

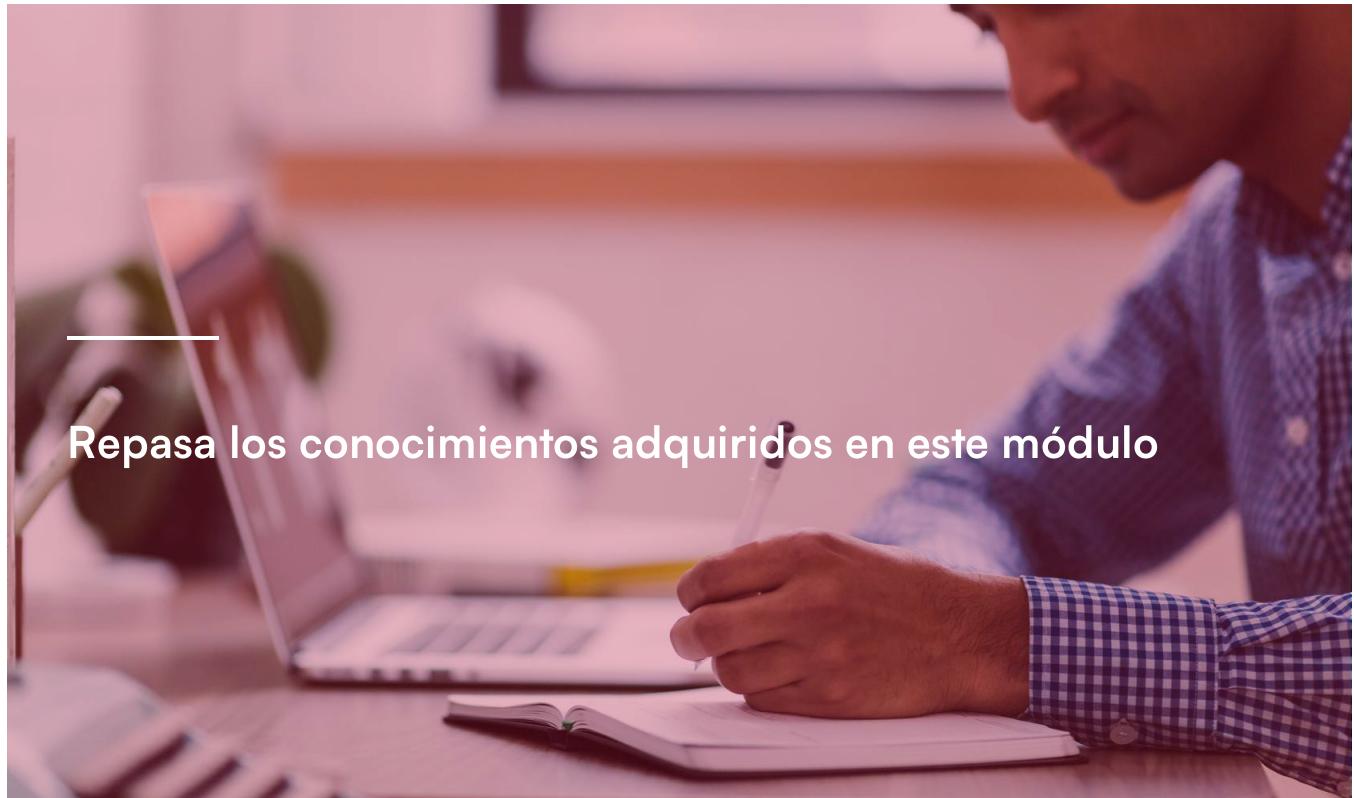
A photograph showing several people from behind, sitting at a long table in an office or study room. They are all wearing headphones and looking at their laptops. One laptop screen is clearly visible, displaying what appears to be a software interface or code. The atmosphere is focused and collaborative.

Repaso final de módulo



- ≡ I. Resumen de módulo
- ≡ II. Caso práctico de repaso con solución
- ≡ III. Lecturas recomendadas
- ≡ IV. Bibliografía

I. Resumen de módulo



Repasa los conocimientos adquiridos en este módulo

En este módulo se ha revisado el sistema de inteligencia de negocio como soporte imprescindible en la toma de decisiones de las empresas, así como los procesos y herramientas que forman parte de su arquitectura: almacén de datos y proceso ETL, en datos externos de la compañía y en todo lo relacionado a CRM y CI.

Se ha comenzado dedicando una unidad al concepto de inteligencia de negocio. Tal y como se ha descrito, la arquitectura tradicional de inteligencia de negocio no puede ofrecer un análisis en tiempo real, conocer la situación y proporcionar autoservicio. Es por esto que ha surgido la inteligencia de negocio de siguiente generación.

Después, se ha profundizado en el almacén de datos o *data warehouse* y bases de datos analíticas como motor principal de un sistema de inteligencia de negocio. Tanto en su motivación como objetivos y diseño, se ha comprobado que se prima la eficiencia del sistema para realizar análisis orientados a negocio.

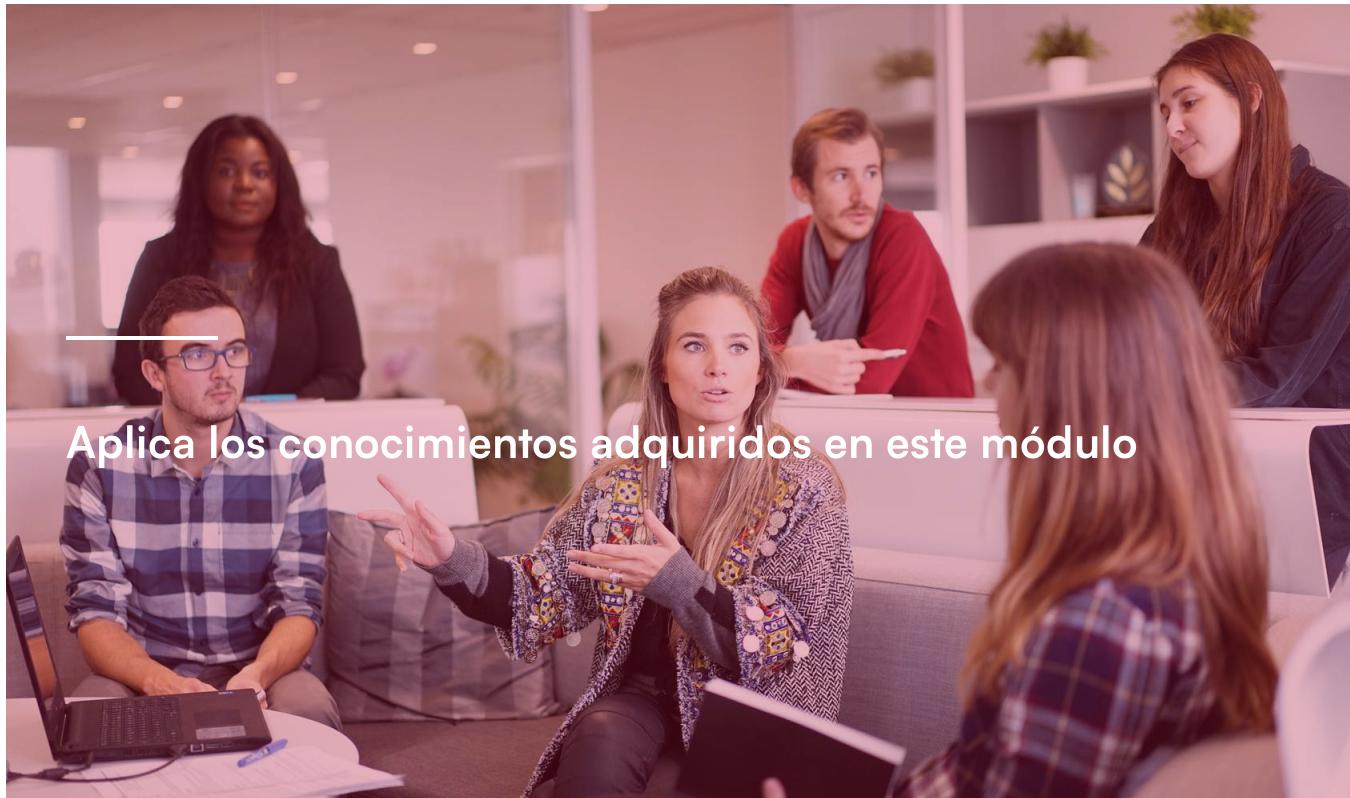
Asimismo, se ha tratado en detalle el proceso de extracción, transformación y carga de datos. Este proceso es pieza clave para extraer y unificar las fuentes, y aporta calidad a los datos en el proceso. Su valor reside en que proporciona fidelidad y credibilidad al sistema generando una visión única y veraz de la información.

También se han expuesto casos de uso (casos de éxito) reales de implementación de soluciones de inteligencia de negocio en empresas, atendiendo a sus necesidades y requisitos, para optimizar el manejo de datos y de información de estas. Por esto, puede considerarse una herramienta básica en la toma de decisiones y mejora del rendimiento y eficiencia de las empresas.

Posteriormente, se ha realizado un análisis de la información externa a la compañía para poder aportar información adicional a la toma de decisiones: open data, datos externos de pago y compartidos, así como datos de redes sociales. Técnicas de *scrapping*, API abiertas, etc., son habituales en estos entornos.

Para finalizar, en la última unidad teórica se ha puesto el foco en los CRM, sistemas de gestión de relaciones con clientes, y se ha estudiado cuáles son sus funcionalidades y principales soluciones comerciales. También se ha visto el concepto de experiencia de cliente y cómo mejorarlo mediante técnicas de inteligencia de clientes.

II. Caso práctico de repaso con solución



Aplica los conocimientos adquiridos en este módulo

ENUNCIADO

1.1. ¿Qué es el Grupo World Bank?

El Grupo World Bank tiene 189 países miembros, personal en más de 170 países y oficinas en más de 130 ubicaciones. El Grupo World Bank es una asociación global única: consta de cinco instituciones que trabajan en soluciones sostenibles que reducen la pobreza y generan prosperidad compartida en los

países en desarrollo. El Grupo World Bank es una de las mayores fuentes de financiación y conocimiento del mundo para los países en desarrollo. Sus cinco instituciones comparten el compromiso de reducir la pobreza, aumentar la prosperidad compartida y promover el desarrollo sostenible.

El grupo recoge y pone a disposición del público general grandes cantidades de datos. Se pueden estudiar [desde aquí](#).

Los Archivos protegen la memoria institucional del Grupo World Bank y proporcionan acceso público a los registros del World Bank de Reconstrucción y Fomento (BIRF) y la Asociación Internacional de Fomento (AIF). Este sitio web ofrece una variedad de recursos históricos e información.

1.2. ¿Qué datos vamos a usar?

En octubre de 2014, el Grupo World Bank lanzó la nueva base de datos SDDS de Estadísticas Trimestrales de Deuda Externa (QEDS). Esta base de datos es consistente con las clasificaciones y definiciones de las Estadísticas de la deuda externa de 2013 (Guía EDS de 2013) y Sexta edición de la Balanza de pagos y el Manual de posición de inversión internacional (BPM6). La base de datos QEDS SDDS proporciona datos detallados de la deuda externa a partir de 1998Q1. Los datos se publican individualmente por países que se suscriben al Estándar Especial de Divulgación de Datos (SDDS) del FMI, así como por los países participantes en el GDDS que están en condiciones de producir los datos de deuda externa prescritos por la SDDS.

El conjunto básico de tablas se basa en la categoría de datos de deuda externa prescrita en SDDS (Tabla 1) y en las categorías de datos de deuda alentadas (Tablas 2, 3 y 4). La Tabla 1 incluye el desglose de la posición de la deuda externa bruta total por sector, vencimiento e instrumento, y las otras tres tablas, Tabla 2, 3 y 4, recopilan datos sobre el desglose de la deuda externa en moneda nacional y extranjera, el programa de servicio de la deuda a plazo y pagos de capital e intereses adeudados en un año o menos. El conjunto restante de tablas comprende estadísticas suplementarias que van más allá de los requisitos de SDDS para proporcionar presentaciones analíticas adicionales, particularmente con respecto a los datos sectoriales, y facilitar el análisis de datos entre países.

La fuente consta de dos ficheros, un maestro de países y otro maestro de datos. El maestro de datos tiene un formato que debemos pivotar para ya que las fechas están en columnas en vez de en filas. Además, se debe proceder a limpiar todas las series cuyos valores sean nulos.

Para la realización del caso práctico se deberán utilizar los archivos "[SDDSDData.csv](#)" y

["SSDSCountryData.csv"](#).

SE PIDE

1. Crear una base de datos de *staging*.
2. Crear una base de datos de DWH.
3. Realizar una pequeño análisis de los datos.
4. Abrir Spoon de PDI y crear las primeras tareas de integración de datos en la base de datos STG.
5. Crear y cargar el DWH, la tabla “Fact” solo debe almacenar las métricas que tengan valores desde 1998Q1.
6. Crear un *job* o trabajo con todas las transformaciones en orden que se hayan generado previamente.
7. Crear el *data source* (cubo) en Pentaho Data Analytics.
8. Crear una visualización con Visualizer.



SDDSDData.csv

93.2 MB





SDDSCountry.csv

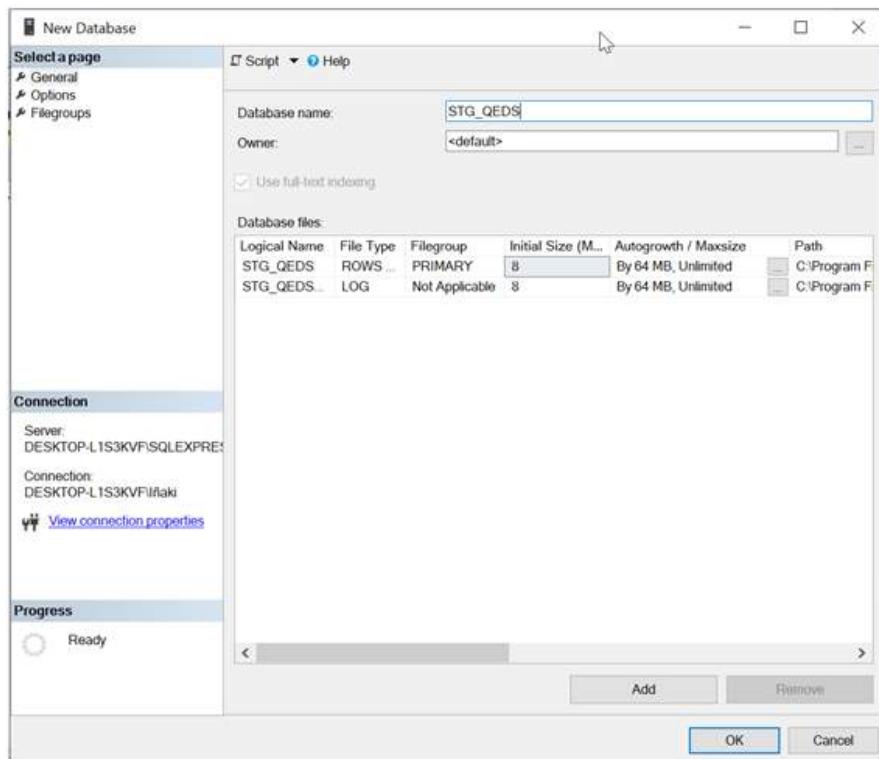
7.4 KB



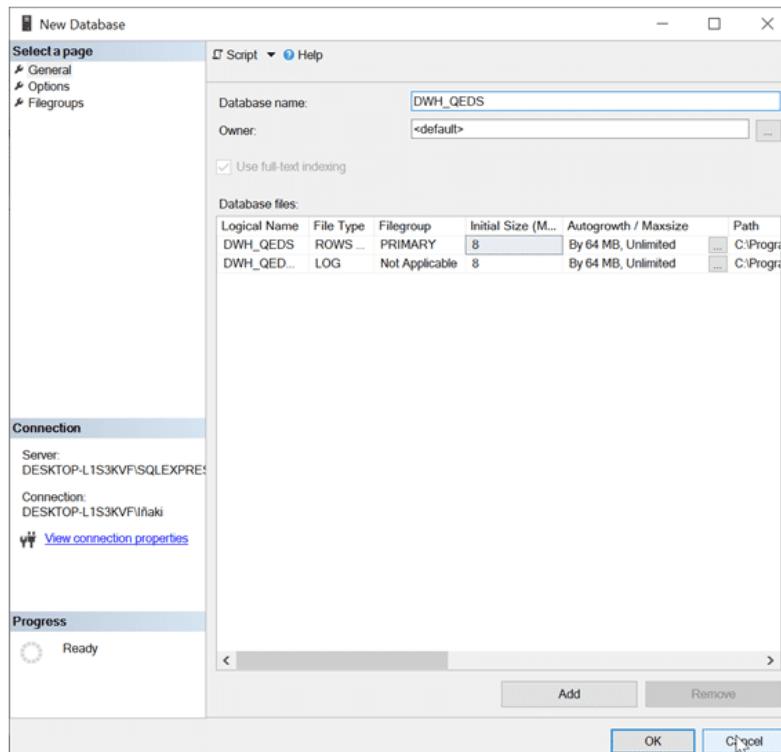
VER SOLUCIÓN

SOLUCIÓN

1. Crear una base de datos de staging.



2. Crear una base de datos de DWH.



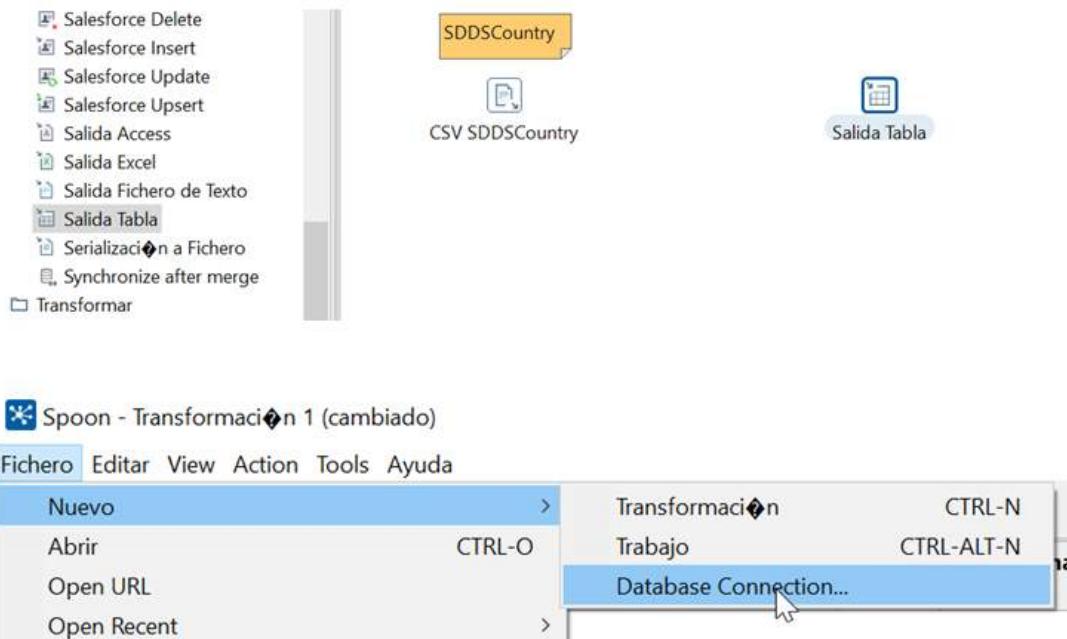
3. Realizar una pequeño análisis de los datos.

En esta fase se debe realizar una explicación de los datos que se tienen y su estructura.

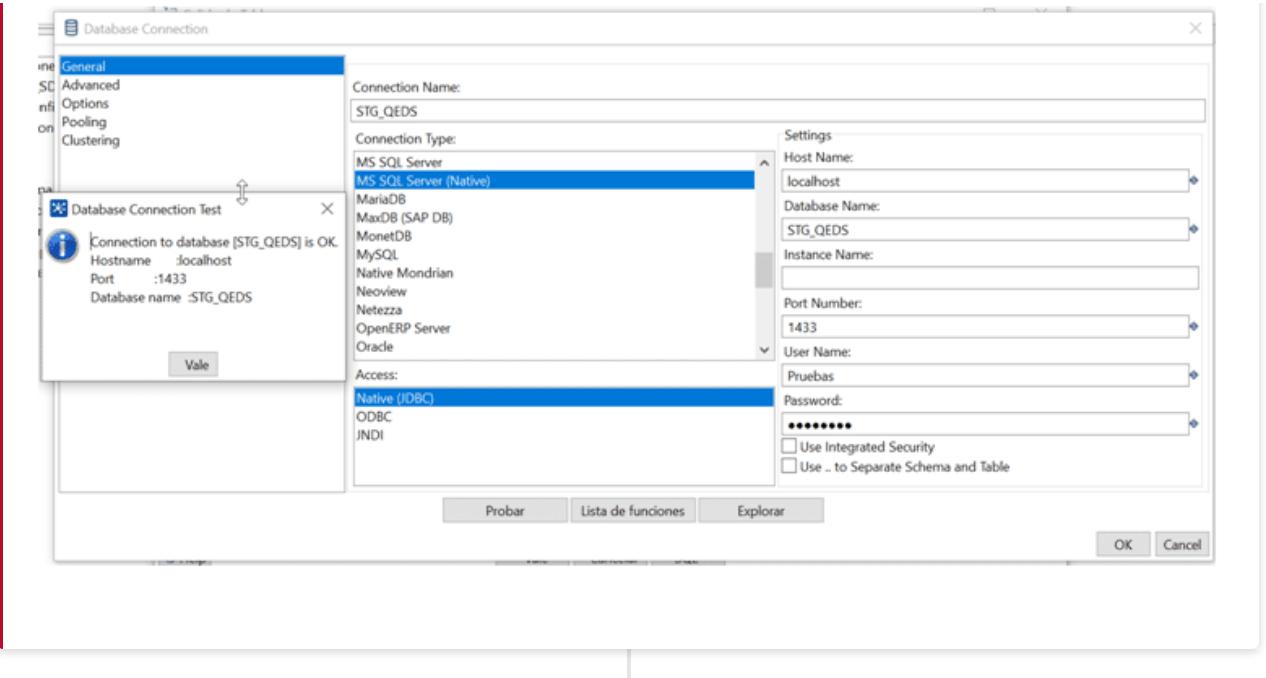
El fichero de datos contiene columnas para país, métricas y una por cada trimestre desde 1998 a 2019. Estos datos se trataran en el paso al DWH.

4. Abrir Spoon de PDI y crear las primeras tareas de integración de datos en la base de datos STG.

Ahora se configura la salida a base de datos, MySQL Server local



Desde aquí se crea la conexión a base de datos.



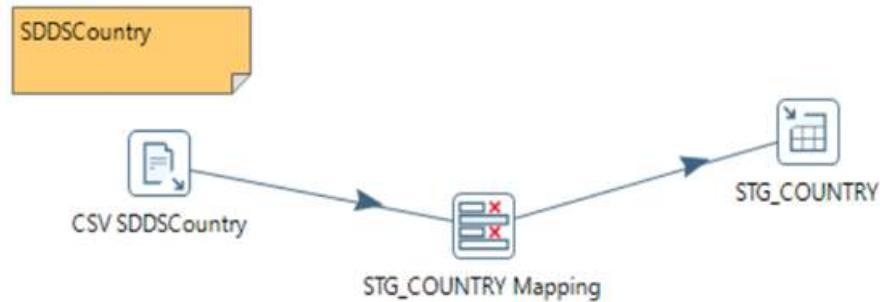
Una vez creada la conexión, seleccionar, con el botón derecho, compartir.

La carga del CSV de datos es como sigue, usando una transformación.



Execution Results													
	Nombre paso	Numero Copia	Lectura	Escrito	Entrada	Salida	Actualizado	Rejected	Errores	Activo	Tiempo	Velocidad (r/s)	Pri/I
1	CSV file input	0	0	219600	219601	0	0	0	0	Finalizado	44.8s	4.900	
2	Salida Tabla Mapping	0	219600	219600	0	0	0	0	0	Finalizado	46.9s	4.685	
3	Salida Tabla	0	219600	219600	0	219600	0	0	0	Finalizado	50.1s	4.386	

Repetir pasos para la tabla “STG_COUNTRY”.



5. Crear y cargar el DWH. La tabla “Fact” solo debe almacenar las métricas que tengan valores desde 1998Q1.

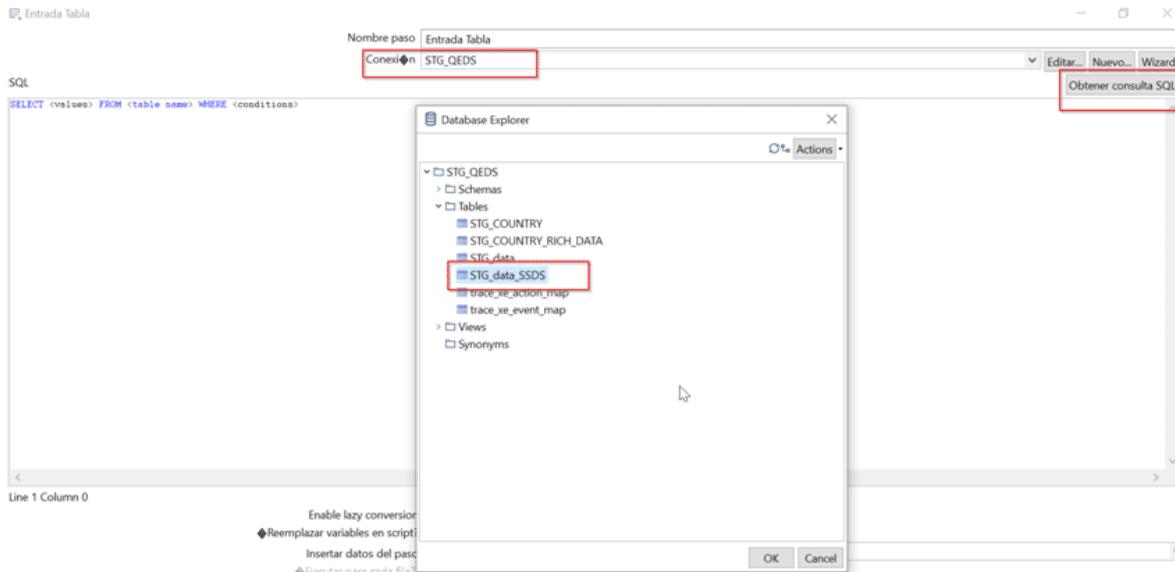
- Se crea la base de datos.
- Se crea la tabla “Fact_QEDS”, con los ID componiendo la PK.

```
CREATE TABLE [dbo].[FACT_QEDS](  
    [ID_PAIS] [varchar](3) NOT NULL ,  [ID_TRIM] [varchar](50) NOT NULL ,  [ID_METRICA] [varchar]  
(50) NOT NULL ,  [IN_VALUE] [float] NULL  PRIMARY KEY ([ID_PAIS],[ID_TRIM],[ID_METRICA])  
) ON [PRIMARY]
```

En PDI se crea una transformación con las dos conexiones de base de datos.

- > █ Run configurations
- ▼ █ Conexiones a base de datos
 - █ DWH_QEDS
 - █ STG_QEDS
- █ Pasos

Se introduce en la transformación una entrada de tabla.



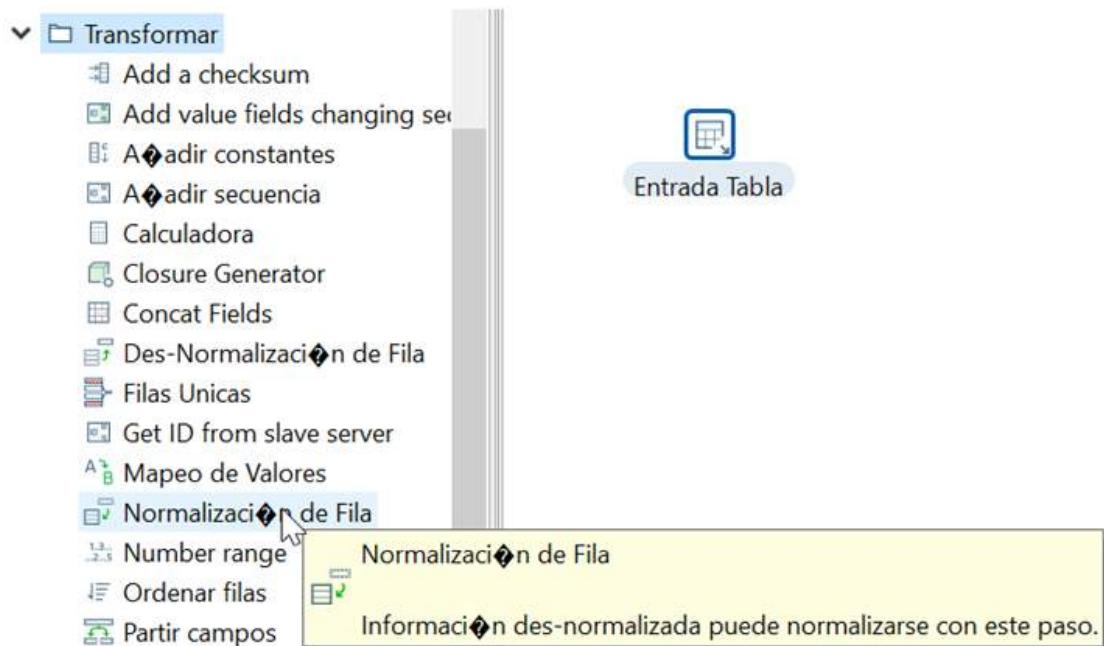
Como existen muchas series de datos que no están completas se van a seleccionar las series que tengan datos desde 1998Q1, añadiendo a la select este filtro:

```
FROM STG_data_SSDSWHERE "Indicator Code" IN (SELECT [Indicator Code]
```

```
FROM [STG_QEDS].[dbo].[STG_data_SSDS] group by [Indicator Code] having count([a1998Q1])>1)
```

Así se pasa de 1800 métricas a 159.

Desde transformaciones con un normalizar filas, se pivotan los datos.



Y se configura de la siguiente forma:

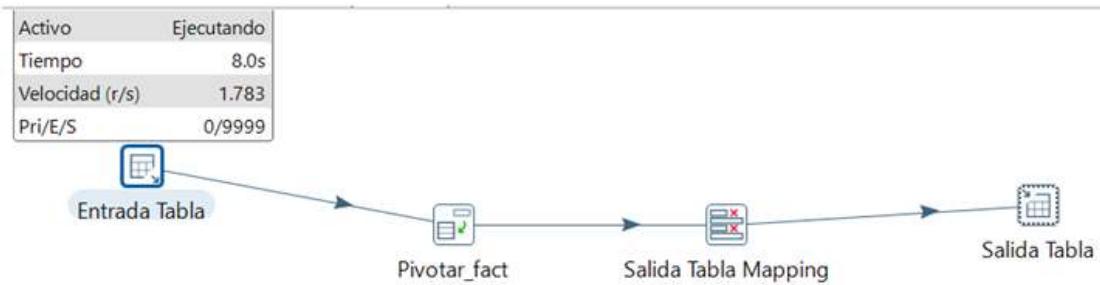
Normalizar Filas

Nombre de paso: Pivolar_fact
Tipo de campo: ID_TRIM

Campos

#	Nombre campo	Tipo	campo nuevo
1	a1998Q1	a1998Q1	IN_VALUE
2	a1998Q2	a1998Q2	IN_VALUE
3	a1998Q3	a1998Q3	IN_VALUE
4	a1998Q4	a1998Q4	IN_VALUE
5	a1999Q1	a1999Q1	IN_VALUE
6	a1999Q2	a1999Q2	IN_VALUE
7	a1999Q3	a1999Q3	IN_VALUE
8	a1999Q4	a1999Q4	IN_VALUE
9	a2000Q1	a2000Q1	IN_VALUE
1.	a2000Q2	a2000Q2	IN_VALUE
1.	a2000Q3	a2000Q3	IN_VALUE
1.	a2000Q4	a2000Q4	IN_VALUE
1.	a2001Q1	a2001Q1	IN_VALUE
1.	a2001Q2	a2001Q2	IN_VALUE
1.	a2001Q3	a2001Q3	IN_VALUE
1.	a2001Q4	a2001Q4	IN_VALUE
1.	a2002Q1	a2002Q1	IN_VALUE
1.	a2002Q2	a2002Q2	IN_VALUE
1.	a2002Q3	a2002Q3	IN_VALUE
2.	a2002Q4	a2002Q4	IN_VALUE
2.	a2003Q1	a2003Q1	IN_VALUE
2.	a2003Q2	a2003Q2	IN_VALUE
2.	a2003Q3	a2003Q3	IN_VALUE
2.	a2003Q4	a2003Q4	IN_VALUE
2.	a2004Q1	a2004Q1	IN_VALUE
2.	a2004Q2	a2004Q2	IN_VALUE

La transformación queda de esta forma.



6. Crear un job o trabajo con todas las transformaciones en orden que se hayan generado previamente.

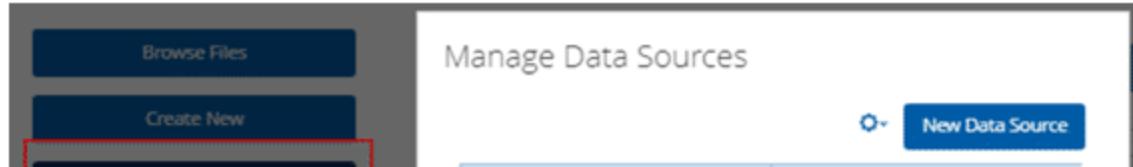
Este trabajo ejecuta las transformaciones del Data Mart



7. Crear el data source (cubo) en Pentaho Data Analytics.

Arrancar el servicio con el script que esta en el escritorio “start-pentaho - Acceso directo.bat”.

Entrar como *admin* evaluador y generar un nuevo cubo y un *data source*. Para ello crear un nuevo *data source* basado en la tabla “FACT_QEDS” que se acaba de crear.



The screenshot shows the 'Manage Data Sources' interface. On the left, there's a sidebar with 'Manage Data Sources' and 'Documentation' buttons, and a 'Recents' section listing 'Webinar' and 'Buyer Report (spa...)' with star icons. The main area displays a table of data sources:

Datasource	Type
AgileBI	JDBC
caso_practico_final	Data Source Wizard
Caso_practico_u4	Data Source Wizard
Casos	JDBC
Northwind	JDBC

A red box highlights the 'Manage Data Sources' button in the sidebar and the 'Casos' row in the table.

Es importante seleccionar opción *reporting* y análisis.

The screenshot shows the 'Data Source Wizard' window. It has three tabs: 'Select Source Type' (selected), 'Select Tables', and 'Define Joins'. The 'Select Source Type' tab contains fields for 'Data Source Name:' (set to 'SDDS') and 'Source Type:' (set to 'Database Table(s)'). Below these fields, a note says: 'Select a database connection and click Next to choose from a list of the available database tables.' A 'Connection:' dropdown menu lists 'AgileBI', 'Casos' (which is highlighted with a red box), 'Northwind', and 'SampleData'. At the bottom, under 'Create data source for:', there are two radio button options: ' Reporting only' and ' Reporting and Analysis (Requires Star Schema)'. A red box highlights the 'Reporting and Analysis (Requires Star Schema)' radio button.

Es importante seleccionar tabla de hechos y joins.

The screenshot shows two steps of the Data Source Wizard:

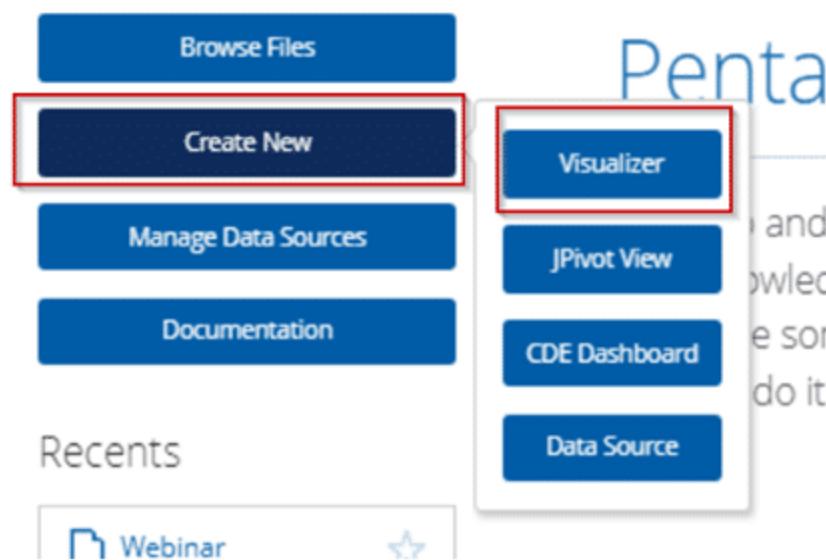
- Step 1: Select Source Type**
 - Schema: `dbo`
 - Available Tables: `dbo..ActivitiesbySalesPerson`, `dbo..FireBuckets`, `dbo..STG_COUNTRIES_DADA`, `dbo..STG_data_SSIS`, `dbo..ventas`
 - Selected Tables: `dbo..FACT_QEDS`, `dbo..STG_COUNTRY`
 - Fact Table: `dbo..FACT_QEDS`
- Step 2: Define Joins**
 - Left Table: `dbo..FACT_QEDS`
 - Right Table: `dbo..STG_COUNTRY`
 - Key Fields: `ID_PAIS`, `ID_TRIM`, `ID_METRICA`, `IN_VALUE`
 - Join Type: `INNER JOIN`
 - Join Condition: `dbo..FACT_QEDS.ID_PAIS = INNER JOIN - dbo..STG_COUNTRY.Country Code`

Editar el cubo por defecto.

The screenshot shows the Data Source Model Editor with the Analysis tab selected:

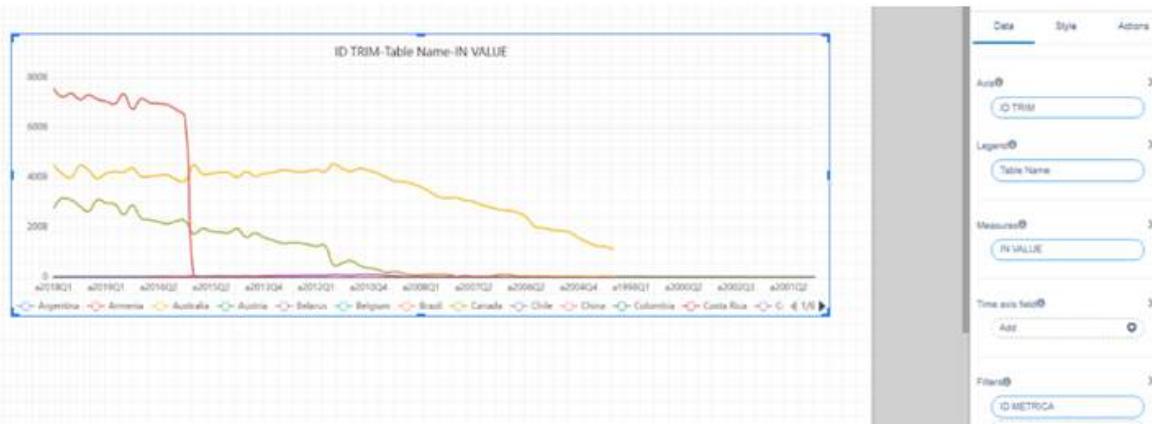
- Available** pane:
 - FACT QEDS** dimension:
 - ID METRICA** measure
 - ID PAIS**
 - ID TRIM**
 - IN VALUE**
 - STG COUNTRY** dimension:
 - Country Code**
 - Table Name**
- Analysis** pane:
 - Measures** group: **IN VALUE** (selected)
 - Dimensions** group:
 - STG COUNTRY** dimension:
 - Country Code**
 - Table Name**
 - ID TRIM**
 - ID METRICA**
 - ID METRICA**
- Properties** pane:
 - Source Column: **IN VALUE**
 - Display Name: **IN VALUE**
 - Default Aggregation: **SUM**
 - Format: **0.00**

8. Crear una visualización con Visualizer.



Para la visualización se usarán los siguientes parámetros:

- Eje x: Id_trim
- Leyenda:TableName(Country)
- Métrica:InValue
- Filtro:Una id_metrica



III. Lecturas recomendadas

En este primer libro se explican los conceptos básicos de la analítica en la nube con Azure, así como el uso de su data warehouse: Azure Synapse Analytics.

- Altaiar, H. *Cloud analytics with Microsoft Azure. Build modern data warehouses with the combined power of analytics and Azure*. Packt Publishing; 2019. Disponible en la biblioteca virtual (lectura en inglés).

Este segundo libro es una completa guía de tratamiento y utilización de datos mediante la exploración, transformación, validación e integración con Pentaho Data Integration 8.

- Roldan, M. C. *Learning Pentaho Data Integration 8 CE (3rd ed.) An end-to-end guide to exploring, transforming, and integrating your data across multiple sources*. Packt Publishing; 2017. Disponible en la biblioteca virtual (lectura en inglés).

IV. Bibliografía

- Alavi, M. *Knowledge Management Systems: Issues, Challenges, and Benefits*. Communications of the AIS. Armstrong-Smith, D. a. (s. f.). Oracle Discoverer 10g Handbook. San Francisco; 1999.
- Anderson, G; Heller, S. *Libro de ideas para el diseño gráfico*. Art Blume; 2016, 5.^a ed.
- Arnott, D.; Pervan, G. "A critical analysis of Decision Support Systems research". *Journal of Information Technology*; 2005; pp. 67-87
- Braunschweig, K.; Eberuis, J.; Thiele, M.; Lehner, W. *The state of open data. Limits of current open data platforms*. Technische Universität Dresden; 2012.
- *Business Intelligence: evolución hacia la empresa inteligente*. Ibermática; 2015.
- *Business Intelligence, el conocimiento compartido*. Ibermática; 2007. [En línea] URL [disponible aquí](#).
- Cano, J. L. *Business intelligence: competir con información*. Banesto, Fundación Cultural; 2007.
- Castellanos, M.; Gupta, C.; Wang, S.; Dayal, U.; Durazo, M. "A platform for situational awareness in operational BI". *Decision Support Systems*; 2012; vol. 52, n. 4: pp. 869-883.
- Chaudhuri, S.; Dayal, U.; Narasayya, V. "An overview of business intelligence technology". *Communications of the ACM*; 2011; vol. 54, n. 8: pp. 88-98.
- D. M. International, *The DAMA guide to the data management body of knowledge*. Bradley Beach: Technics Publications; 2009.

- Darshan. Institute of Engineering and Technology. [En línea] URL [disponible aquí](#).
- Documentación de Salesforce. Hitachi. [En línea] URL [disponible aquí](#).
- Elmasri, R.; Navathe, S. B. *Fundamentals of Database Systems*. Pearson; 2014.
- Gray, P; Watson, H. J. *Decision Support in the Data Warehouse*. Prentice Hall PTR; 1998.
- Hitachi. Pentaho 9.2. [En línea] URL [disponible aquí](#).
- Ibermática. *Business Intelligence, el conocimiento compartido*. 2007.
- Inmon, W. H. *Building the Data Warehouse* (4.^a ed.). Wiley; 2005.
- Inmon, W. H.; Strauss, D.; Neushloss, G. *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. Amsterdam: Morgan Kafmann; 2008.
- Jiawei, H.; Kamber, M. *Data Mining: Concepts and Techniques* (2.a ed.). Morgan Kaufmann Publishers; 2001.
- Jones, B. *Comunicating Data with Tableau* (1.^a ed.). O'Reilly Media; 2014.
- Kaplan, R. S.; Norton, D. *Cuadro de mando integral o balance scorecard (BSC)*. Grupo Planeta; 2009. 3.^a ed.
- Khan, R. A.; Quadri, S. K. "Business intelligence: An integrated approach". *Business Intelligence Journal*; 2012; vol. 5, n. 1: pp. 64-70.
- Khatri, V.; Brown, C. V. "Designing Data Governance". *Communications of the ACM*; 2010; vol. 53, n. 1: pp. 148-152.
- Kimball, R. *The Data Warehouse Toolkit*. Ed. John Wiley; 1996.
- Kimball, R.; Caserta, J. *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data*. John Wiley & Sons; 2011.

- Krasikov, P.; Legner, C.; Harbich, M.; Eurich, M. *Open data use cases. Framework for the generation and documentation of open data use cases*; 2020.
- Krug, S. *No me hagas pensar* (2.^a ed.). Pearson Prentice Hall; 2006.
- Löser, A; Hueske, F.; Markl, V. "Situational business intelligence". *Business Intelligence for the Real-Time Enterprise*. Springer; 2009; pp. 1-11.
- Luhn, H. P. "A Business Intelligence System". *IBM Journal*; 1958; pp. 314-319.
- "Maximizing information through multiple kernel-based heterogeneous data integration and applications to ovarian cancer". *6th International Conference on Emerging Databases (EDB)*. ACM Press; 2016; pp. 97-100.
- Pendse, N. *The OLAP Report. What is OLAP?* Business Application Research Center; 2007. [En línea] URL [disponible aquí](#).
- Pentaho. *Official Documentation Pentaho Business Analytics & Data Integration Tools*. [En línea] URL [disponible aquí](#).
- Ponniah, P. *Data Warehousing Fundamentals. A Comprehensive Guide for IT Professionals*. John Wiley & Sons; 2004.
- Reddy, M. S.; Khan, V. *Comparative Analysis of On-Line Analytical Processing Tools*. Göteborg: IT University of Göteborg; 2007.
- Rojas Bartomeu, P. *Estudio comparativo de bases de datos analíticas*. TFM Máster en Tecnologías de la Información, Especialidad en Ingeniería del Software y Sistemas de Información. Everis, Facultat d'Informàtica de Barcelona, UPC; 2009.
- Ross, M. "Design Tip #152 Slowly Changing Dimension Types 0, 4, 5, 6 and 7". Kimball Group. 5 de febrero de 2013. [En línea] URL [disponible aquí](#).
- Sabherwal, R.; Becerra-Fernández, I. *Business Intelligence: Practices, Technologies and Management*. Estados Unidos: John Wiley & Sons; 2011.

- Seibold, M.; Kemper, A.; Jacobs, D. "Strict SLAs for Operational Business Intelligence". *Cloud Computing (CLOUD), IEEE International Conference*; 2011; pp. 25-32.
- Shanklin, C. "Managing Slowly Changing Dimensions". Cloudera; 23 de agosto de 2017. [En línea] URL [disponible aquí](#).
- Sundeep Reddy, M.; Khan, V. *Comparative Analysis of On-Line Analytical Processing Tools*. Göteborg: IT University of Göteborg; 2007.
- Thomas, J.; Sael, L. "Maximizing information through multiple kernel-based heterogeneous data integration and applications to ovarian cancer". *6th International Conference on Emerging Databases (EDB)*. ACM Press; 2016; pp. 97-100.
- Thomsen, E. *OLAP Solutions, Building Multidimensional Information Systems*. John Wiley & Sons; 2002.
- Veloso, F. "Web scraping para data scientist". *Feeding the Machine*; 30 de junio de 2019. [En línea] URL [disponible aquí](#).
- Vitt, E. *Business Intelligence. Técnicas de análisis para la toma de decisiones estratégicas*. McGraw-Hill; 2003.