4a7efb257921e6236ef5ce2dbc13c50c059/paises.csv</a>              |
| env_bio1.tsv.gz   | Biodiversity - Natura 2000 protected areas.   | Eurostat<br><a href="https://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?file=data/env_bio1.tsv.gz">https://ec.europa.eu/eurostat/estat-navtree-portlet-prod/BulkDownloadListing?file=data/env_bio1.tsv.gz</a> |
| WorldEnergyBalancesHighlights_final.xlsx<br>(Tratar solo la pestaña "TimeSeries_1971-2019") | Contiene variables relevantes del balance energético mundial.   | International Energy Agency (IEA)<br><a href="https://www.iea.org/subscribe-to-data-services/world-energy-balances-and-statistics">https://www.iea.org/subscribe-to-data-services/world-energy-balances-and-statistics</a>              |
| DataGeneric.xml   | Municipal waste, Generation and Treatment.  | Organización para la Cooperación y el Desarrollo Económicos (OECD)<br><a href="https://stats.oecd.org/viewhtml.aspx?datasetcode=MUNW&amp;lang=en">https://stats.oecd.org/viewhtml.aspx?datasetcode=MUNW&amp;lang=en</a>                 |
| ODS.xlsx<br>(Tratar ambas pestañas)   | Objetivos de Desarrollo Sostenible y su relación con los ámbitos medioambientales.  | Fuente propia   |

Se constata que los datos de consumo energético se recibirán anualmente y, por tanto, serán necesarias cargas incrementales para su integración en el Data Warehouse. El desarrollo de estos procesos futuros queda fuera del alcance de esta actividad.

## Enunciado

### 1. PRA1: Análisis y diseño del Data Warehouse.

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 permita y facilite el análisis del impacto ambiental y consumo energético de la actividad económica.

Mediante la metodología de diseño de un Data Warehouse (DW) propuesta en la asignatura, el estudiante debe llevar a cabo:

- **El análisis de requerimientos:** como resultado se generará un documento que describa las preguntas para las cuales los usuarios potenciales esperan respuesta del sistema.
- **El análisis de fuentes de datos:** se deben revisar las fuentes de datos proporcionadas, qué tipo de información contienen, cuál es su formato y qué cantidad representan para la carga inicial.
- **El análisis funcional:** se debe proponer el tipo de arquitectura para la factoría de información que mejor se adecue al proyecto (por ejemplo, si es necesario un *data mart* operacional o una estructura de carga intermedia).
- **El diseño del modelo conceptual, lógico y físico del almacén de datos:** se deben identificar y diseñar las tablas de hechos, las dimensiones y atributos que describen la información.

Para este apartado, el estudiante debe preparar un documento (solución PRA1) en que se detalle cada una de las secciones anteriores.

Se deberá tener en cuenta que para el desarrollo del DW es preciso definir correctamente los hechos (*facts*), las dimensiones de análisis (*dimensions*) y los atributos que nos permitan tener el nivel de granularidad suficiente para la medida y presentación de los objetivos que se definan en el análisis de requerimientos.

### 2. PRA2: Carga de los datos

A partir de la solución oficial de la primera práctica (PRA1), el estudiante deberá diseñar, implementar y ejecutar los procesos de extracción, transformación y carga de los datos (ETL) de las fuentes de datos proporcionadas.

Veamos las tareas que deberá llevar a cabo el estudiante:

- Identificación de los procesos de extracción, transformación y carga de los datos (ETL) hacia el almacén de datos.

- Diseño y desarrollo de los procesos ETL mediante las herramientas de diseño proporcionadas.
- Implementación con los trabajos de los procesos ETL para su carga efectiva planificada.

### 3. PRA3: Explotación de los datos

Tras la carga efectiva de los datos en el almacén (PRA2), se debe implementar un cubo multidimensional para explotar la información como apoyo a la toma de decisiones de los usuarios potenciales.

La finalidad del modelo *multidimensional on line analytical processing* (MOLAP) será responder a las preguntas definidas en el análisis de requerimientos.

## Programas

Para el presente caso, la UOC proporciona un entorno VDI con todo el *software* preconfigurado con las siguientes características:

- Sistema operativo: Windows 10
- Base de datos: Base de datos remota Microsoft SQL Server 2016 accesible desde cliente mediante SQL Server Management Studio 17
- Herramienta para la creación de cubos OLAP: Visual Studio 2017
- Herramienta de diseño de ETLs: Spoon – Pentaho Data Integration 9.0
- Herramienta de creación de informes: PowerBI Desktop

## Bibliografía

*Material de la asignatura Data Warehouse de la UOC.*

**Kimball, R. (2013).** *The Data Warehouse Toolkit*. 3.<sup>a</sup> ed. Nueva York (NY): John Wiley & Sons Inc.

**Inmon, W.H., Imhoff, C. y Sousa, R. (2001).** *Corporate Information Factory*. Nueva York (NY): John Wiley & Sons Inc.

**Inmon, W.H. (2005).** *Building the Data Warehouse*. Nueva York (NY): John Wiley & Sons Inc.

**Inmon, W.H., Strauss, D. y Neushloss, G. (2010).** *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. The Morgan Kaufman Series in Data Management Systems. Burlington (MA): Morgan Kaufmann Publishers.

**Krish, K. (2013).** *Data Warehousing in the Age of Big Data*. The Morgan Kaufmann Series on Business Intelligence. Burlington (MA): Morgan Kaufmann Publishers.

### *Enlaces a internet*

#### **MSDN Analysis Services tutorial:**

<https://docs.microsoft.com/es-es/analysis-services/analysis-services-tutorials-ssas?view=asallproducts-allversions>

#### **Tutorial Pentaho Data Integration:**

<http://wiki.pentaho.com/display/EAI/Latest+Pentaho+Data+Integration+%28aka+Kettle%29+Documentation>