

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

Solución PRA1–Análisis y diseño del *data warehouse*

A partir del análisis del contexto del caso y de las fuentes de datos disponibles, el estudiantado deberá diseñar y proponer un almacén de datos que permita el análisis del impacto ambiental y del consumo energético derivados de la actividad económica.

ÍNDICE

1.	ANÁLISIS DE LOS REQUERIMIENTOS	2
2.	ANÁLISIS DE LAS FUENTES DE DATOS	3
	ESTIMACIÓN DE VOLUMETRÍA	7
3.	ANÁLISIS FUNCIONAL	8
4.	DISEÑO DEL MODELO CONCEPTUAL, LÓGICO Y FÍSICO DEL ALMACÉN DE DATOS	10
	DISEÑO CONCEPTUAL	10
	DISEÑO LÓGICO	14
	DISEÑO FÍSICO	16

1. Análisis de los requerimientos

El análisis de los requerimientos se basa en identificar las necesidades específicas que tiene una organización particular respecto al análisis de la información. Normalmente, en esta fase, se debe ser previsor y pensar más allá de las necesidades actuales para poder cubrir las futuras.

La necesidad principal de la organización encargada del análisis del impacto ambiental y del consumo energético es disponer de la información integrada para su análisis y su posterior difusión mediante las herramientas de inteligencia de negocio. Estas ayudarán a facilitar la toma de decisiones a todos los usuarios potenciales para garantizar el cumplimiento de los siguientes objetivos:

- Realizar el seguimiento de ciertas mediciones relacionadas con políticas medioambientales, orientadas a un desarrollo sostenible.
- Analizar el balance energético mundial de los países productores y consumidores de energía.

El diseño de un almacén de datos de un proyecto incluye la creación e implementación de un modelo dimensional o multidimensional, el diseño y la implementación de procesos ETL y del modelo OLAP y, por último, el diseño de las consultas establecidas en el enunciado.

A continuación, se indica la información necesaria identificada:

1. Analizar las mediciones medioambientales desde diferentes perspectivas:

- por fecha,
- por región,
- por sector de actividad económica,
- por tipo o equipo de instalación,
- por ODS (objetivos de desarrollo sostenible) o
- por unidad de medida.

2. Analizar el balance energético mundial desde diferentes perspectivas:

- por fecha,
- por país,
- por categoría de energía (producto) o
- por unidad de medida.

Si se tiene en cuenta toda esta información, el sistema podrá responder a múltiples preguntas y de esta manera conseguirá cubrir las necesidades de los usuarios potenciales.

A continuación, se indican de forma específica las preguntas que, como mínimo, el sistema debe ser capaz de responder:

- Realizar un análisis evolutivo de la producción en ktoe (tonelada equivalente –energía–) por categoría de energía (producto) de los Estados Unidos.
- Análisis del *top five* de países con unas áreas protegidas mayores, tanto marinas como terrestres.
- Identificar aquellos países con mayor impacto sobre el ODS «Ciudades y comunidades sostenibles».
- Análisis del promedio de inversión de los años 2014 al 2018 en cuanto al ámbito medioambiental de la gestión de aguas residuales en la comunidad autónoma de la Comunidad Valenciana.
- Comparativa de los países de Alemania, Francia, España e Italia en cuanto a área protegida en el año 2019.

2. Análisis de las fuentes de datos

En este apartado se deben revisar las fuentes de datos proporcionadas, qué tipo de información contienen, cuál es su formato y qué datos deben ser cargados. Véase a continuación un análisis detallado para cada tipo de formato.

- 1) **02002.xlsx**. Encuesta del gasto de la industria en protección ambiental. Serie 2008-2018. Evolución de la inversión en protección ambiental por tipo de equipo e instalación, ámbito medioambiental y sector de actividad económica.

A continuación, se analizan los campos del fichero .xlsx (Excel):

Nombre campo	Tipo	Ejemplo
Periodo	Numérico	2018
Sector de actividad económica	Texto	«C. industria manufacturera»
Tipo de equipo e instalación	Texto	«INVERSIÓN EN EQUIPOS E INSTALACIONES INDEPENDIENTES»
Ámbito medioambiental	Texto	«Protección del aire y el clima»
Comunidad autónoma	Texto	«Asturias, Principado de»
Inversión	Numérico	6.280.443

Observaciones:

- Las columnas desde la E hasta la U corresponden a la inversión de cada comunidad autónoma.
- La columna V contiene el agregado de *Total nacional*.
- La unidad de la inversión son euros.

Total de registros: 154 registros.

- 2) **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.

La estructura del fichero es la siguiente:

Nombre campo	Tipo	Ejemplo
nombre	Texto	«España»
name	Texto	«Spain»
nom	Texto	«Espagne»
iso2	Texto	«ES» (código de país de 2 letras)
iso3	Texto	«ESP» (código de país de 3 letras)
phone_code	Texto	«34»

Total de registros: 246.

- 3) **env_bio1.tsv.** Contiene datos de la biodiversidad, información de áreas protegidas tanto marinas como terrestres.

A continuación, se analizan los campos del fichero .tsv (*Tab Separated Values* – Valores separados por tabulaciones):

Nombre campo	Tipo	Ejemplo
areaprot	Texto	«AREA_KM2»
geo\time	Texto	«AT»
Year	Numérico	2019
Value	Numérico	83.944

Terminología y aclaraciones:

- El campo «areaprot» (área protegida) contiene los siguientes valores:
 - MPA_KM2: Marine Protected Area (en km2).
 - TPA_KM2: Terrestrial Protected Area (en km2).
 - AREA_KM2: MPA_KM2 + TPA_KM2.
 - TPA_PC: Terrestrial Protected Area (en %).
- El campo geo\time contiene el código de país de 2 letras (código ISO 3166-1-alpha-2).

Total de registros: 121 registros.

4) WorldEnergyBalancesHighlights_final.xlsx. Contiene todos aquellos aspectos relevantes del balance energético mundial.

A continuación, se analizan los campos del fichero .xlsx (Excel):

Nombre campo	Tipo	Ejemplo
Country	Texto	«Australia»
Product	Texto	«Renewables and waste»
Flow	Texto	«Production (ktoe)»
Year	Numérico	1971
Value	Numérico	4.554

Observaciones:

- Solo se debe tratar la hoja «TimeSeries_1971-2019».
- La primera fila de esta hoja anterior es informativa y no se debe tratar.
- Las hojas «Contents», «Definitions», «IEA Energy balance explained» y «Interactive_PivotChart» no se deben tratar, pero se pueden consultar para clarificar los datos.
- Las columnas desde la G hasta la BC corresponden a los valores desde el año 1971 al 2019.
- Cada «Flow» tiene su propia unidad de medida. Por ejemplo, «Electricity output» está en GWh (gigavatio-hora), mientras que «Imports» está en ktoe (tonelada equivalente –energía–).
- El campo «Country» contiene los países de la OECD («Australia», «Austria»...), países de asociación («Brazil», «People's Republic of China», «India»...) y regiones («Non-OECD Total», «Africa», «Middle East», «World»...).

Total de registros: 6.050 registros.

5) DataGeneric.xml. Contiene datos de los residuos urbanos, su generación y el tratamiento.

A continuación, se analizan los campos del fichero .xml:

Nombre campo	Tipo	Ejemplo
COU (SeriesKey)	Texto	«AUS» (Australia)
VAR (SeriesKey)	Texto	«MUNICIPAL»
TIME_FORMAT (Attributes)	Texto	«P1Y»
UNIT (Attributes)	Texto	«TONNE»
POWERCODE (Attributes)	Texto	«3»
Time1 (Obs)	Numérico	1992

Value1 (Obs)	Numérico	12.000
OBS_STATUS1 (Obs)	Texto	«E»
Time2 (Obs)	Numérico	2000
Value2 (Obs)	Numérico	13.200
OBS_STATUS2 (Obs)	Texto	«E»
...

Terminología y aclaraciones:

- COU: corresponde al país en código oficial de 3 letras (iso3).
- VAR: diferentes variables a tratar según el país.
- TIME_FORMAT: indica el formato de tiempo («P1Y»: un año).
- UNIT: unidades de la medición («TONNE»: toneladas).
- POWERCODE: valor según la unidad de medición.
- TIME1: año específico de la medición.
- VALUE1: valor de la medición para el año indicado en *TIME1*.
- OBS_STATUS1: atributo que especifica «valor estimado» (*Estimated Value*):
 - Un país (COU) y variable (VAR) tienen [n] conjuntos de TIME/VALUE.
 - No todos los registros TIME/VALUE tienen el atributo OBS_STATUS.

A continuación, se muestra una captura del XML para clarificar sus campos:

```
<DataSet keyFamilyURI="https://stats-2.oecd.org/RestSDMX/sdmx.ashx/GetKeyFamily/MUNW">
  <KeyFamilyRef>MUNW</KeyFamilyRef>
  <Series>
    <SeriesKey>
      <Value concept="COU" value="AUS" />
      <Value concept="VAR" value="MUNICIPAL" />
    </SeriesKey>
    <Attributes>
      <Value concept="TIME_FORMAT" value="P1Y" />
      <Value concept="UNIT" value="TONNE" />
      <Value concept="POWERCODE" value="3" />
    </Attributes>
    <Obs>
      <Time>1992</Time>
      <ObsValue value="12000" />
      <Attributes>
        <Value concept="OBS_STATUS" value="E" />
      </Attributes>
    </Obs>
    <Obs>
      <Time>2000</Time>
      <ObsValue value="13200" />
      <Attributes>
        <Value concept="OBS_STATUS" value="E" />
      </Attributes>
    </Obs>
    <Obs>
      <Time>2007</Time>
      <ObsValue value="12873" />
    </Obs>
    <Obs>
      <Time>2008</Time>
      <ObsValue value="13096.5" />
    </Obs>
  </Series>
</DataSet>
```

Total de registros: 20.779 registros.

6) ODS.xlsx. Este fichero relaciona los objetivos de desarrollo sostenible con los ámbitos medioambientales. Contiene dos pestañas.

A continuación, se analizan los campos del fichero .xlsx (Excel):

Pestaña ODS:

Nombre campo	Tipo	Ejemplo
Objetivo	Numérico	13
Nombre	Texto	«Acción por el clima»
Descripción	Texto	«El cambio climático es un reto global que no respeta las fronteras nacionales»

Total de registros: 17 registros.

Pestaña Ambito_VAR_Flow-ODS:

Nombre campo	Tipo	Ejemplo
Codigo	Texto	«Protección del aire y el clima»
Ambito/VAR/Flow	Texto	Inversión en protección del aire y el clima en €
ODS principal	Numérico	13

Total de registros: 50 registros.

Observaciones:

- ODS es el acrónimo de los objetivos de desarrollo sostenible.
- El campo [ODS principal] de la 2.ª pestaña (Ambito_VAR_Flow-ODS) corresponde al campo [Objetivo] de la 1.ª pestaña (ODS).

Estimación de volumetría

En los proyectos de diseño de factoría de información corporativa existe una primera fase en la que se realiza una carga inicial y, *a posteriori*, una segunda fase para realizar las cargas incrementales de los datos nuevos que van llegando.

Una posible estimación del volumen de datos del almacén para la carga inicial de los datos sería la siguiente:

Fichero	Registros	Valores	Datos
02002.xlsx	154	22	3.388
Countries.json	246	6	1.476
env_bio1.tsv	121	11	1.331
WorldEnergyBalancesHighlights_final.xlsx	6.050	52	314.600
DataGeneric.xml	20.779	8	166.232
ODS.xlsx	17	3	51
	50	3	150
Total	27.417	105	487.228

3. Análisis funcional

A continuación, se propone el tipo de arquitectura para la factoría de información que mejor se adecua al proyecto. 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 pide el enunciado, y los deseables, los que complementan la actividad.

Además, en términos de la escala de prioridades, se asigna una prioridad del 1 al 3, siendo 1 completamente prioritario para la actividad y 3 no prioritario.

A continuación, se describen los requerimientos funcionales para el diseño de una factoría de información para la organización, teniendo en cuenta las consideraciones del enunciado:

#	Requerimiento	Prioridad	Exigible / deseable
1	Se extraerá de forma adecuada la información de las fuentes de datos.	1	E
2	Se creará un almacén de datos.	1	E
3	Se cargará la información sobre el impacto ambiental y el consumo energético derivados de la actividad económica.	1	E
4	Se creará un modelo OLAP para consultas multidimensionales de los usuarios.	2	E
5	Se crearán los informes estáticos solicitados.	2	E
6	Se redactará un manual de carga de datos inicial e incremental.	3	D

Cabe comentar que, en un caso genérico real, se pueden encontrar también otros requerimientos funcionales, como los que se muestran a continuación:

- Análisis de viabilidad y análisis de riesgos.

- Creación de procesos de calidad de datos.
- Creación de *data marts* (si se analizan otras áreas).
- Creación de procesos de cargas incrementales.
- Creación de un repositorio de metadatos de gestión del almacén de datos, así como de los procesos ETL, que permita realizar la trazabilidad a lo largo del ciclo de vida de los datos.

Asimismo, dado que estos sistemas frecuentemente forman parte de la implementación de un sistema de inteligencia de negocio, la lista de requerimientos funcionales sería mucho mayor, como puede ser la administración de seguridad en cuanto a datos y usuarios.

En términos de la arquitectura funcional, existen los siguientes elementos:

- Las fuentes de datos de las que se dispone son las siguientes:
 - 3 ficheros Excel (xlsx).
 - 1 fichero en formato JSON.
 - 1 fichero en formato TSV.
 - 1 fichero en formato XML.
- La arquitectura de la factoría de información puede estar formada por varios elementos alojados en la misma máquina:
 - *Staging area* (opcional): en el caso de tener múltiples fuentes (ficheros, bases de datos, servicios RSS, etc.), es conveniente cargarlas para consolidar la información en una estructura de carga intermedia que puede ser creada en la misma base de datos.

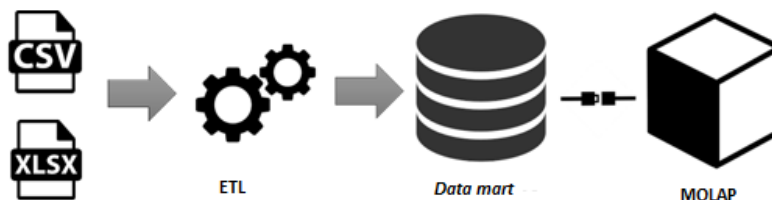
Esta área del DW también puede servir para entender, simplificar y consolidar los procesos ETL.

- *Data mart* del impacto ambiental y consumo energético derivados de la actividad económica: al centrarnos en una única área temática, es más correcto considerar que se está creando un *data mart* en lugar de un almacén de datos corporativo.
- MOLAP: a partir de la información del *data mart*, se creará un cubo multidimensional.

Según lo comentado anteriormente, se podría elegir entre dos diseños para la arquitectura funcional. Por un lado, tenemos una arquitectura funcional que usa un área intermedia (*staging area*) y se crearía dentro de la misma base de datos, cuyos objetos se identificarán con un prefijo en los nombres. El siguiente gráfico resume los elementos de la arquitectura necesarios para esta actividad:



Por otro lado, también sería correcto utilizar una arquitectura sin área intermedia (*staging area*), que identifique las tablas intermedias en el *data mart* con un prefijo en el nombre, como, por ejemplo, «IN_nombre_tabla_intermedia».



En esta solución se propone un diseño que utiliza un área intermedia (*staging area*) que se crearía dentro de la misma base de datos, cuyos objetos se identificarán con un prefijo en los nombres (STG_).

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

Diseño conceptual

Para el correcto desarrollo del DW es preciso definir los hechos (*facts*), las dimensiones de análisis (*dimensions*), las métricas y los atributos que permitan tener el nivel de granularidad suficiente para la presentación de los objetivos. Estos se han definido en el análisis de requerimientos y de las fuentes de datos.

Del análisis de las fuentes de datos y de los requerimientos iniciales, se puede determinar que los hechos que se deben considerar son los siguientes:

- **Mediciones ambientales.** Hace referencia a ciertas mediciones ambientales relevantes para un desarrollo sostenible.
- **Balances energéticos.** Hace referencia al balance energético mundial de los países productores y consumidores de energía.

El **análisis de las mediciones ambientales** determina el diseño de la primera tabla de hechos, como se puede observar a continuación:

Tabla de hechos	Descripción
FACT_EnvironmentalMeasurements	Mediciones ambientales para un desarrollo sostenible

Las **métricas** de la tabla de hechos «FACT_EnvironmentalMeasurements» son las siguientes:

Métricas	Descripción
Value	Valor de la medición

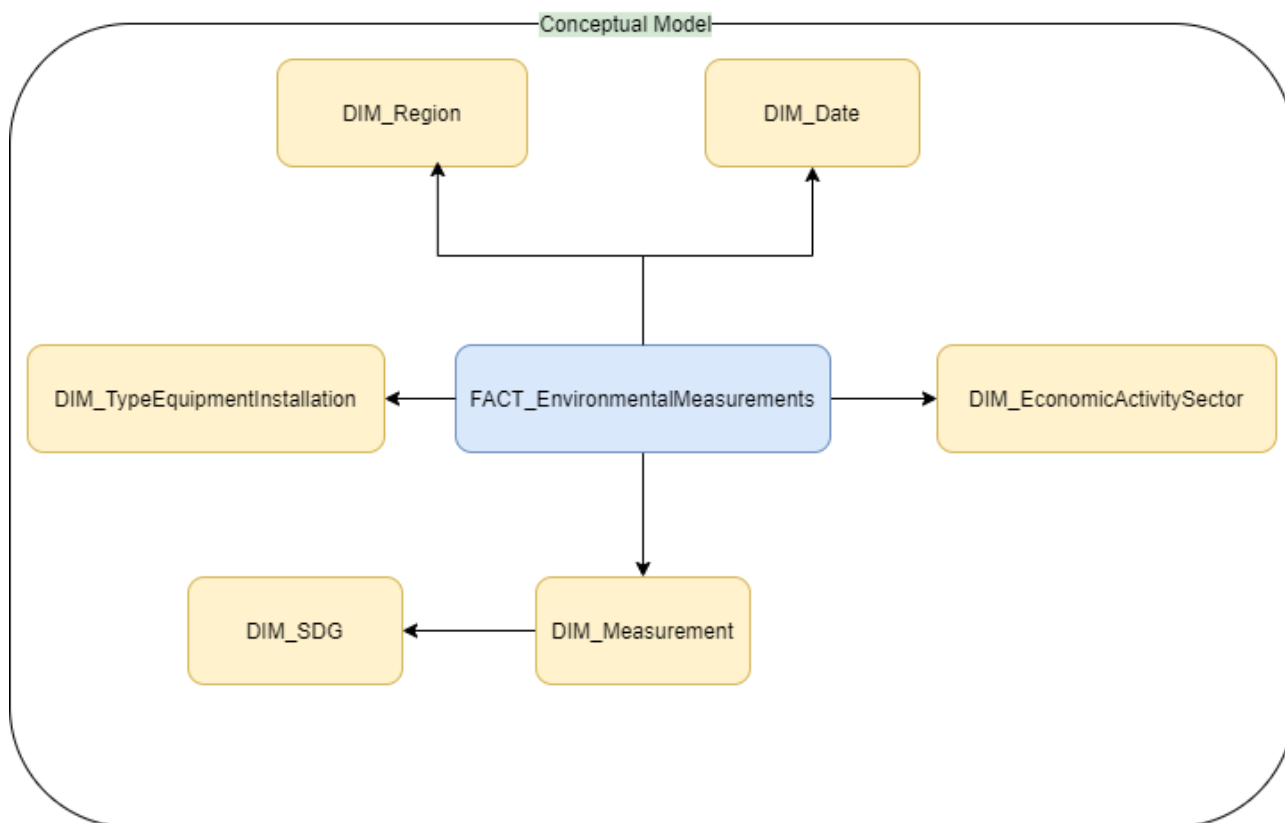
Las métricas de esta tabla de hechos podrán ser analizadas desde las diferentes perspectivas, a partir de las siguientes dimensiones:

Dimensiones	Descripción
Date	Fecha en la que se realiza la medición.
Region	Región o ámbito geográfico donde se localiza la medición.
Economic Activity Sector	Sector de la actividad económica al que se clasifica la medición.
Type of Equipment or Installation	Distinción del tipo de equipo o instalación de la medición.
SDG (Sustainable Development Goals)	Objetivos de desarrollo sostenible (ODS) con los que se relaciona la medición.
Measurement	Medición a tratar.

En las dimensiones con jerarquía se puede optar por las siguientes alternativas:

- **Diseño en copo de nieve:** se aplican las técnicas de normalización de las bases de datos para optimizar el espacio y eliminar las redundancias. Esto requiere la creación de nuevas tablas y relaciones, lo que empeora el rendimiento.
- **Diseño en estrella:** no se normalizan las bases de datos y se mantienen las jerarquías dentro de una misma tabla.

El diseño conceptual para esta tabla de hechos (FACT_EnvironmentalMeasurements) y sus dimensiones con un **diseño en copo de nieve** es el siguiente:



Este modelo considera las fuentes de datos siguientes:

- 02002.xlsx
- Countries.json
- env_bio1.tsv
- DataGeneric.xml
- ODS.xlsx

Para el análisis de los **balances energéticos mundiales** se identifica una segunda tabla de hechos, como la siguiente:

Tabla de hechos	Descripción
FACT_EnergyBalances	Balance energético mundial de los países productores y consumidores de energía.

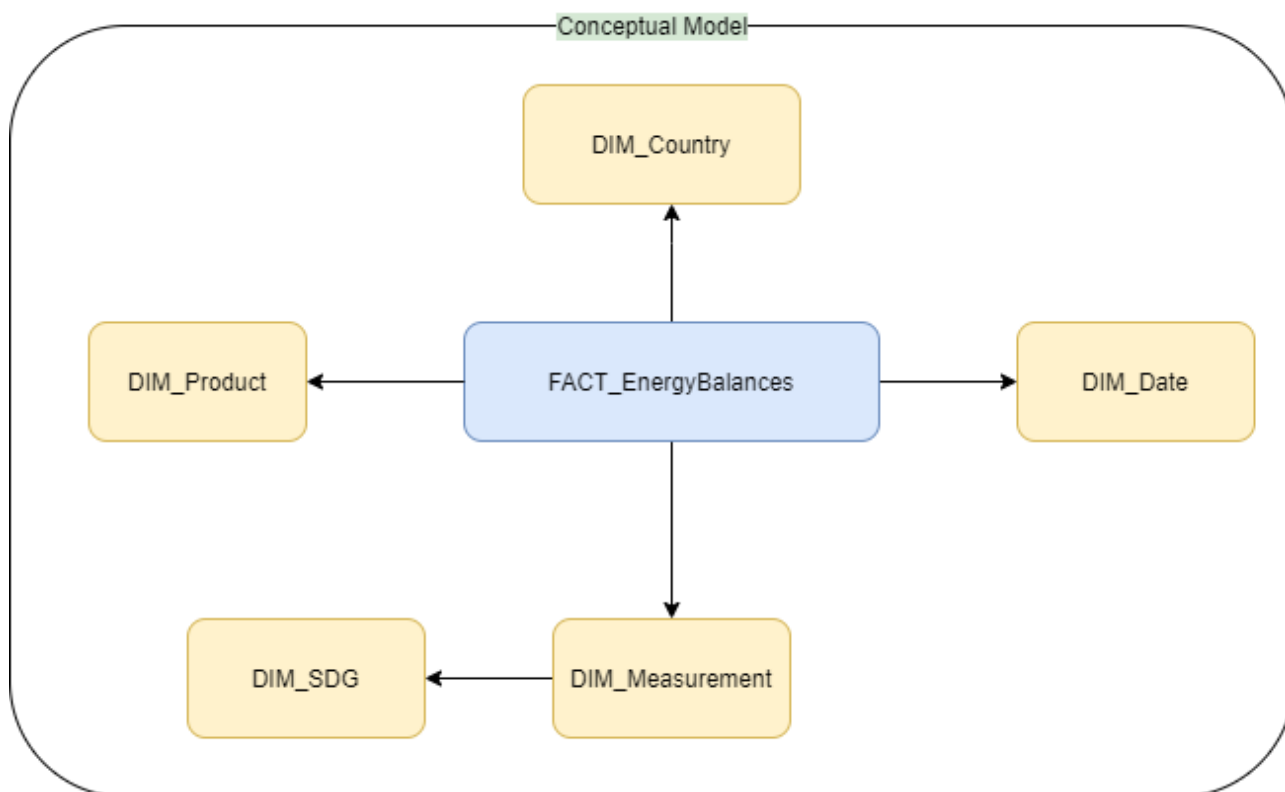
Esta tabla de hechos almacenará la siguiente **métrica**:

Métricas	Descripción
Value	Valor de la medición del balance energético.

Esta métrica puede ser analizada desde diferentes perspectivas a través de las siguientes dimensiones:

Dimensiones	Descripción
Date	Fecha en la que se realiza la medición del balance energético.
Country	País donde se localiza la medición del balance energético.
Product	Productos que intervienen en la medición del balance energético (ejemplos: carbón, gas natural, energía nuclear, electricidad...).
SDG (Sustainable Development Goals)	Objetivos de desarrollo sostenible (ODS) con los que se relaciona la medición.
Measurement	Medición del balance energético.

El diseño conceptual para esta tabla de hechos (FACT_EnergyBalances) y sus dimensiones con un **diseño en copo de nieve** es el siguiente:



Este modelo considera las fuentes de datos siguientes:

- Countries.json
- WorldEnergyBalancesHighlights_final.xlsx

Diseño lógico

Una vez obtenido el modelo conceptual del almacén de datos para el análisis del impacto ambiental, pasamos a realizar el diseño lógico del mismo.

Teniendo en cuenta que vamos a utilizar tecnología relacional y que el modelo de datos va a ser el multidimensional, pasamos a describir el modelo lógico en términos de tablas, atributos y claves primarias y foráneas.

El primer paso corresponde a identificar las métricas de cada hecho que va a representarse mediante una tabla de hechos. En nuestro caso, hemos identificado estos dos hechos y sus métricas:

Tabla de hechos	Métricas
FACT_EnvironmentalMeasurements	Value
FACT_EnergyBalances	Value

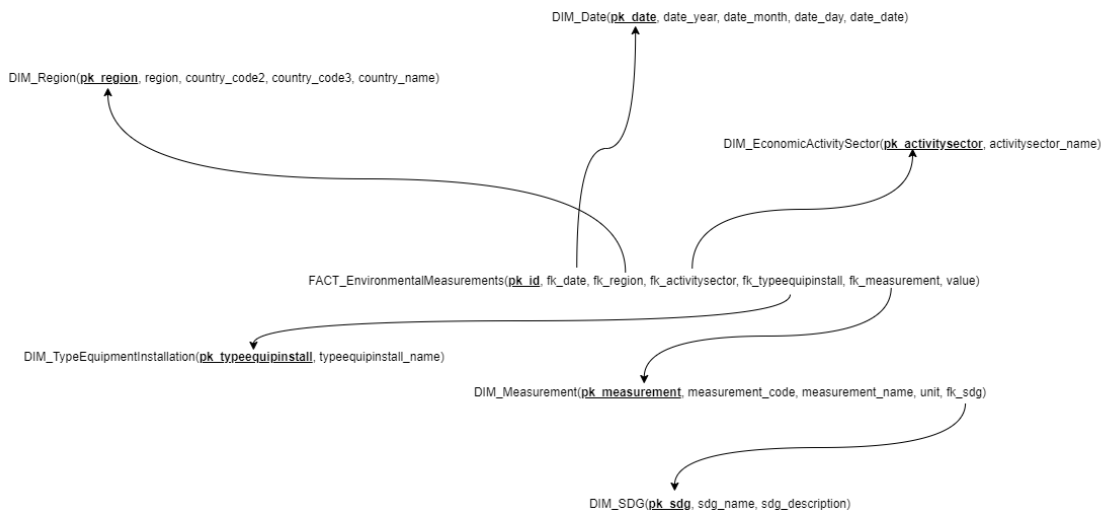
Después se detallan los atributos de las dimensiones de cada hecho. Específicamente, los atributos de las dimensiones de la tabla de hechos «FACT_EnviromentalMeasurements» se muestran en la siguiente tabla:

Dimensiones	Atributos
DIM_Date	Año, mes y día
DIM_Region	Comunidad autónoma o región, Codigo2digitos del país, Codigo3digitos del país y nombre del país (oficial en inglés)
DIM_EconomicActivitySector	Código, nombre
DIM_TypeEquipmentInstallation	Código, nombre
DIM_SDG	Código, nombre, descripción
DIM_Measurement	Código, nombre de la medición y unidad de medida

Estos serían algunos ejemplos para las mediciones ambientales:

Nombre fichero	Nombre medición	Unidad medida
02002.xlsx	Inversión en protección del aire y el clima en € (protección del aire y el clima)	euros
env_bio1.tsv	Área protegida marina (MPA_KM2)	km2
env_bio1.tsv	Área protegida terrestre (TPA_KM2)	km2
env_bio1.tsv	Área protegida terrestre (TPA_PC)	%
DataGeneric.xml	Municipal waste generated (MUNICIPAL)	Tonnes (toneladas)

La representación visual del **modelo lógico** para las mediciones ambientales sería la siguiente:



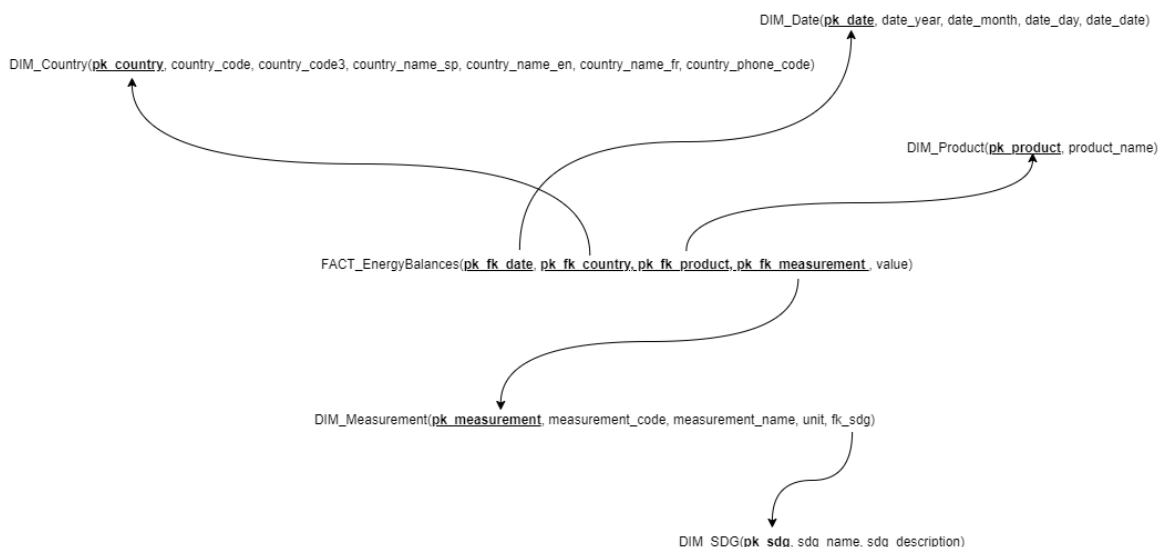
Procedemos igual con el hecho «FACT_EnergyBalances». Se muestran en la siguiente tabla los atributos de las dimensiones de esta tabla de hechos:

Dimensiones	Atributos
DIM_Date	Año, mes, día
DIM_Country	Codigo2digitos del país, Codigo3digitos del país, nombre del país (en español), nombre del país (en inglés), nombre del país (en francés) y prefijo telefónico
DIM_Product	Código, nombre
DIM_SDG	Código, nombre, descripción
DIM_Measurement	Código, nombre de la medición y unidad de medida

Estos serían algunos ejemplos para el balance energético mundial:

Nombre fichero	Nombre medición	Unidad medida
WorldEnergyBalancesHighlights_fi nal.xlsx	Production (ktoe)	ktoe
WorldEnergyBalancesHighlights_fi nal.xlsx	Electricity output (GWh)	GWh

La representación visual del **modelo lógico** para el balance energético mundial sería la siguiente:



En el caso de que unas fuentes de datos indiquen solamente el año, otras el nivel de mes y otras el nivel de día, conviene pedagógicamente tratar fechas completas, trasladando los datos al día 1 de enero de cada año tratado, en el caso de no disponer de los niveles de mes y día.

Así, para cargar la dimensión «DIM_Date», podemos hacerlo solo con las fechas que nos aparecen en las fuentes de datos o con todas las fechas desde el 01/01/1971 hasta el 01/01/2019.

Diseño físico

Una vez que se han determinado las tablas de hechos, las dimensiones, las métricas y los atributos que existen en el modelo lógico, podemos pasar a realizar el diseño físico, lo cual significa obtener una implementación del modelo lógico en términos del sistema gestor de bases de datos escogido.

Además, para el correcto diseño físico del almacén, se deben tener en cuenta los siguientes aspectos:

- El **sistema gestor de bases de datos** con el que vamos a trabajar implementará de una manera concreta los distintos elementos del modelo lógico.
- El ajuste del **diseño físico** a las particularidades de nuestro sistema gestor de bases de datos, con el fin de obtener buen rendimiento en el procesamiento de consultas.
- La **revisión periódica del diseño físico inicial**, para validar que continúa dando respuesta a las necesidades del cliente.

Puesto que utilizaremos SQL Server, y este es un sistema gestor de bases de datos relacional, en esta etapa deberemos tener en cuenta, entre otras cosas, la implementación de las claves primarias y foráneas en las tablas de hechos y en las de dimensiones.

En este paso, también es necesario tener en cuenta el tamaño adecuado de los atributos (por ejemplo, qué longitud tiene una cadena o si los valores numéricos contienen decimales).

Para ello, vamos a detallar los tipos de datos de cada campo que forma parte de las tablas de hechos y dimensiones.

Dado que el modelo de almacén está compuesto por más de una tabla de hechos (*facts*), también se deben revisar las dimensiones que se han definido en el diseño conceptual y en el lógico de cada *fact* y aplicar una visión conjunta del modelo. Esto permitirá definir las dimensiones comunes, como la «DIM_Date», y así simplificar el modelo final y conseguir un rendimiento óptimo en la ejecución de los análisis.

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

Dimensiones

Las dimensiones del modelo podrán estar referenciadas en las tablas de hechos utilizando sus claves primarias o, en inglés, *primary keys* (PK). El modelo físico de las dimensiones es el siguiente:

- **DIM_Date:** corresponde a la dimensión temporal del almacén. La dimensión temporal es común en todo el modelo diseñado y permite analizar los hechos desde un punto de vista temporal, como el análisis de tendencias o los evolutivos. Este tipo de análisis no se puede realizar si el modelo no cuenta con una dimensión de tiempo. Se puede ampliar con más información como, por ejemplo, los días de la semana, los festivos, los semestres, los cuatrimestres, etc.

Nombre campo	Tipo	Tamaño	Ejemplo
pk_date (PK)	Numérico	4	25
date_year	Numérico	4	2018
date_month	Numérico	2	1
date_day	Numérico	2	1
date_date	Fecha	10	1/1/2018

- **DIM_Region:** contiene los datos del ámbito geográfico.

Nombre campo	Tipo	Tamaño	Ejemplo
pk_region (PK)	Numérico	4	105
region	Texto	100	«Andalucía»
country_code2	Texto	2	«ES»
country_code3	Texto	3	«ESP»
country_name	Texto	100	«Spain»

- **DIM_EconomicActivitySector:** contiene los datos del sector de actividad económica.

Nombre campo	Tipo	Tamaño	Ejemplo
pk_activitysector (PK)	Numérico	4	1
activitysector_name	Texto	100	«C. industria manufacturera»

- **DIM_TypeEquipmentInstallation:** contiene los datos del tipo de equipo o instalación.

Nombre campo	Tipo	Tamaño	Ejemplo
pk_typeequipinstall (PK)	Numérico	4	1
typeequipinstall_name	Texto	100	«Inversión en equipos e instalaciones independientes»

- **DIM_SDG:** contiene los datos de los objetivos de desarrollo sostenible.

Nombre campo	Tipo	Tamaño	Ejemplo
pk_sdg (PK)	Numérico	4	13
sdg_name	Texto	50	«Acción por el clima»
sdg_description	Texto	500	«El cambio climático es un reto global que no respeta las fronteras nacionales»

- **DIM_Measurement:** contiene los datos de la medición.

Nombre campo	Tipo	Tamaño	Ejemplo
pk_measurement (PK)	Numérico	4	1
measurement_code	Texto	100	«Protección del aire y el clima»
measurement_name	Texto	200	«Inversión en protección ambiental»
Unit	Texto	25	«euros»
fk_sdg	Numérico	4	13

- **DIM_Country:** contiene los datos de los países.

Nombre campo	Tipo	Tamaño	Ejemplo
pk_country (PK)	Numérico	4	10
country_code2	Texto	2	«ES»
country_code3	Texto	3	«ESP»
country_name_sp	Texto	100	«España»
country_name_en	Texto	100	«Spain»
country_name_fr	Texto	100	«Espagne»
country_phone_code	Texto	5	«34»

- **DIM_Product:** contiene los datos de los productos.

Nombre campo	Tipo	Tamaño	Ejemplo
pk_product (PK)	Numérico	4	6
product_name	Texto	100	«Industry»

Tablas de hechos

La composición del modelo físico de las tablas de hechos consistirá en la creación de tablas cuyos campos serán las métricas, los atributos y los atributos referenciales definidos en el modelo conceptual y en el modelo 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 en copo de nieve definido.

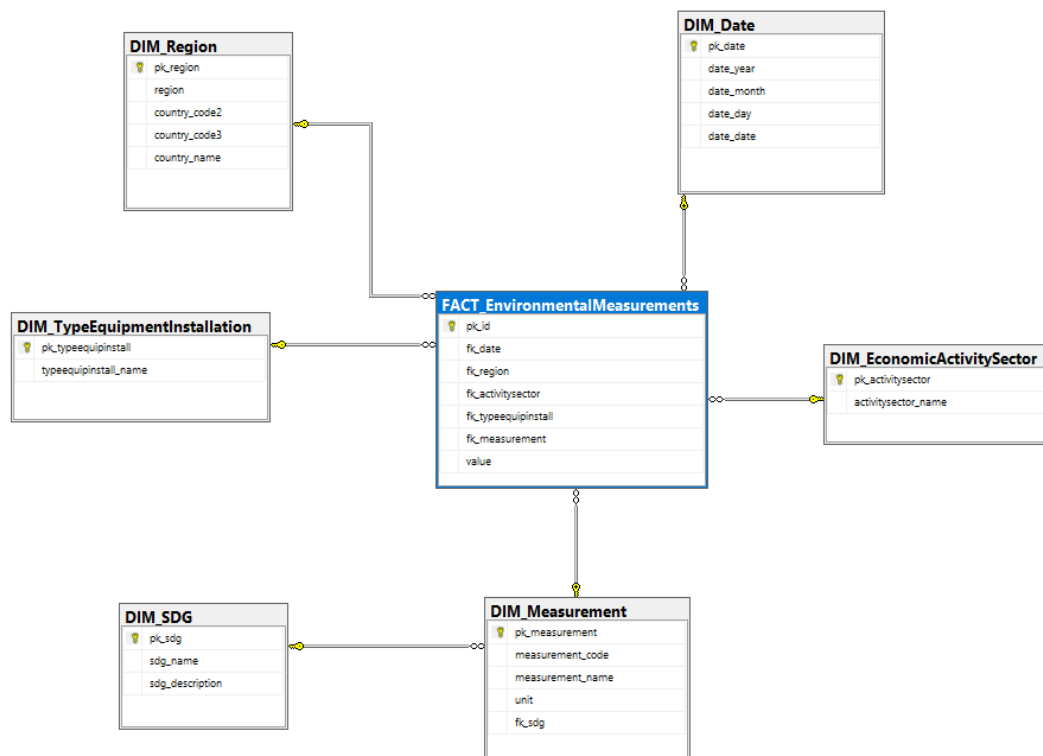
El modelo físico de las tablas de hechos del almacén de datos para el impacto ambiental y el consumo energético está compuesto de las siguientes tablas:

- **FACT_EnvironmentalMeasurements:** es la tabla física que contendrá la información que permitirá realizar el análisis de los datos de las mediciones ambientales para un desarrollo sostenible. Tendrá los siguientes campos:

Nombre campo	Tipo	Tamaño	Ejemplo
pk_id (PK)	Numérico	4	17
fk_date (FK)	Numérico	4	25
fk_region (FK)	Numérico	4	10
fk_activitysector (FK)	Numérico	4	1

fk_typeequipinstall (FK)	Numérico	4	1
fk_measurement (FK)	Numérico	4	1
value	Numérico	8	6.390.920

En la siguiente imagen se muestra una posible implementación del **diseño del modelo físico** para la tabla de hechos **FACT_EnvironmentalMeasurements**:



- **FACT_EnergyBalances**: es la tabla física que contendrá la información que permitirá realizar el análisis del balance energético mundial de los países productores y consumidores de energía. Tendrá los siguientes campos:

Nombre campo	Tipo	Tamaño	Ejemplo
pk_fk_date (PK y FK)	Numérico	4	25
pk_fk_country (PK y FK)	Numérico	4	10
pk_fk_product (PK y FK)	Numérico	4	1
pk_fk_measurement (PK y FK)	Numérico	4	1
value	Numérico	8	3.835

En la siguiente imagen se muestra una posible implementación del **diseño del modelo físico** para la tabla de hechos **FACT_EnvironmentalMeasurements**:

