



**Facultad de Ingeniería
Departamento de Ingeniería en Sistemas**

Clase y Sección:

Bases De Datos II

IS-601 0900

PROYECTO FINAL

Catedrático:

Ing. Emilson Acosta

Alumna:

Ana Evelin Hernández Martínez

Número de Cuenta:

20171001620

Fecha de Entrega:

Jueves 15 de abril de 2021

2. ÍNDICE

2. ÍNDICE	2
3. INTRODUCCTIÒN	5
4. MARCO TEÓRICO	6
4.1 Conceptos Técnicos	6
4.1.1 Data Warehouse	6
4.1.2 Base de Datos OLTP	7
4.1.3 Business Intelligence	8
4.1.4 ETL	9
4.1.5 Base de Datos OLAP	10
4.1.6 Cubo OLAP	10
4.1.7 Analizador Pentaho	11
4.1.8 Pentaho Data Integration	11
5. ESQUEMAS DE BASES DE DATOS	12
5.1 Esquema Base de Datos OLTP	12
5.2 Esquema Base de Datos OLAP	13
6. CAPTURAS DE PANTALLA DEL PROCESO ETL	14
6.1 Ítems Utilizados	14
6.2 Consultas utilizadas para el proceso ETL	15
6.2.1 Consulta para llenar la Tabla de Dimensión de Clientes:	15
6.2.2 Consulta para llenar la Tabla de Dimensión de Empleados:	15
6.2.3 Consulta para llenar la Tabla de Dimensión de Propiedades:	16
6.2.4 Consulta para llenar la Tabla de Dimensión de Proveedores:	16
6.2.5 Consulta para llenar la Tabla de Dimensión de Tiempo:	17
6.2.6 Consulta para llenar la Tabla de Hechos de Ventas:	17
6.3 Configuración de Ítems para el ETL	18
6.3.1 ETL Tabla de Dimensión de Propiedades	18
6.3.1.1 Entrada de datos	18

6.3.1.2 Transformación de Datos	18
6.3.1.3 Salida de Datos	19
6.3.1.4 Conexión exitosa de los 3 ítems	19
6.3.2 ETL Tabla de Dimensión de Proveedores	20
6.3.2.1 Entrada de datos	20
6.3.2.2 Transformación de Datos	20
6.3.2.3 Salida de Datos	21
6.3.2.4 Conexión exitosa de los 3 ítems	21
6.3.3 ETL Tabla de Dimensión de Clientes	22
6.3.3.1 Entrada de datos	22
6.3.3.2 Transformación de Datos	22
6.3.3.3 Salida de Datos	23
6.3.3.4 Conexión exitosa de los 3 ítems	23
6.3.4. ETL Tabla de Dimensión de Tiempo	24
6.3.4.1 Entrada de datos	24
6.3.4.2 Transformación de Datos	24
6.3.4.3 Salida de Datos	25
6.3.4.4 Conexión exitosa de los 3 ítems	25
6.3.5 ETL Tabla de Dimensión de Empleados	26
6.3.5.1 Entrada de datos	26
6.3.5.2 Transformación de Datos	26
6.3.5.3 Salida de Datos	27
6.3.5.4 Conexión exitosa de los 3 ítems	27
6.3.6 ETL Tabla de Hechos de Ventas	28
6.3.6.1 Entrada de datos	28
6.3.6.2 Transformación de Datos	28
6.3.6.3 Salida de Datos	29
6.3.6.4 Conexión exitosa de los 3 ítems	29
6.4 Ejecución exitosa de todos los ETL	30
6.4.1 Ejecución de ETL de Llenar Dimensiones	30
6.4.2 Ejecución de ETL de Llenar Hechos	30

6.4.3 Ejecución de ETL de Limpiar Hechos	31
6.4.4 Ejecución de ETL de Limpiar Dimensiones	31
6.5 Ejecución exitosa del JOB para ejecutar los ETL	32
6.6 Ejecución de consultas para ver registros	32
6.6.1 Tabla D_CLIENTES	32
6.6.2 Tabla D_EMPLEADOS	33
6.6.3 Tabla D_PROPIEDADES	33
6.6.4 Tabla D_PROVEEDORES	34
6.6.5 Tabla D_TIEMPO	34
6.6.6 Tabla H_VENTAS	35
7. Creación del Cubo OLAP	35
7.1 Conexión con la base de datos OLAP	35
7.2 Data Source Model / Cubo OLAP	36
7.2.1 Creación	36
7.2.2 Selección	38
7.2.3 Edición	38
7.3 Generación de Reportes	39
7.3.1 Reporte – Total de ventas en base a tipo de producto	39
7.3.2 Reporte – Nombre y Código de empleado que realiza más ventas	39
7.3.3 Reporte – Total de ventas en base al empleado que las realiza	40
7.3.4 Reporte – Total de cada propiedad vendida en base al cliente	40
7.3.5 Reporte – Analizar las ventas por mes (DrillDown)	41
7.3.6 Reporte – Analizar las ventas por año(DrillUp)	41
7.4 Asignación de Reportes	42
7.4.1 Usuario Admin	42
7.4.2 Usuario Suzy	42
8. Preguntas de Negocio Utilizadas	43
9. Explicación de Métricas Utilizadas	43
10. Conclusiones	44
11. Recomendaciones	45
12. Bibliografía	46

3. INTRODUCCIÓN

En el siguiente informe se contempla un resumen e investigación de lo que se realizó para poder completar el proyecto de DATAWAREHOUSE para una empresa inmobiliaria, partiendo de la creación de una base de datos OLTP, un modelo en estrella para la generación de un datamart y una base de datos OLAP.

La información de la base de datos OLAP será alimentada mediante procesos ETL desarrollados en Pentaho, dicha información servirá para llegar al objetivo final que es poder generar reportes de una manera más comprensible para todo tipo de usuario y que estos nos permitan crear una toma de decisiones más precisa acerca de las necesidades y éxitos que esté teniendo la empresa.

Este proyecto está orientado a la generación del Business Intelligence más apropiado y enfocado a la administración y creación de información mediante análisis de datos existentes.

4. MARCO TEÓRICO

4.1 Conceptos Técnicos

4.1.1 Data Warehouse

Data warehouse es un sistema que agrega y combina información de diferentes fuentes en un almacén de datos único y centralizado; consistente para respaldar el análisis empresarial, la minería de datos, inteligencia artificial (IA) y Machine Learning. Data warehouse permite a una organización o empresa ejecutar análisis potentes en grandes volúmenes (petabytes y petabytes) de datos históricos de formas que una base de datos estándar simplemente no puede.

Los data warehouses han sido parte de las soluciones de inteligencia empresarial (BI) durante más de tres décadas, pero han evolucionado significativamente en los últimos años. Tradicionalmente, un data warehouse tenía una implementación on-premises, a menudo en un mainframe central, y su funcionalidad se centraba en extraer datos de otras fuentes, limpiar y preparar la información, y cargar y mantener los documentos en una base de datos relacional (Relational Database). Hoy en día, un Data warehouse puede estar alojado en un dispositivo dedicado o en la nube, y la mayoría de los data warehouses han agregado capacidades de análisis y herramientas de visualización y presentación de datos. Todo lo anterior hace que los reportes provenientes de este data warehouse se vean enriquecidos debido a la diversidad de las fuentes. Un ejemplo, en lugar de solo tener información proveniente de marketing, tendremos información proveniente de producción, marketing, ventas, business Partners y más, haciendo posible la detección de patrones, tendencias e insights que nos ayuden a tomar mejores decisiones de negocio.

4.1.2 Base de Datos OLTP

Un sistema OLTP se puede definir como una herramienta tecnológica capaz de soportar el procesamiento, administración y mantenimiento diario de transacciones generadas por los negocios de una compañía a nivel corporativo, para ofrecer altos niveles de disponibilidad, seguridad y confiabilidad. El proceso transaccional es típico de las bases de datos operacionales. ¿Para qué sirven? ¿Cómo ayudan a tomar decisiones?

Los sistemas OLTP, siglas de inglés On Line Transaction Processing, son bases de datos orientadas al procesamiento de transacciones. Como ejemplos, se pueden citar las operaciones en un hipermercado, las reservas aéreas, las operaciones bancarias, la fabricación masiva, las ventas de las grandes tiendas. Además de ser los sistemas soporte de las operaciones diarias de una compañía, los sistemas OLTP hoy día asumen un papel muy importante:

- Proporcionan información confiable y precisa a los grandes almacenes de datos, que son la fuente principal para generar información de gestión.
- Proveen acceso optimizado a los datos para tareas frecuentes de lectura y escritura. Por ejemplo, la enorme cantidad de transacciones que tienen que soportar las bases de datos de bancos o hipermercados diariamente.
- Estructurar los datos según el nivel de aplicación. Por ejemplo, sistema de gestión a medida, ERP o CRM implantado, sistema de información departamental, etcétera.

Algunas desventajas potenciales de los sistemas OLTP son las siguientes:

- Es posible que los datos no sean coherentes en la empresa.
- El acceso a los datos puede resultar complicado.
- Los formatos de los datos no son necesariamente uniformes en los diferentes departamentos. Es común la falta de compatibilidad y la existencia de islas de datos.

- El historial de datos suele limitarse a los datos actuales o recientes.

4.1.3 Business Intelligence

Business Intelligence es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios. Desde un punto de vista más pragmático, y asociándolo directamente con las tecnologías de la información, podemos definir Business Intelligence como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporting, análisis OLTP / OLAP, alertas...) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio.

La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, etc...

Los principales productos de Business Intelligence que existen hoy en día son:

- Cuadros de Mando Integrales (CMI)
- Sistemas de Soporte a la Decisión (DSS)
- Sistemas de Información Ejecutiva (EIS)

Por otro lado, los principales componentes de orígenes de datos en el Business Intelligence que existen en la actualidad son:

- Datamart

- Datawarehouse

Los sistemas y componentes del BI se diferencian de los sistemas operacionales en que están optimizados para preguntar y divulgar sobre datos. Esto significa típicamente que, en un datawarehouse, los datos están desnormalizados para apoyar consultas de alto rendimiento, mientras que en los sistemas operacionales suelen encontrarse normalizados para apoyar operaciones continuas de inserción, modificación y borrado de datos. En este sentido, los procesos ETL (extracción, transformación y carga), que nutren los sistemas BI, tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén completamente integrados.

4.1.4 ETL

Extraer, transformar y cargar (ETL, Extract, Transform, Load) es el proceso de compilación de datos a partir de un número ilimitado de fuentes, su posterior organización y centralización en un único repositorio.

En la mayoría de empresas los datos potencialmente útiles resultan inaccesibles; un estudio ha revelado que dos terceras partes de las sociedades o bien obtiene “pocos resultados tangibles” de sus datos o no obtiene ninguno. Los datos suelen estar aislados en compartimentos estancos, sistemas heredados o aplicaciones muy poco utilizadas. El ETL es el proceso por el que se ponen a disposición esos datos extrayéndolos de múltiples fuentes (como ilustra el diagrama de arriba) y transformándolos en datos útiles para la limpieza, la transformación y, por último, la obtención de información corporativa.

Hay quien efectúa los procesos de ETL programando manualmente en SQL o Java, pero existen herramientas que simplifican el proceso. En este artículo se analizan varios casos de aplicación de ETL, las ventajas que reporta el uso de una herramienta de ETL en lugar de programar a mano y las ventajas que deberían esperar encontrar los clientes en las herramientas de ETL.

4.1.5 Base de Datos OLAP

OLAP - On-Line Analytical Processing; Los sistemas OLAP son bases de datos orientadas al procesamiento analítico. Este análisis suele implicar, generalmente, la lectura de grandes cantidades de datos para llegar a extraer algún tipo de información útil: tendencias de ventas, patrones de comportamiento de los consumidores, elaboración de informes complejos... etc. Este sistema es típico de los datamarts.

1. El acceso a los datos suele ser de sólo lectura. La acción más común es la consulta, con muy pocas inserciones, actualizaciones o eliminaciones.
2. Los datos se estructuran según las áreas de negocio, y los formatos de los datos están integrados de manera uniforme en toda la organización.

4.1.6 Cubo OLAP

En el análisis multidimensional, los datos se representan mediante dimensiones como destino, autobús, operador y tiempo. En general, las dimensiones se relacionan en jerarquías, por ejemplo, ciudad, estado, región, país y continente. El tiempo también es una dimensión estándar con sus propias jerarquías tales como: día, semana, mes, trimestre y año. Para tener una idea más simple de la función de los cubos OLAP dentro de una base de datos multidimensional, cabe destacar que cada una de las dimensiones o escalas del cubo corresponde básicamente a una jerarquía de datos.

Pueden utilizarse otras dimensiones del cubo para recabar información referente a situaciones geográficas, clasificación de los productos por categorías, gastos realizados por la empresa, y demás. Esta confluencia de la información permite llevar a cabo un análisis completo de diversas situaciones, para obtener las soluciones correctas a los problemas de negocios.

Mediante la incorporación de estos vectores o cubos, se han ampliado las posibilidades de las bases de datos relacionales, permitiendo el procesamiento de importantes volúmenes de información, de lo contrario sería imposible realizar dicha actividad. Cada una de las dimensiones que posee la base de datos, incorpora un campo determinado para cierto tipo específico, que luego podrá ser comparado con la información contenida en el resto de dimensiones, para hacer posible la evaluación y posteriores informes de las actividades relevantes para una compañía.

4.1.7 Analizador Pentaho

Analyzer tiene un entorno de diseño de arrastrar y soltar, fácil de usar, basado en la web, que puede ser utilizado por cualquiera que desee explorar datos dinámicamente y profundizar para descubrir detalles previamente ocultos.

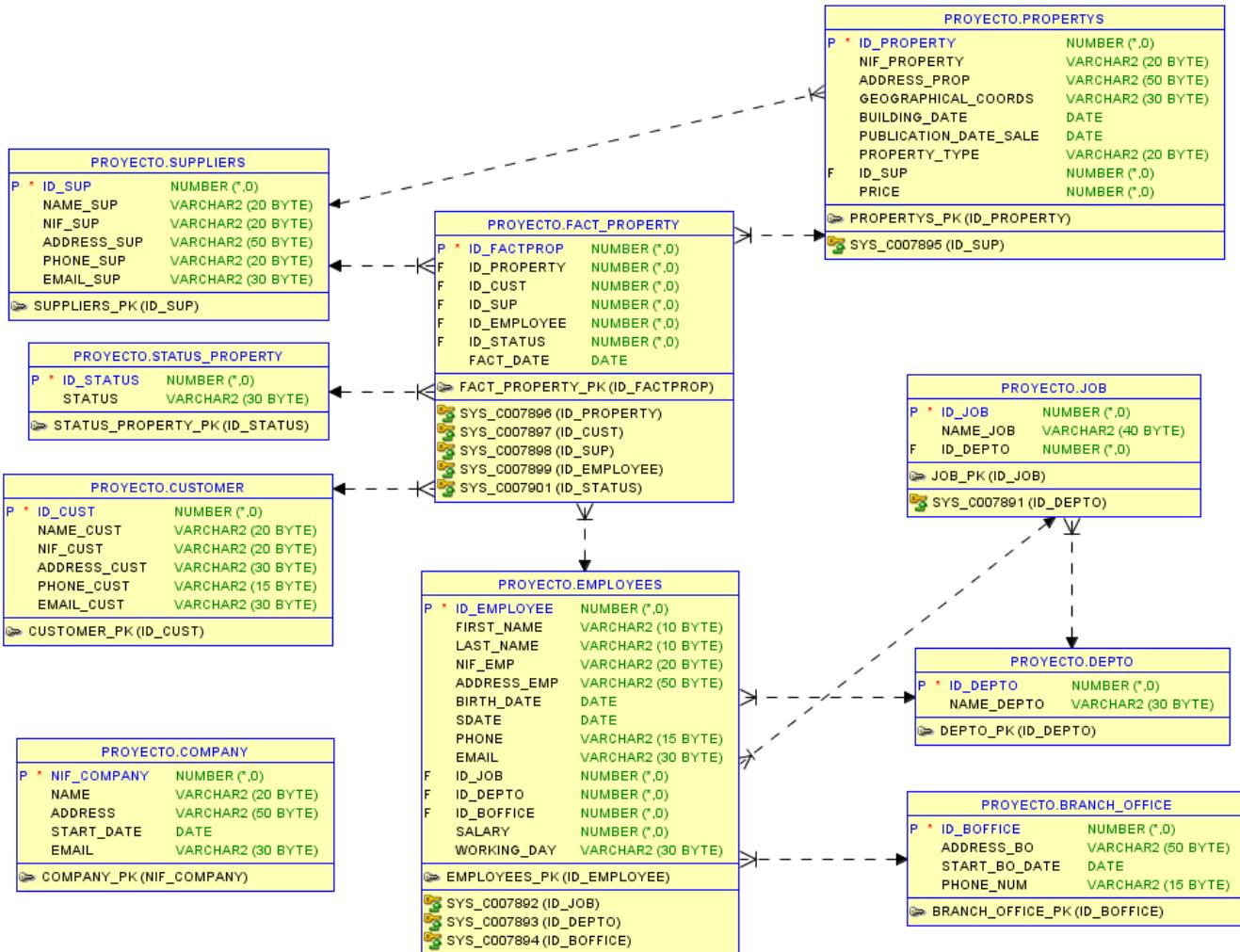
Antes de poder utilizar Analyzer, debe tener acceso a una fuente de datos. Solo los administradores del sistema pueden crear fuentes de datos. La fuente de datos de Analyzer se basa en el modelo de datos multidimensionales de Mondrian. El modelo de datos de Mondrian le permite elegir qué dimensiones y medidas desea explorar en sus datos.

4.1.8 Pentaho Data Integration

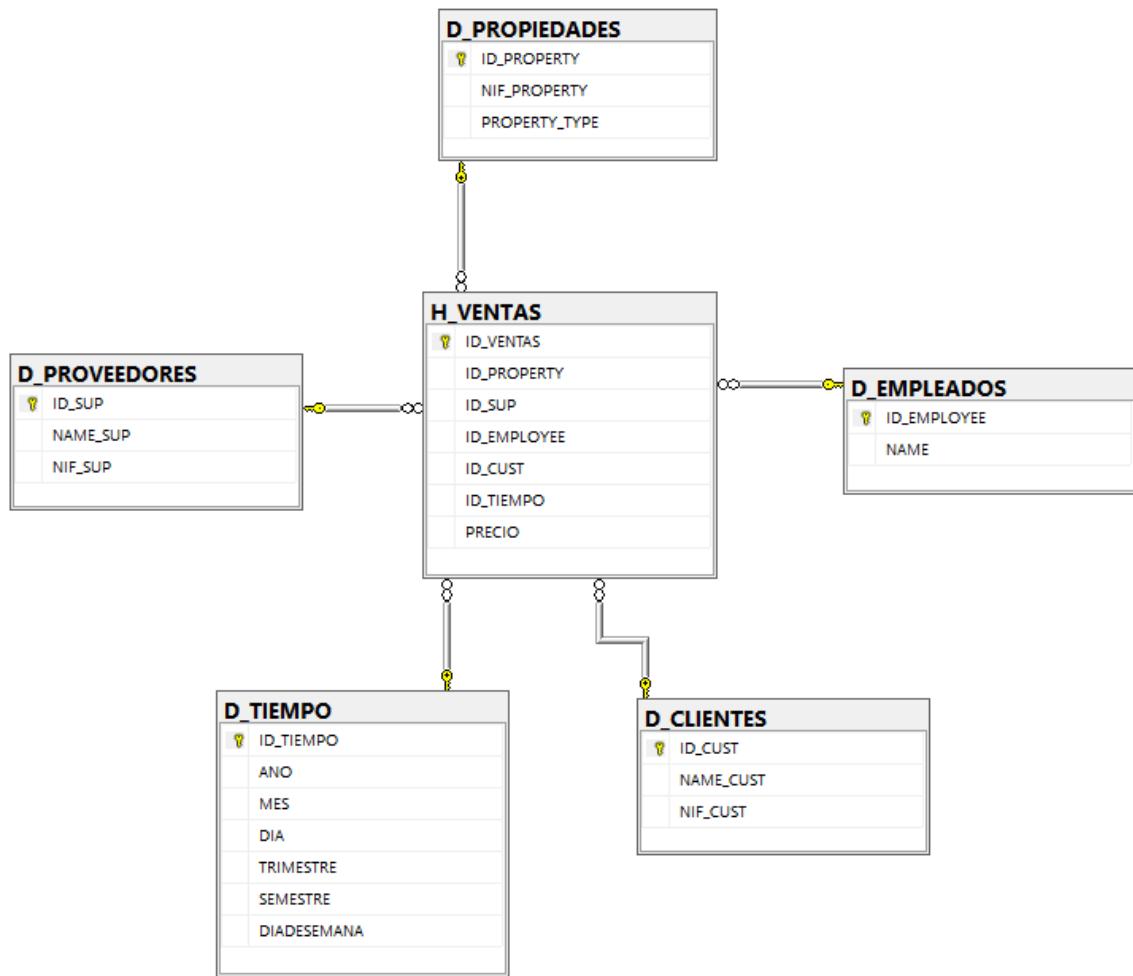
Pentaho Data Integration, cuyo nombre clave es Kettle, es una herramienta de la suite de Pentaho de las que se denomina ETL (Extract – Transform – Load), es decir, una herramienta de Extracción de datos de una fuente, Transformación de esos datos y Carga de esos datos en otro sitio. El uso de Kettle permite evitar grandes cargas de trabajo manual frecuentemente difícil de mantener y de desplegar. Estas tareas son típicas en procesos de migración, integración con terceros, explotación de Big Data, etc, y en general se podría decir que son necesarias en casi cualquier proyecto mediano o grande.

5. ESQUEMAS DE BASES DE DATOS

5.1 Esquema Base de Datos OLTP



5.2 Esquema Base de Datos OLAP



6. CAPTURAS DE PANTALLA DEL PROCESO ETL

Al comenzar con el proceso de ETL, debemos empezar definiendo los ítems que se van a utilizar para hacer la transformación. Se comienza definiendo los procesos para las tablas de dimensiones, ya que son las tablas independientes de otras, posteriormente se definen los procesos para la tabla de hechos ya que su información depende de las tablas de dimensión.

6.1 Ítems Utilizados

- ✓ Este es un paso que nos permite a través de una consulta



Table input

obtener información de la base de datos OLTP para así poder llenar o agregar la información necesaria en una tabla dentro de la base de datos OLAP.

- ✓ En cuanto a la selección de valores es un paso que nos



Select values

permite la conexión de datos de la base de datos OLTP con la base de datos OLAP; entre las cosas que se nos permiten hacer al usar este ítem están cambiar tipos de datos, cambiar el nombre de los valores, configurar la longitud, etc.

- ✓ Este es un paso para la salida de datos en el cual los datos



Table output

que han sido obtenidos a través de la carga de datos de los pasos anteriormente seleccionados son insertados en la tabla de la base de datos OLAP.

- ✓ Este paso lo que hace es buscar una fila en la tabla buscando



Insert / Update

claves de búsqueda, en el caso de no ser encontradas, estas se insertan, para cuando la clave es encontrada entonces los campos que se actualizan son los mismos y no se hace nada.

6.2 Consultas utilizadas para el proceso ETL

6.2.1 Consulta para llenar la Tabla de Dimensión de Clientes:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure with the 'BD_PROYECTO' database selected. The central pane displays a T-SQL query:

```
SELECT ID_CUST, NAME_CUST, NIF_CUST FROM CUSTOMER
```

The results pane shows the output of the query, listing 20 rows of customer data:

ID_CUST	NAME_CUST	NIF_CUST
1	Tarah August	234-04-7095
2	Reyna Kellbell	618-77-3007
3	Arly Pidge	734-99-5913
4	Amin Folkard	461-01-2871
5	Rodrigo Esmead	214-53-8686
6	Fredrika Redfield	462-51-3861
7	Andy Edgcombe	845-93-2242
8	Fredi Coolahan	674-20-0355
9	Eugene Beadey	536-96-2403
10	Bill Serpent	344-37-8962
11	Tam Cankett	368-16-1288
12	Emmeline Jeffs	719-13-4746
13	Alen Brng	460-85-2467
14	Elfrieda Hiddley	108-82-2856
15	Angl Clemmons	689-41-6407
16	Kerwin Padignton	116-97-0681
17	Earvin Benham	102-03-5149
18	Melania Weafer	156-17-9782
19	Mary Mewis	716-30-3664
20	Hildegaard De Buyne	755-94-0767

At the bottom, a message indicates the query was executed successfully.

6.2.2 Consulta para llenar la Tabla de Dimensión de Empleados:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure with the 'BD_PROYECTO' database selected. The central pane displays a T-SQL query:

```
SELECT ID_EMPLOYEE, CONCAT(FIRST_NAME, ' ', LAST_NAME) Nombre FROM EMPLOYEES
```

The results pane shows the output of the query, listing 27 rows of employee data:

ID_EMPLOYEE	Nombre
1	Melanie Douglass
2	Ashla Turville
3	Rollins Tipping
4	Kely Hounsom
5	Geradina Hyder
6	Jocko Angell
7	Dell Barthup
8	Randolf Gaynor
9	Nonnah Giorgio
10	Frances Garaway
11	Cosmo MacShany
12	Rochette Sinne
13	Yancy Teuler
14	Beme Downie
15	Rane Shitland
16	Jena Gullen
17	Valentia Iacovone
18	Serge Ottey
19	Theodore Hudd
20	Miton Blincoe
21	Camella Lawday
22	Tobie Krolch
23	Lacy Fred
24	Robinet Chatelot
25	Ainc Laws
26	Luciana Huffston
27	Karolta Hellmore

At the bottom, a message indicates the query was executed successfully.

6.2.3 Consulta para llenar la Tabla de Dimensión de Propiedades:

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the BD_PROYECTO database. The central pane displays the results of a query against the PROPERTY table:

```
SELECT ID_PROPERTY, NIF_PROPERTY, PROPERTY_TYPE FROM PROPERTYS
```

ID_PROPERTY	NIF_PROPERTY	PROPERTY_TYPE
1	282-82-8223	Apartment
2	653-55-2525	Apartment
3	242-39-0737	Condominium
4	550-07-7411	Home
5	227-84-7012	Home
6	448-45-9290	Condominium
7	451-41-3542	Residential
8	359-28-6324	Apartment
9	726-95-3289	Condominium
10	860-24-5116	Home
11	450-14-176	Commercial
12	791-69-9541	Apartment
13	253-36-5394	Parking
14	528-18-1243	Commercial
15	744-75-2664	Home
16	639-90-2858	Residential
17	606-53-9720	Apartment
18	639-46-5207	Home
19	688-65-6181	Condominium
20	658-07-6108	Residential
21	568-17-2953	Apartment
22	558-93-8817	Parking
23	677-59-7772	Condominium
24	121-37-2288	Residential
25	573-62-6063	Apartment
26	800-42-7999	Commercial
27	129-37-4524	Home
28	496-68-1181	Commercial

Query executed successfully.

6.2.4 Consulta para llenar la Tabla de Dimensión de Proveedores:

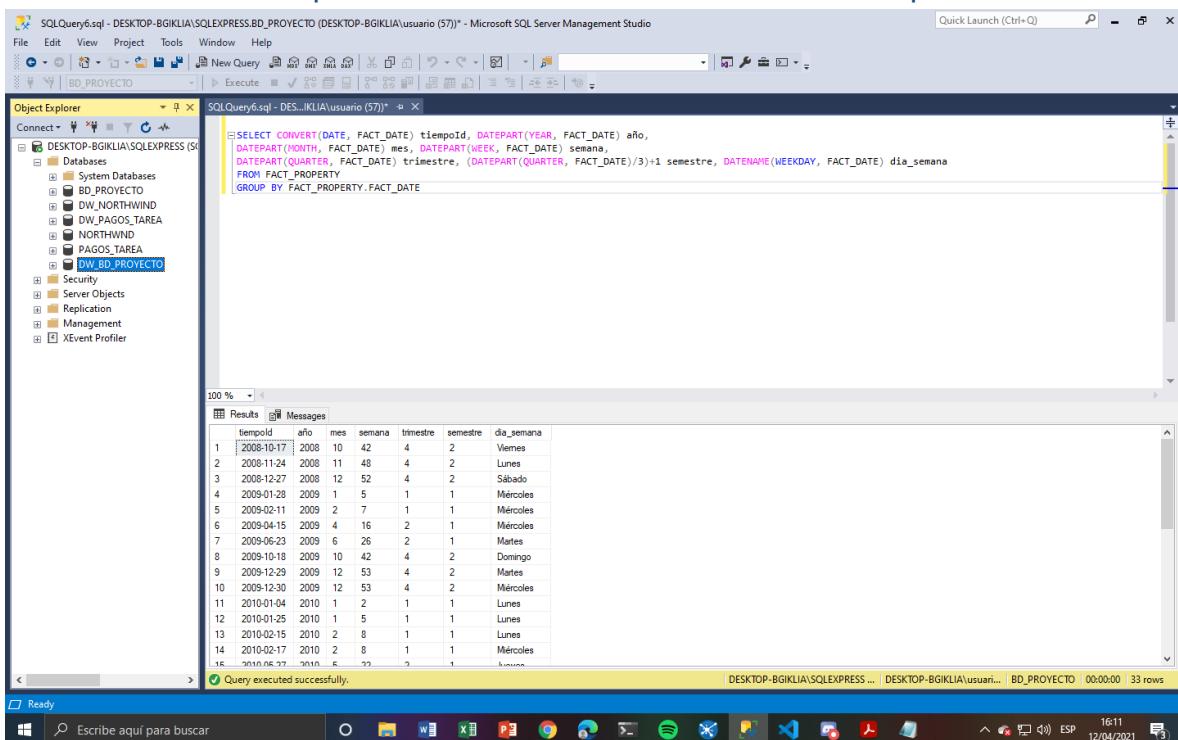
The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the BD_PROYECTO database. The central pane displays the results of a query against the SUPPLIERS table:

```
SELECT ID_SUP, NAME_SUP, NIF_SUP FROM SUPPLIERS
```

ID_SUP	NAME_SUP	NIF_SUP
1	Odelinda Millott	767-44-9561
2	Roven McClecole	475-12-7450
3	Levey Ferson	814-04-6811
4	Katharine Greich	895-31-3867
5	Leland Uverseeep	889-21-0148
6	Rlin Wilatton	493-39-0301
7	Fidela Milbum	189-85-7831
8	Rowan Jayne	283-78-0751
9	Trenna Davydchey	810-21-1267
10	Casie Smythman	627-27-5773
11	Stephan Ahfeld	396-64-5002
12	Sue Liderbrauss	878-30-3681
13	Bekky Glaston	576-35-5107
14	Ingaborg Wellard	459-89-7548
15	Angie McOwan	605-75-3259
16	Taddeus Abella	278-20-2789
17	Mylene Gusticke	438-69-3736
18	Felicidad Jaques	189-65-3733
19	Sheila Abet	742-06-7474
20	Gard Haston	102-88-8347
21	Florette Nozzolini	483-87-7098
22	Tera Eringey	515-33-7821
23	Gaye Peleg	623-30-3344
24	Lef Rookwell	689-73-2743
25	Pandora Gracewood	360-83-9686
26	Garett Note	231-96-0191
27	Lelia Bartunek	195-36-8415
28	Stilmann Ivesedko	300-24-8366
29	Teodor Bixley	334-98-4173
30	Uywelyn Tatleton	723-75-5271

Query executed successfully.

6.2.5 Consulta para llenar la Tabla de Dimensión de Tiempo:

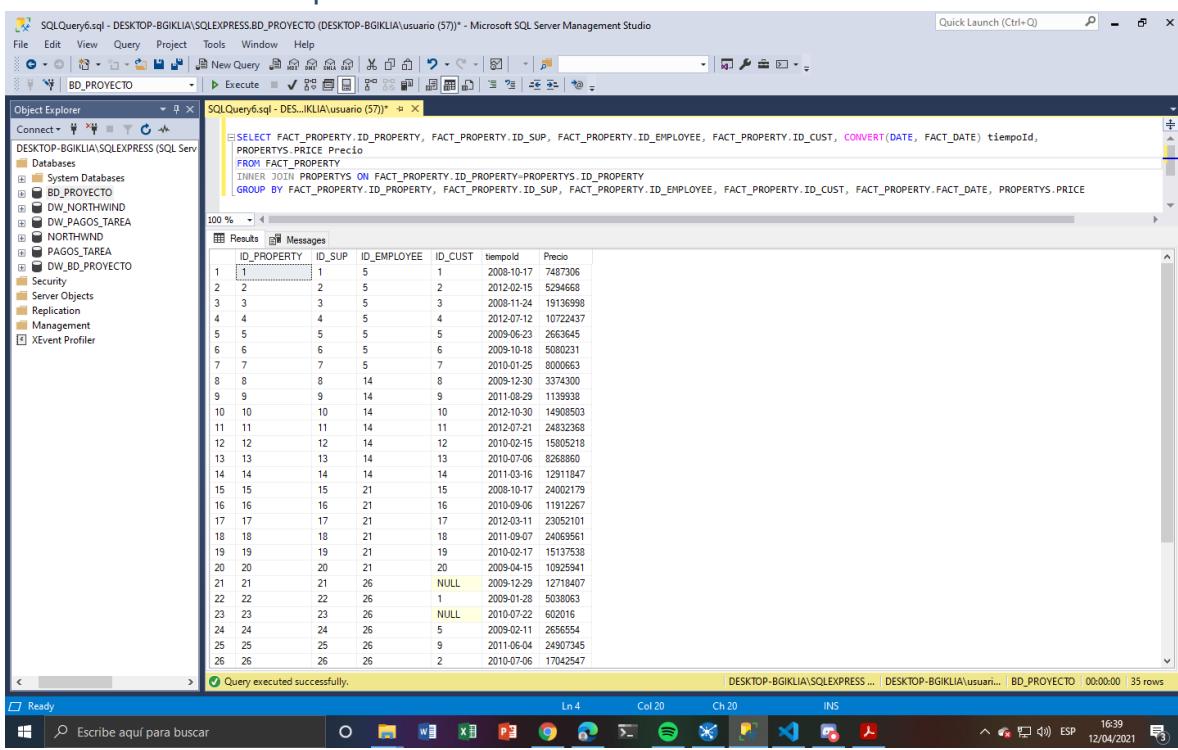


```

SELECT CONVERT(DATE, FACT_DATE) tiempoId, DATEPART(YEAR, FACT_DATE) año,
DATEPART(MONTH, FACT_DATE) mes, DATEPART(WEEK, FACT_DATE) semana,
DATEPART(QUARTER, FACT_DATE) trimestre, (DATEPART(QUARTER, FACT_DATE)/3)+1 semestre, DATENAME(WEEKDAY, FACT_DATE) dia_semana
FROM FACT_PROPERTY
GROUP BY FACT_PROPERTY.FACT_DATE
    
```

The screenshot shows the Object Explorer on the left with the database 'BD_PROYECTO' selected. The results pane displays a table with columns: tempoid, año, mes, semana, trimestre, semestre, dia_semana. The data starts from 2008-10-17 and ends at 2010-02-17, showing 33 rows successfully executed.

6.2.6 Consulta para llenar la Tabla de Hechos de Ventas:



```

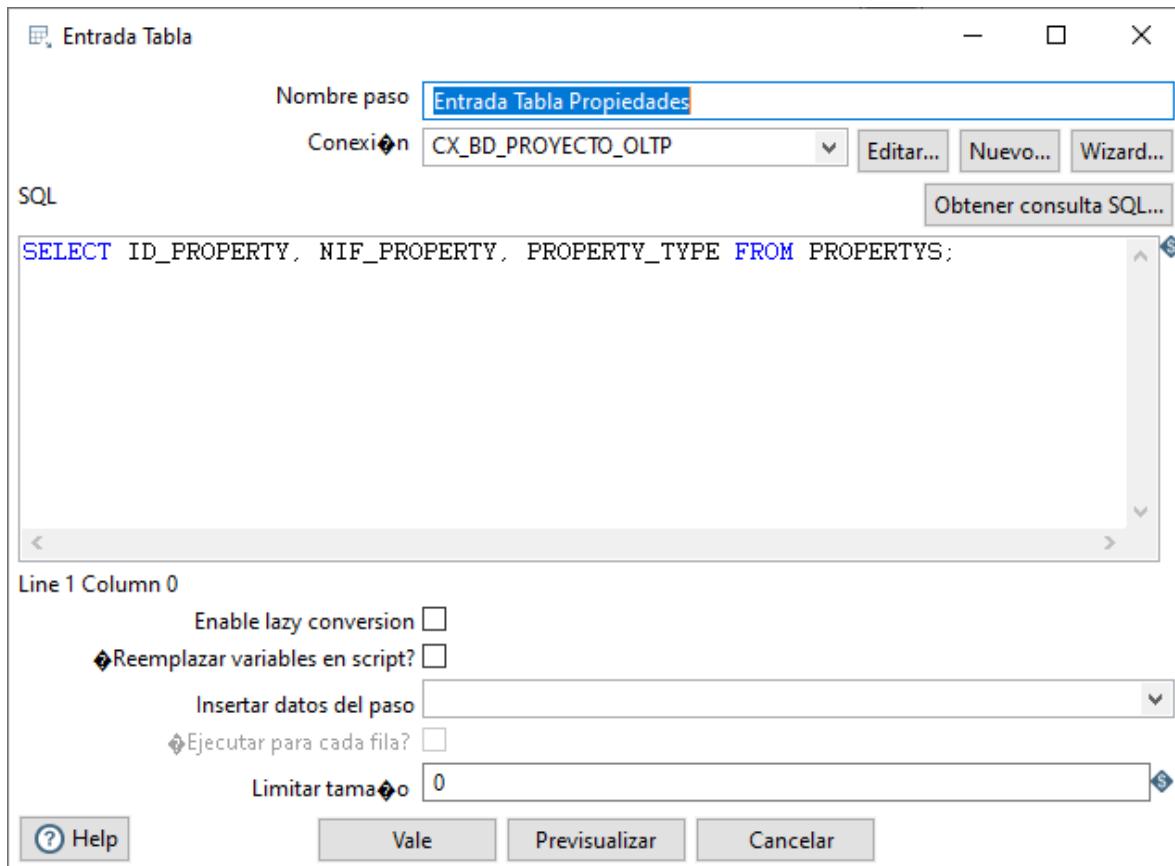
SELECT FACT_PROPERTY.ID_PROPERTY, FACT_PROPERTY.ID_SUP, FACT_PROPERTY.ID_EMPLOYEE, FACT_PROPERTY.ID_CUST, CONVERT(DATE, FACT_DATE) tiempoId,
PROPERTIES.PRICE Precio
FROM FACT_PROPERTY
INNER JOIN PROPERTIES ON FACT_PROPERTY.ID_PROPERTY=PROPERTIES.ID_PROPERTY
GROUP BY FACT_PROPERTY.ID_PROPERTY, FACT_PROPERTY.ID_SUP, FACT_PROPERTY.ID_EMPLOYEE, FACT_PROPERTY.ID_CUST, FACT_PROPERTY.FACT_DATE, PROPERTIES.PRICE
    
```

The screenshot shows the Object Explorer on the left with the database 'BD_PROYECTO' selected. The results pane displays a table with columns: ID_PROPERTY, ID_SUP, ID_EMPLOYEE, ID_CUST, tempoid, Precio. The data spans from 2008-10-17 to 2010-07-06, showing 35 rows successfully executed.

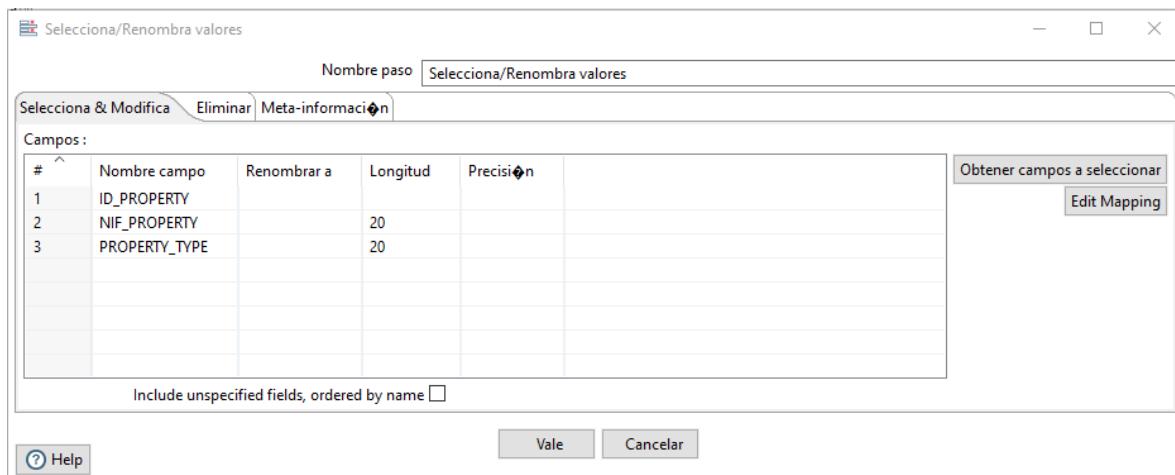
6.3 Configuración de Ítems para el ETL

6.3.1 ETL Tabla de Dimensión de Propiedades

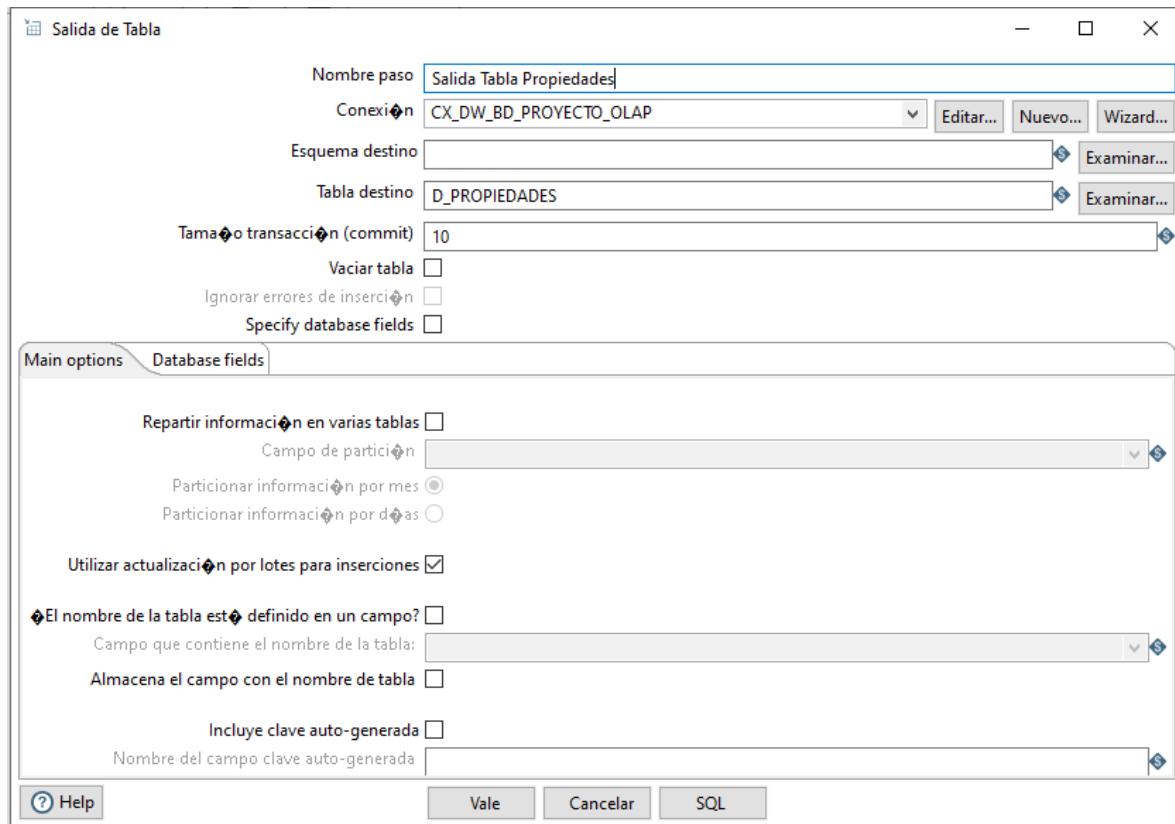
6.3.1.1 Entrada de datos



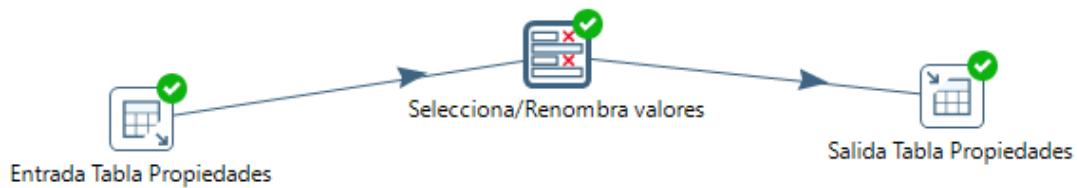
6.3.1.2 Transformación de Datos



6.3.1.3 Salida de Datos

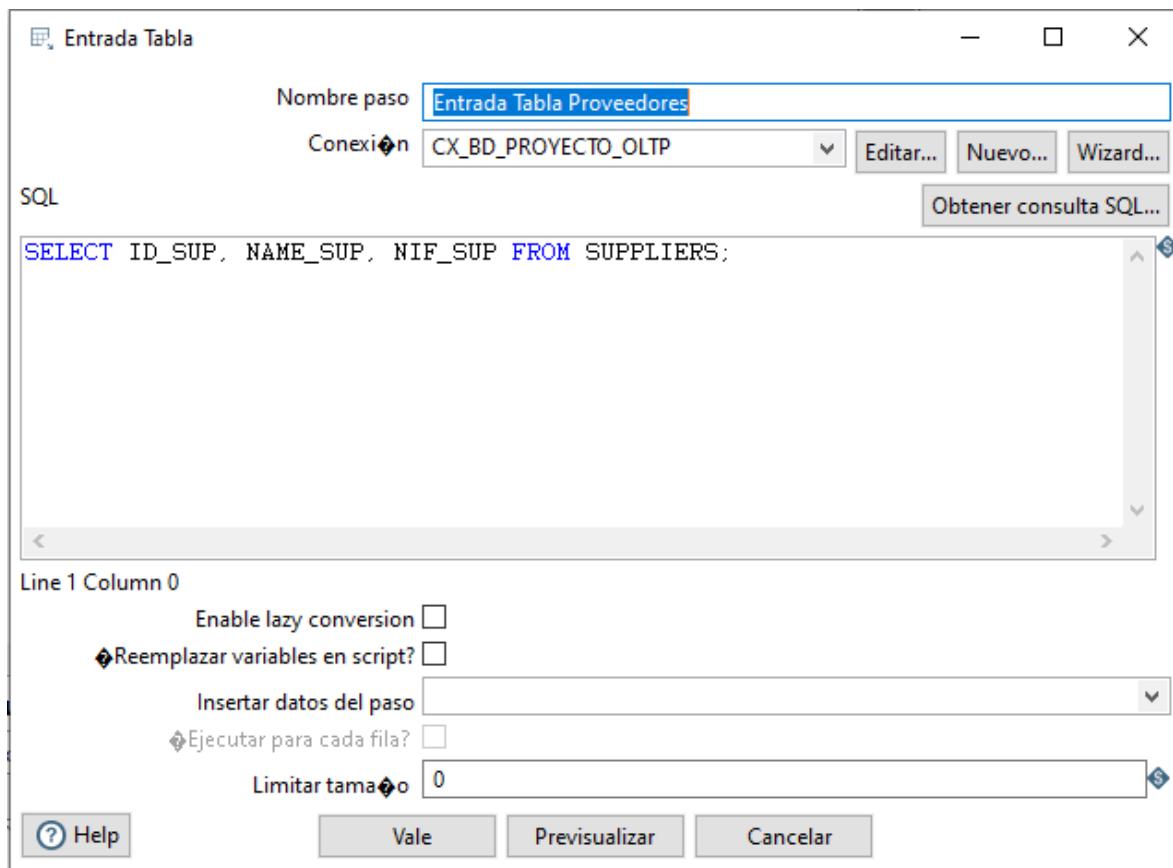


6.3.1.4 Conexión exitosa de los 3 ítems

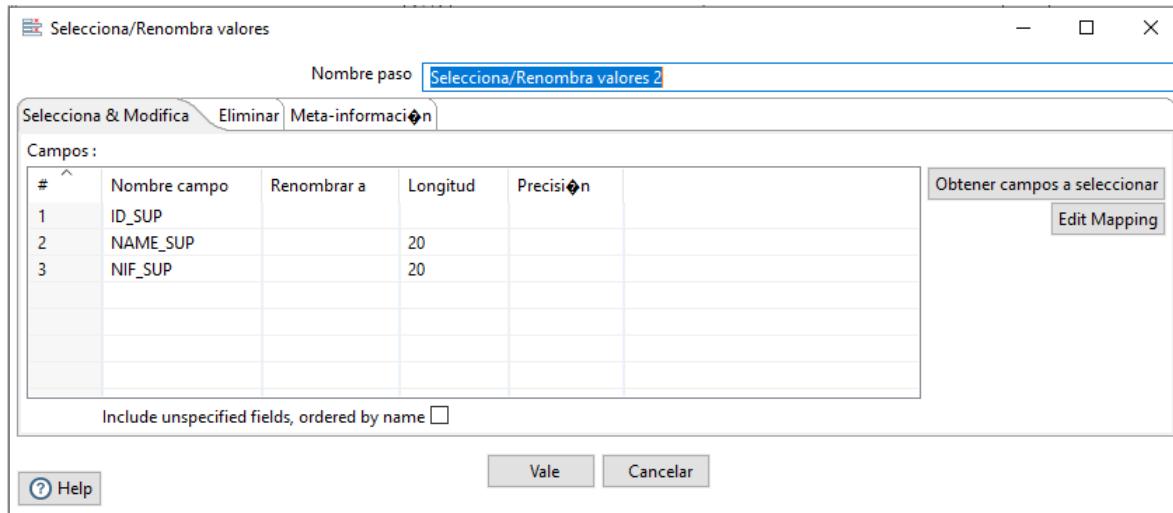


6.3.2 ETL Tabla de Dimensión de Proveedores

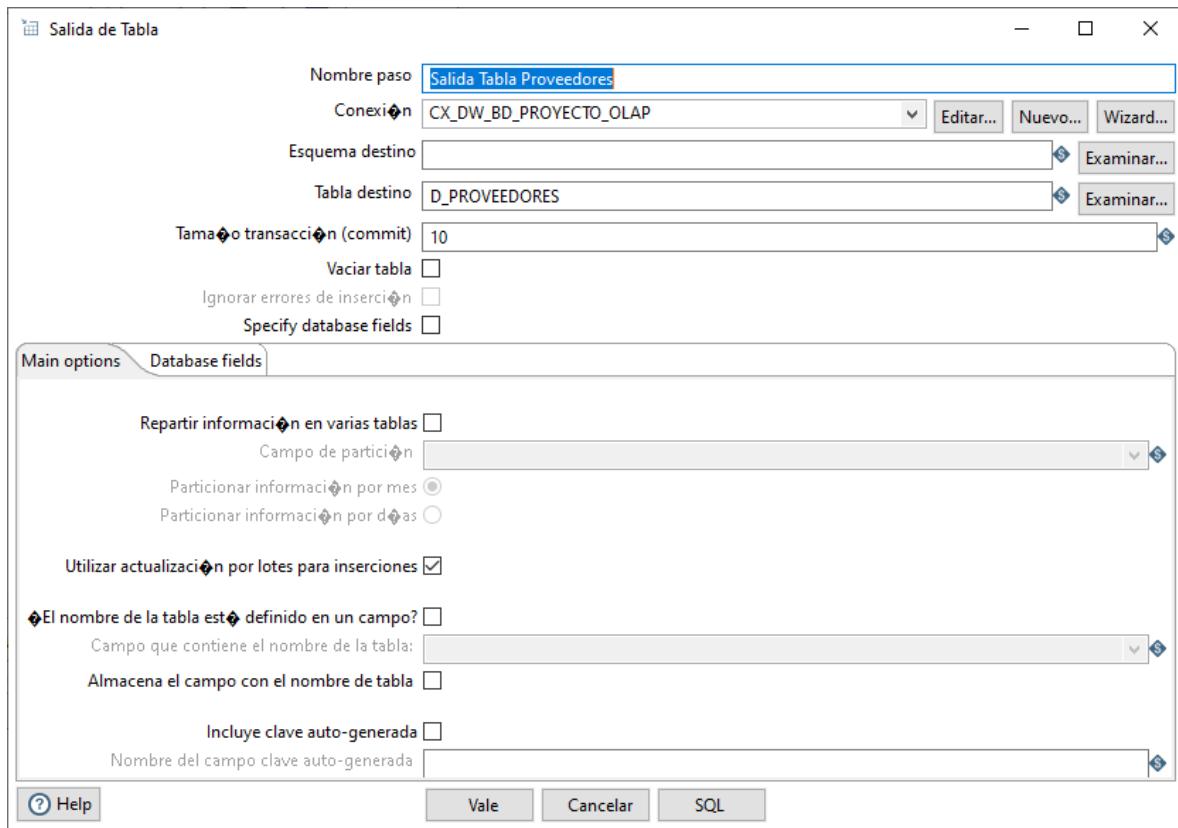
6.3.2.1 Entrada de datos



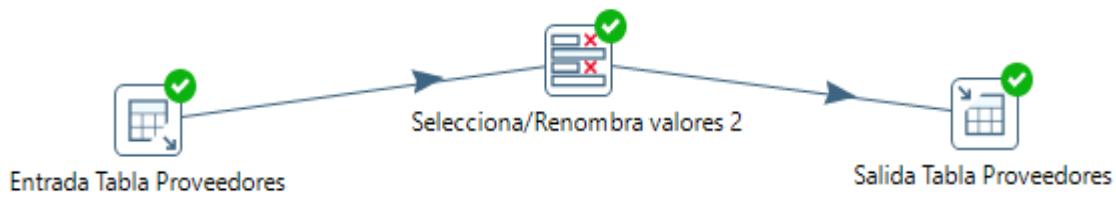
6.3.2.2 Transformación de Datos



6.3.2.3 Salida de Datos

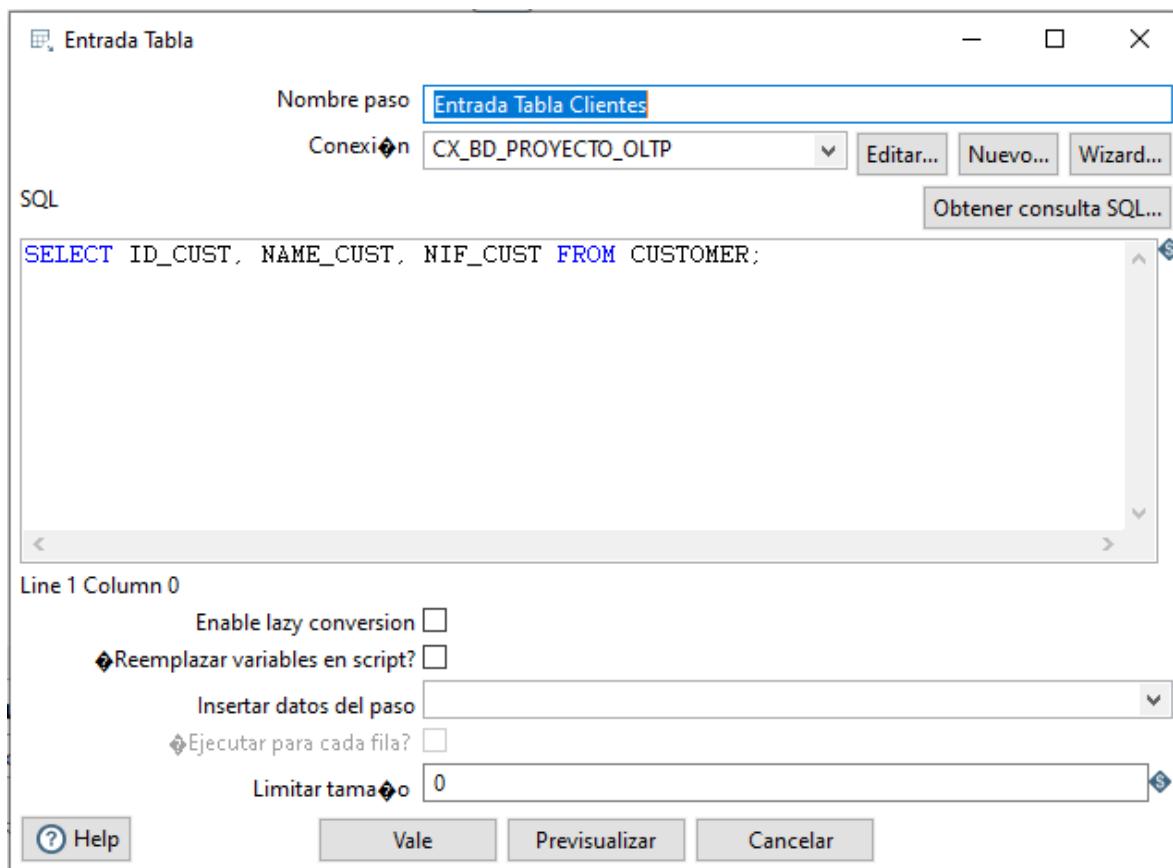


6.3.2.4 Conexión exitosa de los 3 ítems

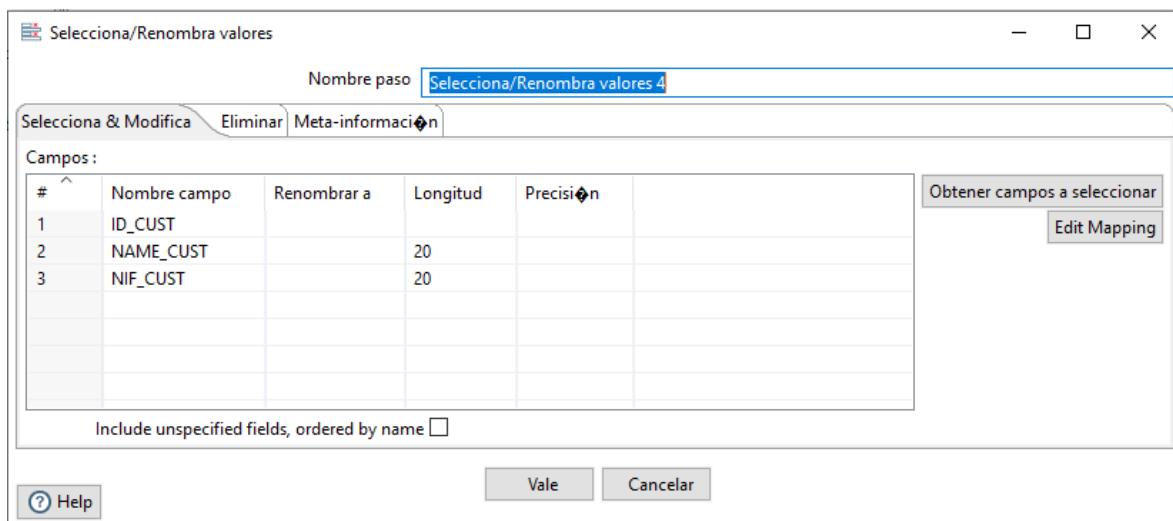


6.3.3 ETL Tabla de Dimensión de Clientes

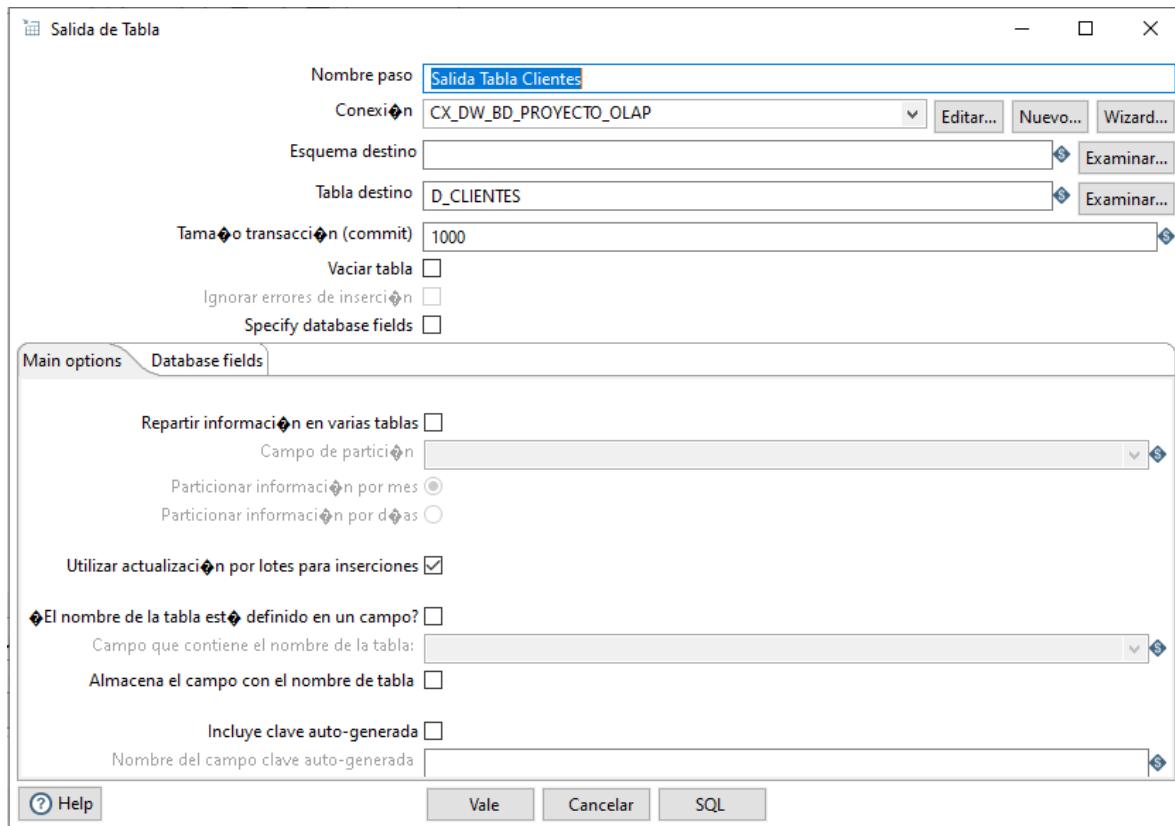
6.3.3.1 Entrada de datos



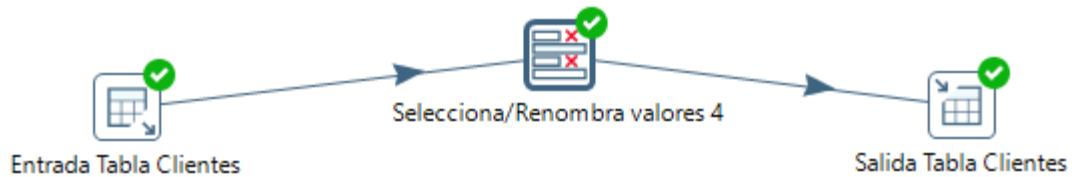
6.3.3.2 Transformación de Datos



6.3.3.3 Salida de Datos

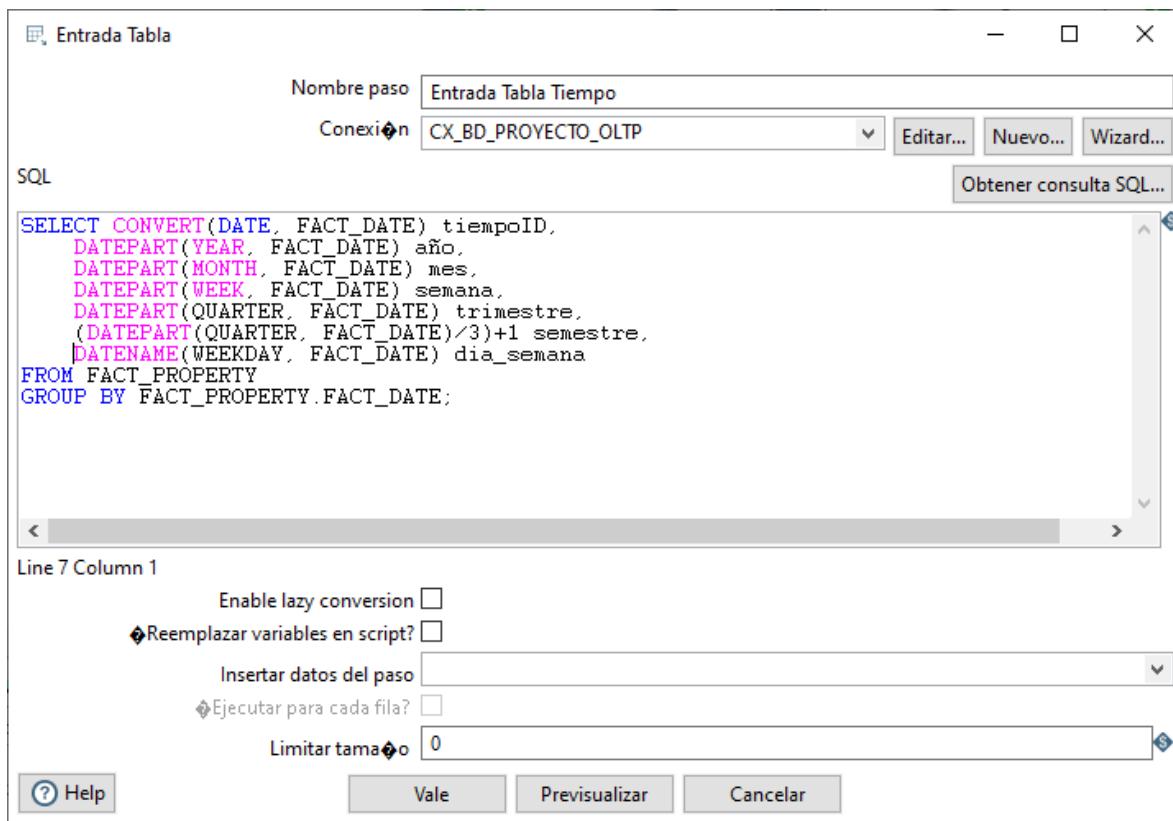


6.3.3.4 Conexión exitosa de los 3 ítems

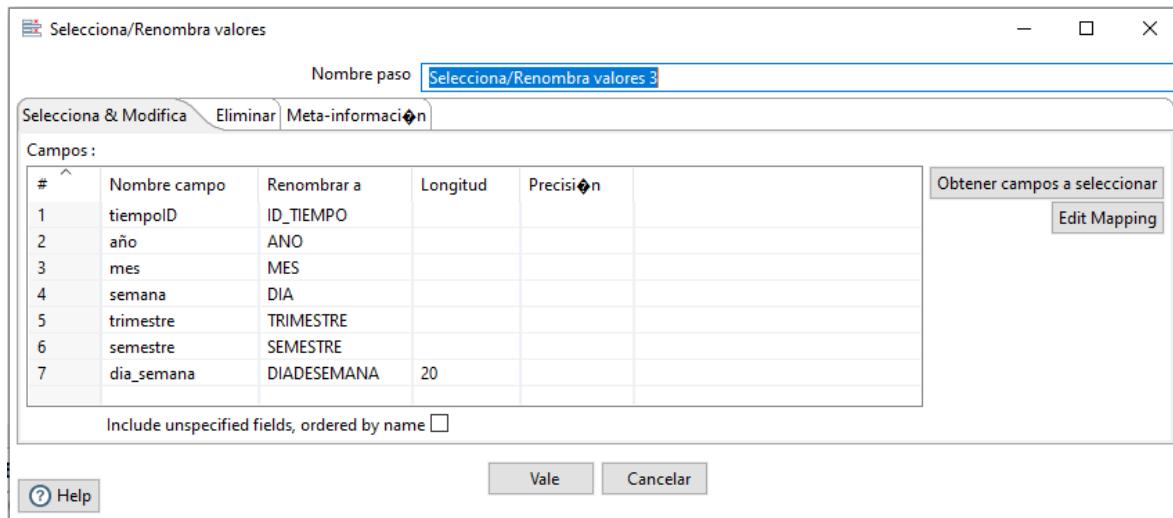


6.3.4. ETL Tabla de Dimensión de Tiempo

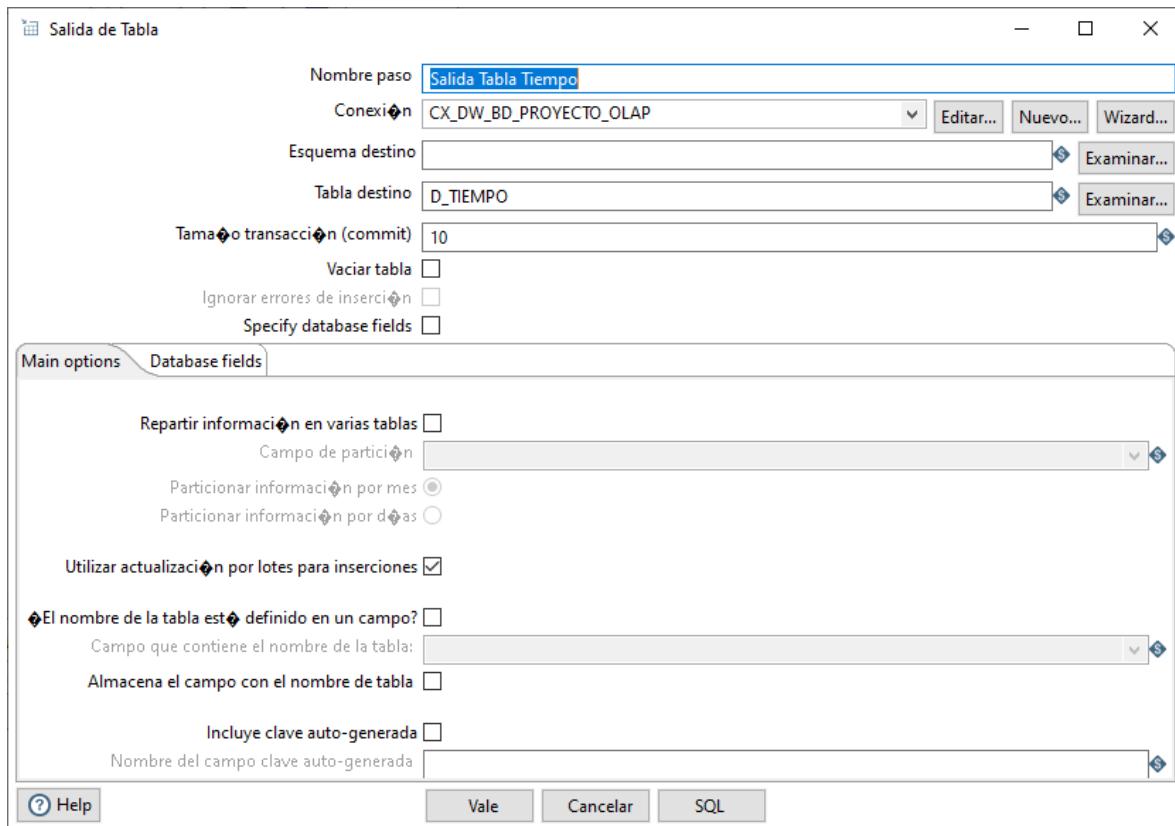
6.3.4.1 Entrada de datos



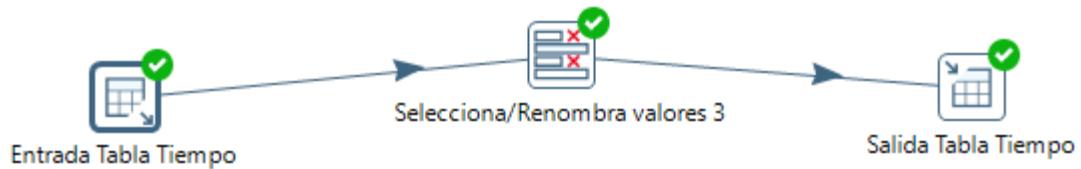
6.3.4.2 Transformación de Datos



6.3.4.3 Salida de Datos

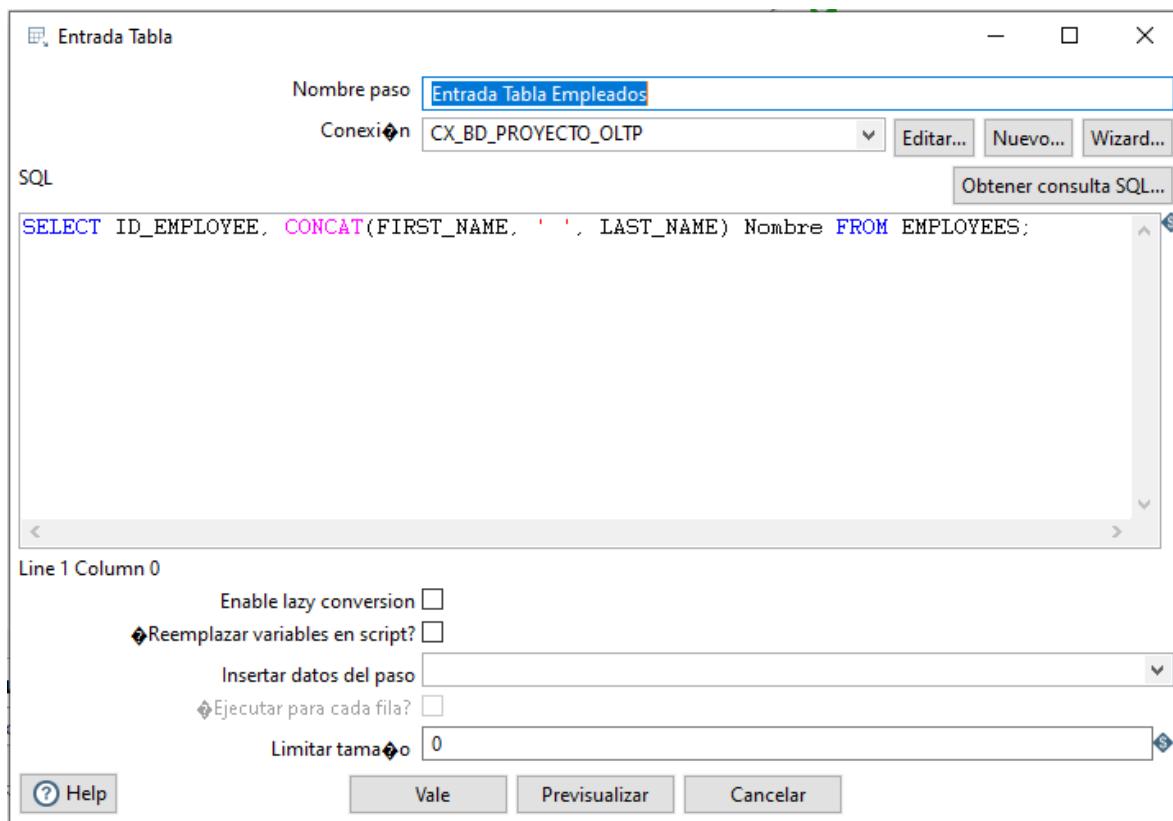


6.3.4.4 Conexión exitosa de los 3 ítems

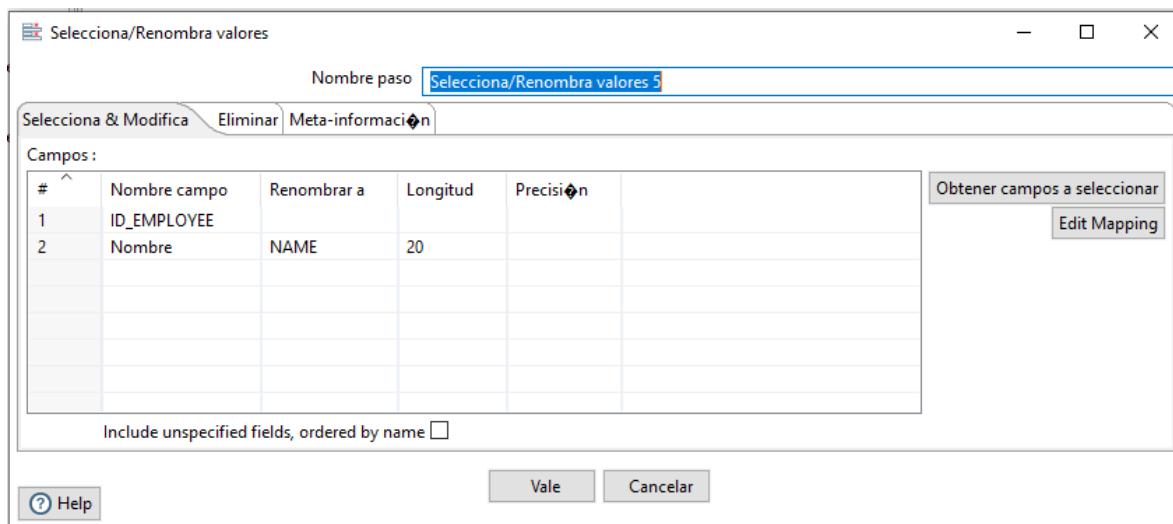


6.3.5 ETL Tabla de Dimensión de Empleados

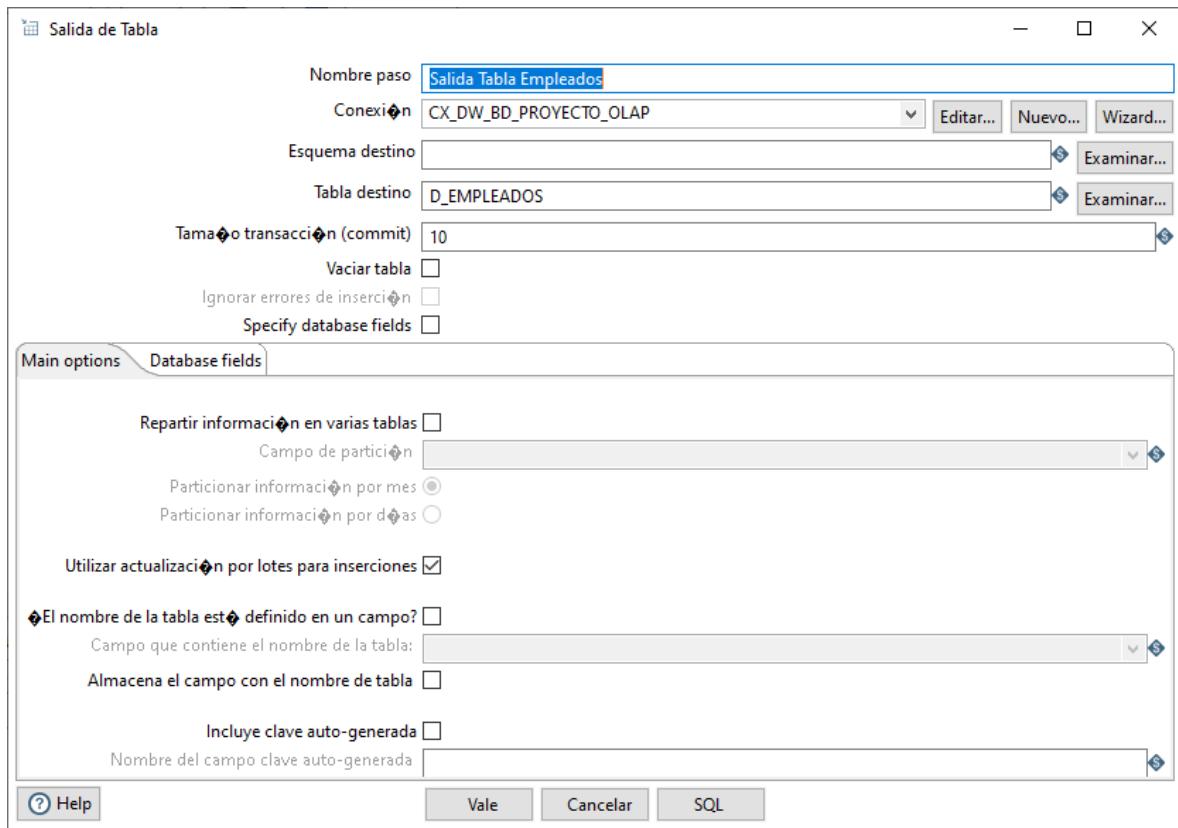
6.3.5.1 Entrada de datos



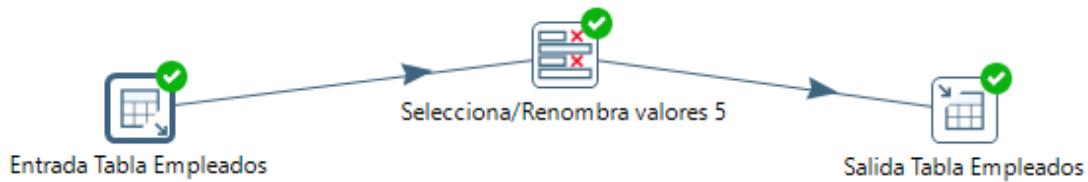
6.3.5.2 Transformación de Datos



6.3.5.3 Salida de Datos



6.3.5.4 Conexión exitosa de los 3 ítems



6.3.6 ETL Tabla de Hechos de Ventas

6.3.6.1 Entrada de datos

The screenshot shows the 'Entrada Tabla' (Table Input) step configuration window. The 'Nombre paso' (Step Name) is set to 'Entrada Tabla H_Ventas'. The 'Conexión' (Connection) is set to 'CX_BD_PROYECTO'. The SQL query is defined as follows:

```
SELECT FACT_PROPERTY.ID_PROPERTY,
       FACT_PROPERTY.ID_SUP,
       FACT_PROPERTY.ID_EMPLOYEE,
       FACT_PROPERTY.ID_CUST,
       CONVERT(DATE, FACT_DATE) tiempoid,
       PROPERTYS.PRICE Precio
  FROM FACT_PROPERTY
 INNER JOIN PROPERTYS ON FACT_PROPERTY.ID_PROPERTY=PROPERTYS.ID_PROPERTY
 GROUP BY FACT_PROPERTY.ID_PROPERTY,
          FACT_PROPERTY.ID_SUP,
          FACT_PROPERTY.ID_EMPLOYEE,
          FACT_PROPERTY.ID_CUST,
          FACT_PROPERTY.FACT_DATE,
          PROPERTYS.PRICE
```

Below the SQL area, there are several configuration options:

- Enable lazy conversion
- Reemplazar variables en script?
- Insertar datos del paso
- Ejecutar para cada fila?
- Limits the number of rows to 0.

At the bottom are 'Vale' (OK), 'Previsualizar' (Preview), and 'Cancelar' (Cancel) buttons.

6.3.6.2 Transformación de Datos

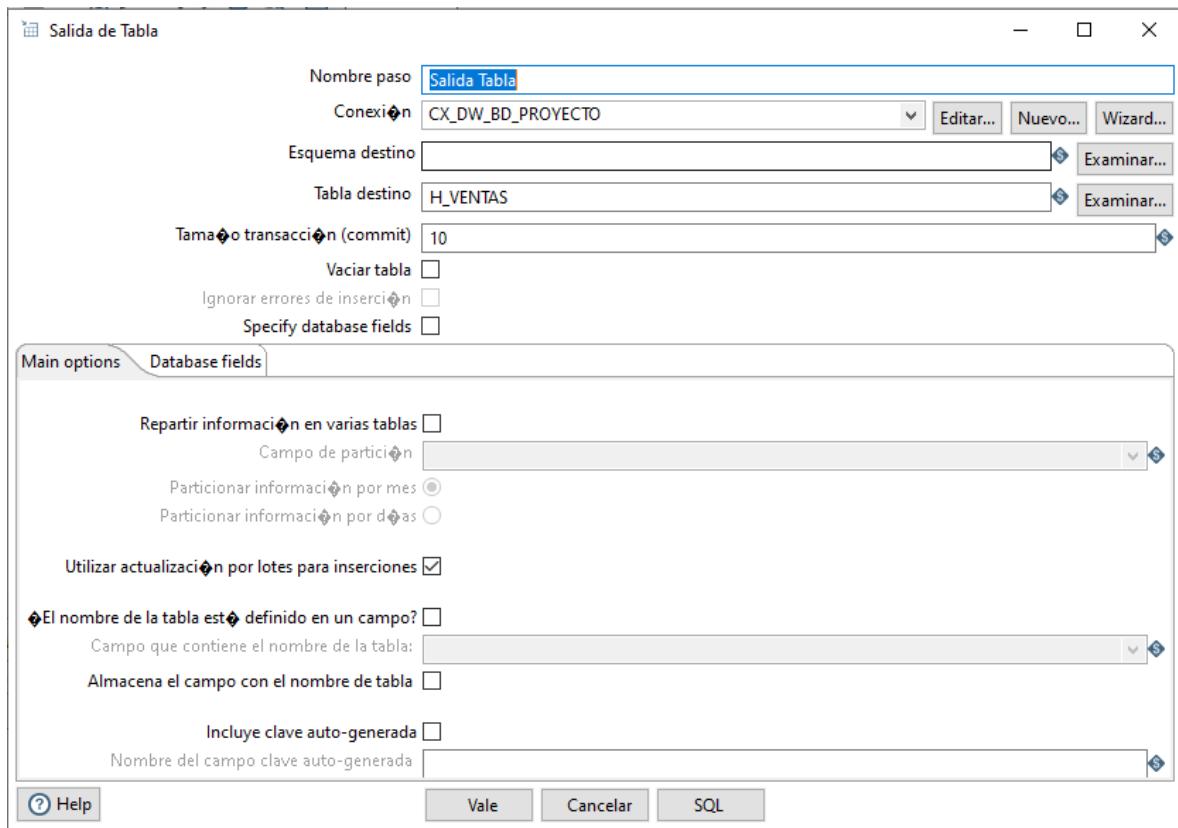
The screenshot shows the 'Selecciona/Renombra valores' (Select/Rename values) transformation configuration window. The 'Nombre paso' (Step Name) is set to 'Selecciona/Renombra valores'. The 'Campos' (Fields) table lists the following mappings:

#	Nombre campo	Renombrar a	Longitud	Precision	
1	ID_PROPERTY				
2	ID_SUP				
3	ID_EMPLOYEE				
4	ID_CUST				
5	tiempold	ID_TIEMPO			
6	Precio	PRECIO			

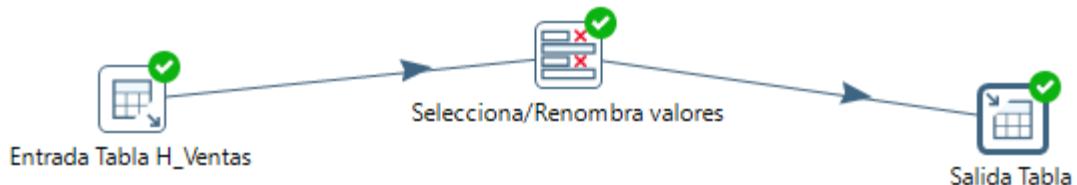
On the right side, there are buttons for 'Obtener campos a seleccionar' (Get fields to select) and 'Edit Mapping'.

At the bottom are 'Vale' (OK), 'Cancelar' (Cancel), and 'Help' buttons.

6.3.6.3 Salida de Datos

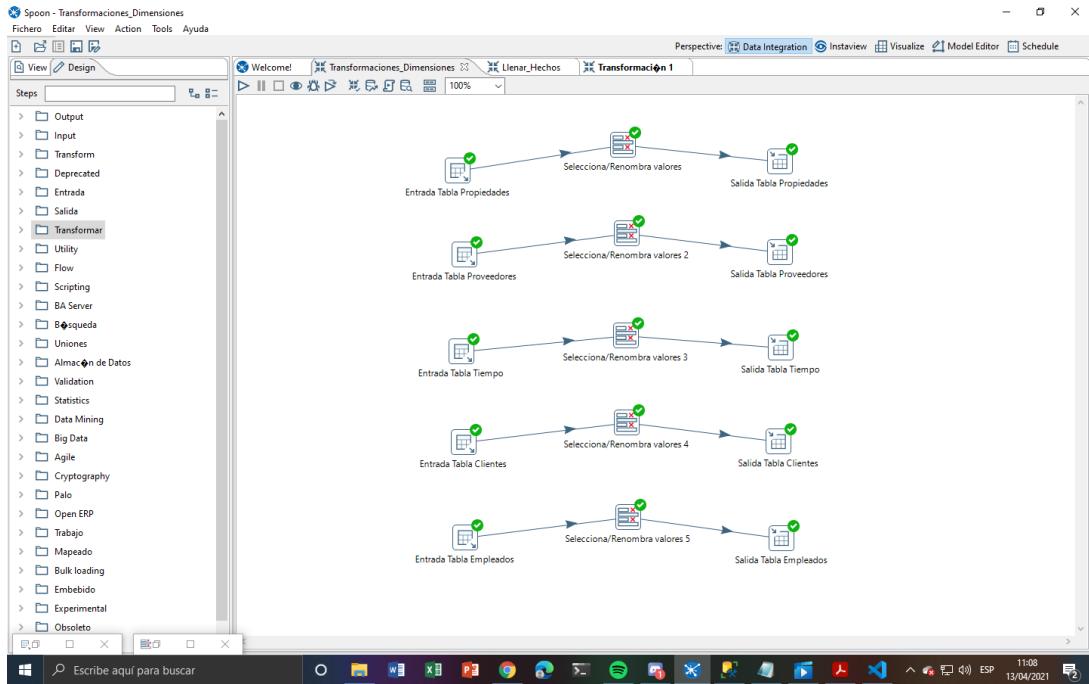


6.3.6.4 Conexión exitosa de los 3 ítems

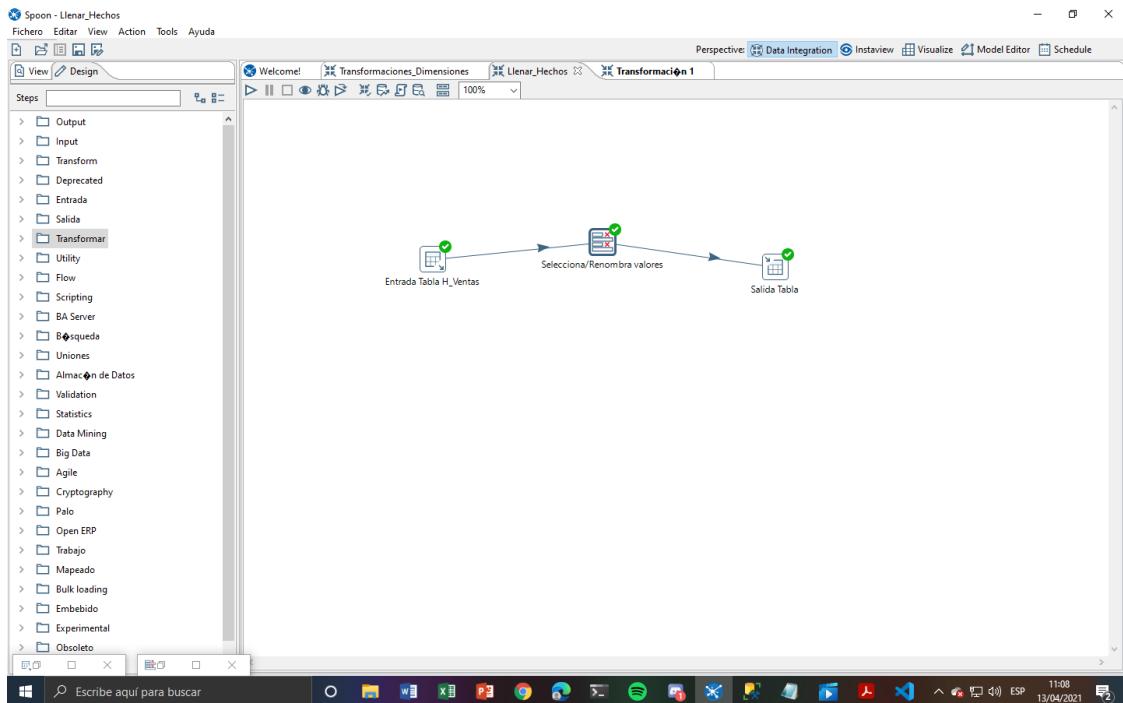


6.4 Ejecución exitosa de todos los ETL

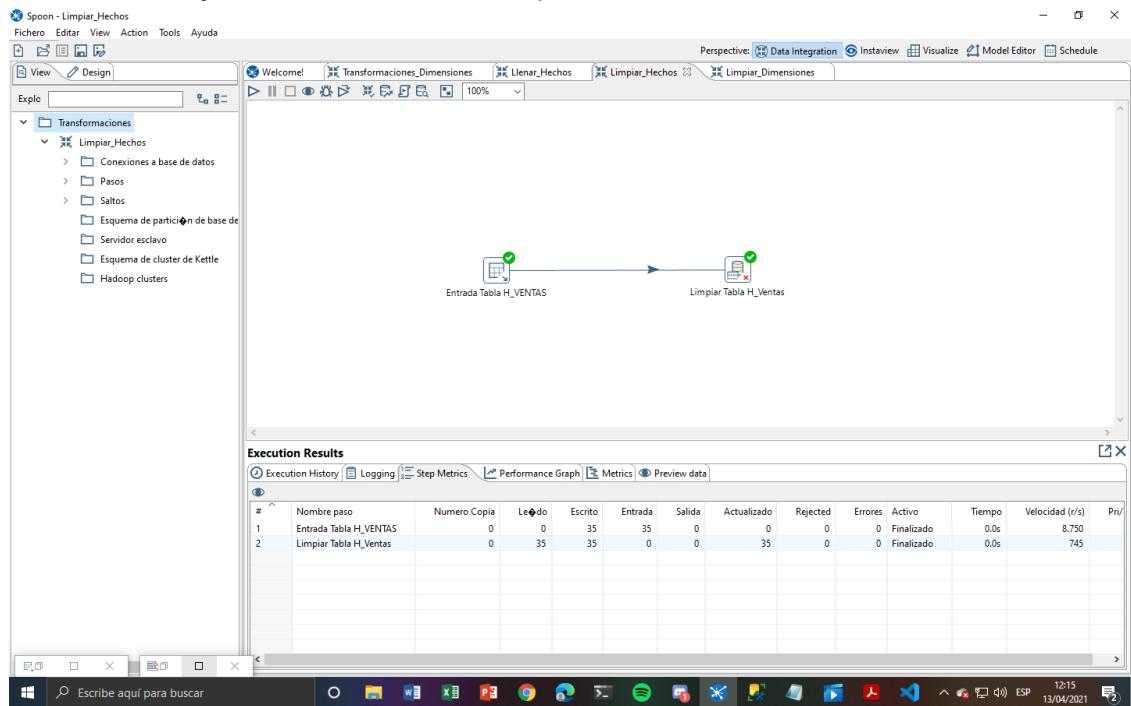
6.4.1 Ejecución de ETL de Llenar Dimensiones



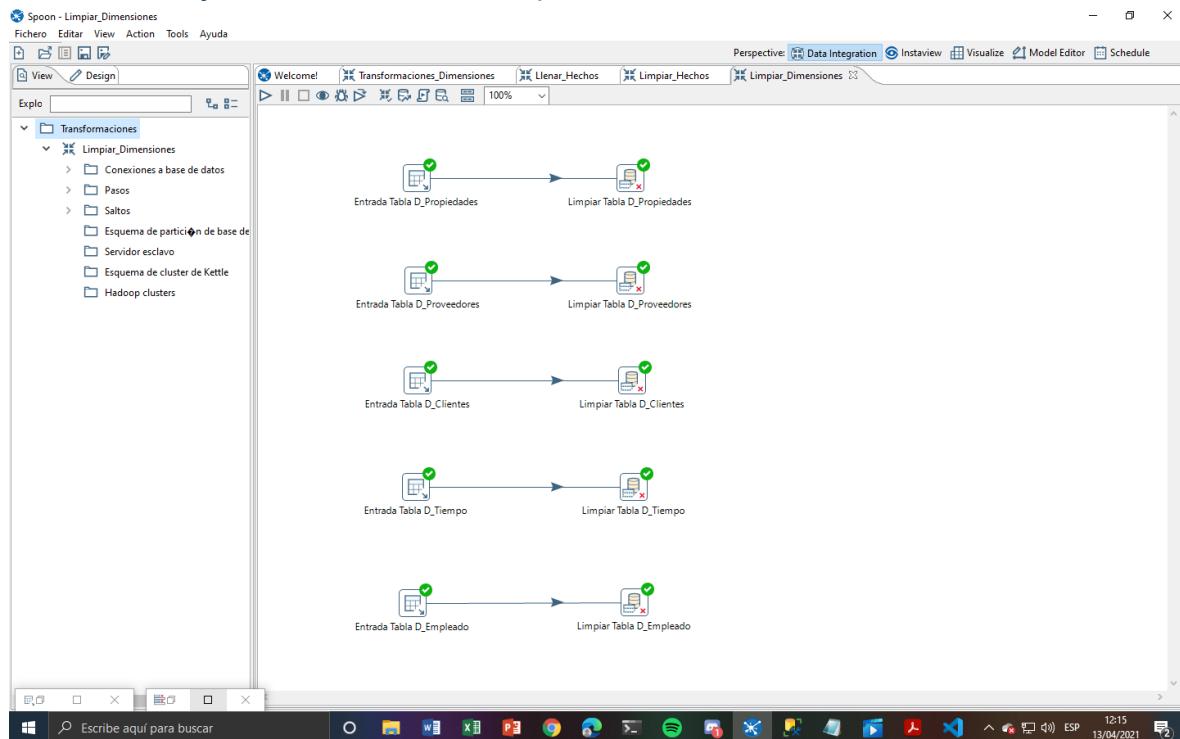
6.4.2 Ejecución de ETL de Llenar Hechos



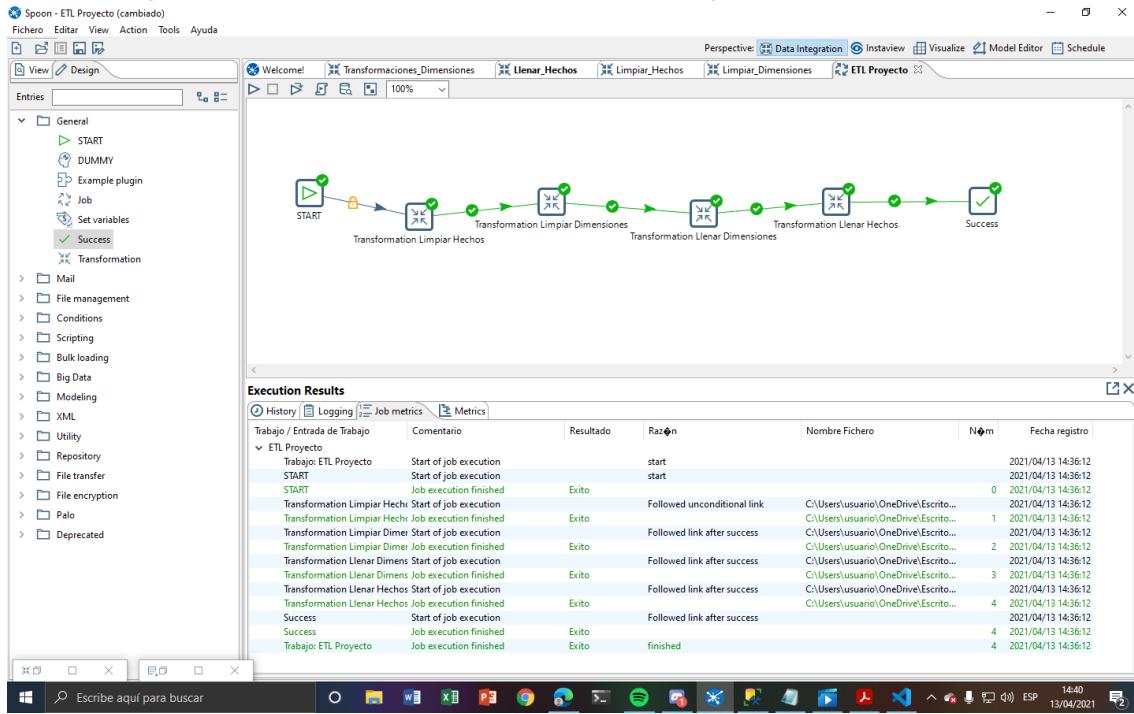
6.4.3 Ejecución de ETL de Limpiar Hechos



6.4.4 Ejecución de ETL de Limpiar Dimensiones



6.5 Ejecución exitosa del JOB para ejecutar los ETL



6.6 Ejecución de consultas para ver registros

6.6.1 Tabla D_CLIENTES

```
Object Explorer
Connect - DESKTOP-BGIKLIA\usuario (55) -> X
DESKTOP-BGIKLIA\SQLEXPRESS (SQL)
  <--> Databases
    BD_PROYECTO
    DW_BD_PROYECTO
      <--> Tables
        dbo.D_CLIENTES
          ID_CUST NAME_CUST NIF_CUST
          1 Taran August 234-04-7095
          2 Reyna Kelsell 618-77-3007
          3 Aly Piddie 734-99-5913
          4 Amin Folkard 461-01-2871
          5 Dodge Esmed 214-53-8698
          6 Fredrika Redfield 462-51-3861
          7 Andy Edcombe 845-93-2242
          8 Fred Coolahan 674-20-0355
          9 Eugene Beazley 536-96-2403
          10 Bill Sergent 344-37-8962
          11 Tam Cankett 358-16-1288
          12 Emmeline Jeffs 719-13-4746
          13 Alen Bing 460-85-2487
          14 Alfreda Hiddlesey 108-82-2856
          15 Angril Clemens 689-41-6407
          16 Kerwin Padig... 116-97-0681
          17 Eavin Berham 102-03-5149
          18 Melania Weifer 156-17-9782
          19 Mary Mewis 716-30-3664
          20 Hildegaard De ... 755-94-0767
```

Query executed successfully.

6.6.2 Tabla D_EMPLEADOS

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the DW_BD_PROYECTO database which contains tables like D_CLIENTES, D_EMPLEADOS, D_PROPIEDADES, etc. The central pane displays the results of a query:

```

SELECT TOP 1000 [ID_EMPLOYEE]
      ,[NAME]
   FROM [DW_BD_PROYECTO].[dbo].[D_EMPLEADOS]
  
```

The results table has columns ID_EMPLOYEE and NAME, containing 40 rows of employee data.

ID_EMPLOYEE	NAME
1	Melanie Douglass
2	Ashia Turville
3	Rollins Tipping
4	Kelly Housom
5	Getadina Hyder
6	Jocko Angell
7	Dell Bathup
8	Randolf Gaynford
9	Norman Gergo
10	Frances Garaway
11	Cozmo MacShany
12	Rochette Sinnie
13	Yancy Teuler
14	Bene Downie
15	Raine Shaland
16	Jena Guillen
17	Valerita Iacovone
18	Serge Ottey
19	Theodore Hudd
20	Milton Bincoe
21	Camella Lawday
22	Tobie Krolch
23	Lacy Fredi
24	Robinet Chatelot
25	Alic Laws
26	Luciana Hurleston
27	Karotta Helmire
28	Cordey Troyes

At the bottom, a message says "Query executed successfully." The status bar shows the connection details: DESKTOP-BGIKLIA\SQLEXPRESS... | DESKTOP-BGIKLIA\usuario... | DW_BD_PROYECTO | 00:00:00 | 40 rows.

6.6.3 Tabla D_PROPIEDADES

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left shows the database structure, including the DW_BD_PROYECTO database which contains tables like D_CLIENTES, D_EMPLEADOS, D_PROPIEDADES, etc. The central pane displays the results of a query:

```

SELECT TOP 1000 [ID_PROPERTY]
      ,[NIF_PROPERTY]
      ,[TYPE_PROPERTY]
      ,[PROPERTY_TYPE]
   FROM [DW_BD_PROYECTO].[dbo].[D_PROPIEDADES]
  
```

The results table has columns ID_PROPERTY, NIF_PROPERTY, TYPE_PROPERTY, and PROPERTY_TYPE, containing 35 rows of property data.

ID_PROPERTY	NIF_PROPERTY	TYPE_PROPERTY	PROPERTY_TYPE
1	282-82-9823		Apartment
2	655-55-2525		Apartment
3	242-39-0737		Condominium
4	550-07-7411		Home
5	227-84-7012		Home
6	448-45-9290		Condominium
7	451-41-3542		Residential
8	359-28-6924		Apartment
9	736-95-3829		Condominium
10	860-24-5116		Home
11	490-41-6176		Commercial
12	791-09-8541		Apartment
13	253-36-9394		Parking
14	528-18-1243		Commercial
15	744-75-2664		Home
16	639-90-2858		Residential
17	606-53-9720		Apartment
18	699-46-5207		Home
19	688-65-6181		Condominium
20	698-07-6108		Residential
21	568-17-2893		Apartment
22	558-93-8817		Parking
23	677-59-7772		Condominium
24	121-37-8288		Residential
25	573-62-9603		Apartment
26	800-42-7999		Commercial
27	129-37-4524		Home
28	496-68-1181		Commercial

At the bottom, a message says "Query executed successfully." The status bar shows the connection details: DESKTOP-BGIKLIA\SQLEXPRESS... | DESKTOP-BGIKLIA\usuario... | DW_BD_PROYECTO | 00:00:00 | 35 rows.

6.6.4 Tabla D_PROVEEDORES

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists the database structure, including the DW_BD_PROYECTO database which contains tables like D_CLIENTES, D_EMPLEADOS, D_PROPEDADES, D_TIEMPO, and D_PROVEEDORES. The Results tab on the right displays the output of a query:

```

SELECT TOP 1000 [ID_SUP]
      ,[NAME_SUP]
      ,[NIF_SUP]
   FROM [DW_BD_PROYECTO].[dbo].[D_PROVEEDORES]
  
```

The results table has columns ID_SUP, NAME_SUP, and NIF_SUP. The data shows 30 rows of supplier information:

ID_SUP	NAME_SUP	NIF_SUP
1	Odeinda Millet	767-44-9561
2	Rowen McIlcole	475-12-4540
3	Levey Ferson	814-04-8311
4	Katharine Gresch	895-31-3867
5	Leland Liversege	889-10-0148
6	Flint Wilson	493-39-0301
7	Fidela Nilburn	189-85-7831
8	Rowan Jaynes	283-78-7571
9	Trenna Davydchev	810-21-2627
10	Cassie Smythyma	627-75-7773
11	Stephan Ashford	396-64-5002
12	Sue Lindenstrasse	878-30-6681
13	Beky Glotter	576-35-1107
14	Ingaborg Welford	469-89-7548
15	Ange Mcowan	605-75-3259
16	Tadeusz Abela	278-20-2789
17	Mylene Gutjich	438-63-7376
18	Feliciad Jaques	189-53-7373
19	Shela Aber	742-06-7474
20	Gard Haston	102-88-3347
21	Florette Nozzolini	483-87-0798
22	Tera Ernigey	515-33-7821
23	Gaye Palos	623-30-3344
24	Lef Roskwell	689-73-7473
25	Pandora Gracewood	360-83-9688
26	Ganett Note	231-96-0191
27	Lella Bartunek	195-36-8415
28	Stilmann Ivasechko	300-24-8366

Query executed successfully.

6.6.5 Tabla D_TIEMPO

The screenshot shows the Microsoft SQL Server Management Studio interface. The Object Explorer on the left lists the database structure, including the DW_BD_PROYECTO database which contains tables like D_CLIENTES, D_EMPLEADOS, D_PROPEDADES, D_PROVEEDORES, and D_TIEMPO. The Results tab on the right displays the output of a query:

```

SELECT TOP 1000 [ID_TIEMPO]
      ,[ANO]
      ,[MES]
      ,[DIA]
      ,[TRIMESTRE]
      ,[SEMANESTRE]
      ,[DIADESEMANA]
   FROM [DW_BD_PROYECTO].[dbo].[D_TIEMPO]
  
```

The results table has columns ID_TIEMPO, ANO, MES, DIA, TRIMESTRE, SEMESTRE, and DIADESEMANA. The data shows 33 rows of time period information:

ID_TIEMPO	ANO	MES	DIA	TRIMESTRE	SEMANESTRE	DIADESEMANA
2008-10-17	2008	10	42	4	2	Vierres
2008-11-24	2008	11	48	4	2	Lunes
3	2008-12-27	2008	12	52	4	Sábado
4	2009-01-28	2009	1	5	1	Mércores
5	2009-02-11	2009	2	7	1	Mércores
6	2009-04-15	2009	4	16	2	Mércores
7	2009-06-23	2009	6	26	2	Martes
8	2009-10-18	2009	10	42	4	Domingo
9	2009-12-29	2009	12	53	4	Martes
10	2009-12-30	2009	12	53	4	Mércores
11	2010-01-04	2010	1	2	1	Lunes
12	2010-01-25	2010	1	5	1	Lunes
13	2010-02-15	2010	2	8	1	Lunes
14	2010-02-17	2010	2	8	1	Mércores
15	2010-05-27	2010	5	22	2	Jueves
16	2010-07-06	2010	7	28	3	Martes
17	2010-07-22	2010	7	30	3	Jueves
18	2010-09-06	2010	9	37	3	Lunes
19	2010-10-04	2010	10	41	4	Lunes
20	2011-03-03	2011	3	10	1	Jueves
21	2011-03-16	2011	3	12	1	Mércores
22	2011-06-04	2011	6	23	2	Sábado
23	2011-07-10	2011	7	28	3	Domingo
24	2011-08-29	2011	8	36	3	Lunes

Query executed successfully.

6.6.6 Tabla H_VENTAS

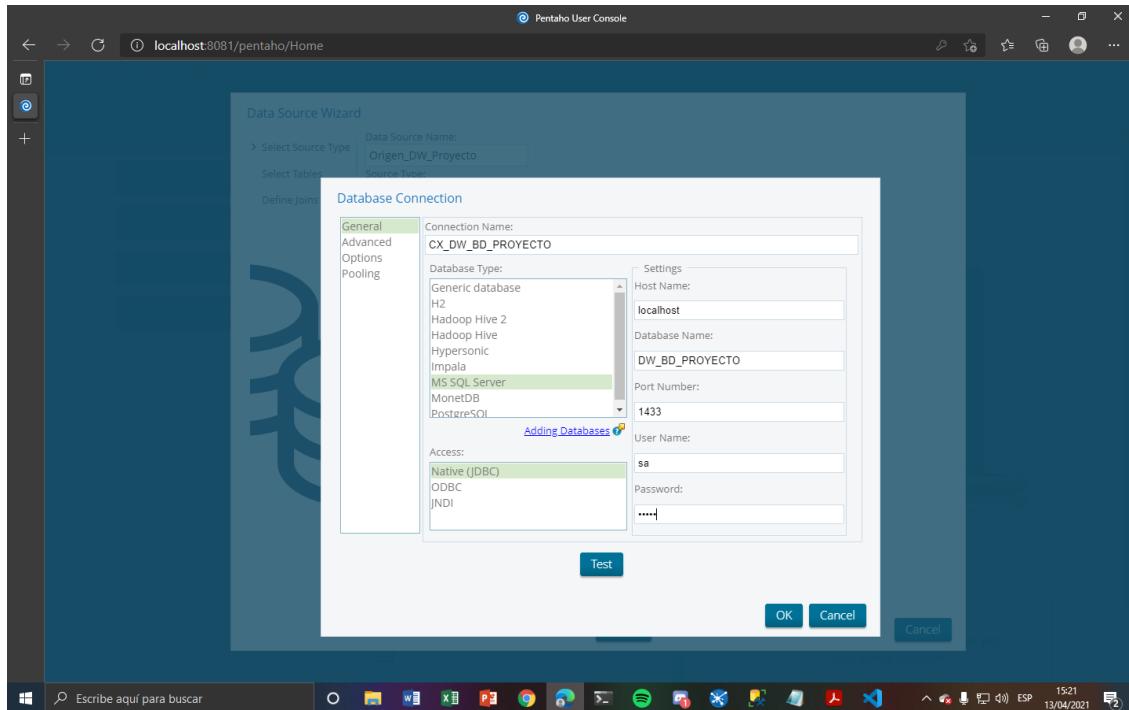
The screenshot shows the Microsoft SQL Server Management Studio interface. In the Object Explorer, the database 'DW_BD_PROYECTO' is selected. In the center pane, a query window displays the following SQL code:

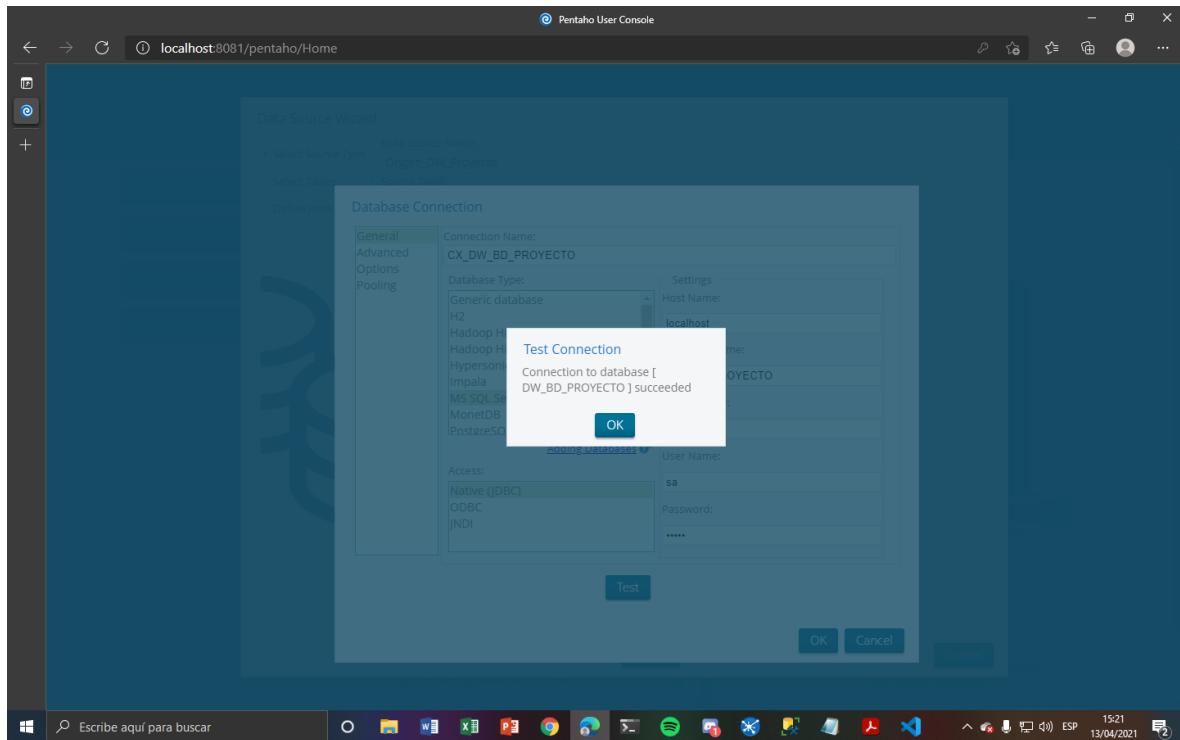
```
SELECT TOP (1000) [ID_VENTAS]
      ,[ID_PROPERTY]
      ,[ID_SUP]
      ,[ID_EMPLOYEE]
      ,[ID_CUST]
      ,[ID_TIEMPO]
      ,[PRECIO]
  FROM [DW_BD_PROYECTO].[dbo].[H_VENTAS]
```

The results grid shows 35 rows of data from the H_VENTAS table. The columns are: ID_VENTAS, ID_PROPERTY, ID_SUP, ID_EMPLOYEE, ID_CUST, ID_TIEMPO, and PRECIO. The data includes various IDs and dates ranging from 2008-10-17 to 2011-09-29.

7. Creación del Cubo OLAP

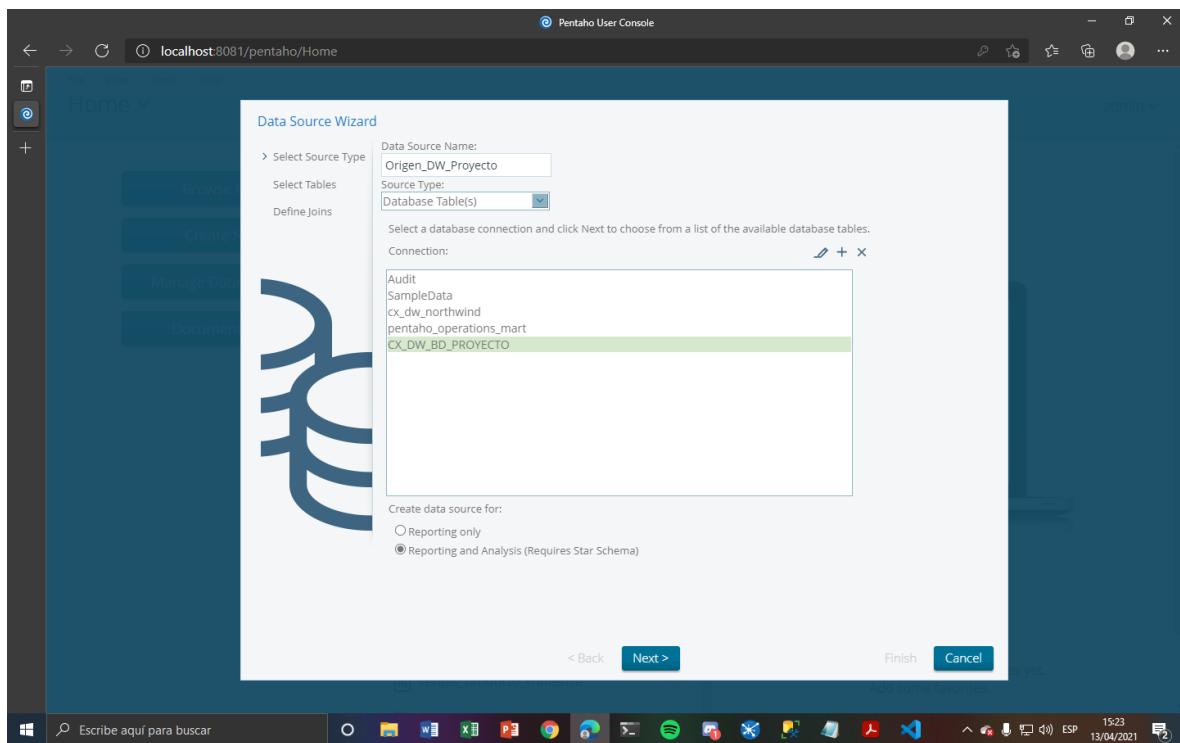
7.1 Conexión con la base de datos OLAP

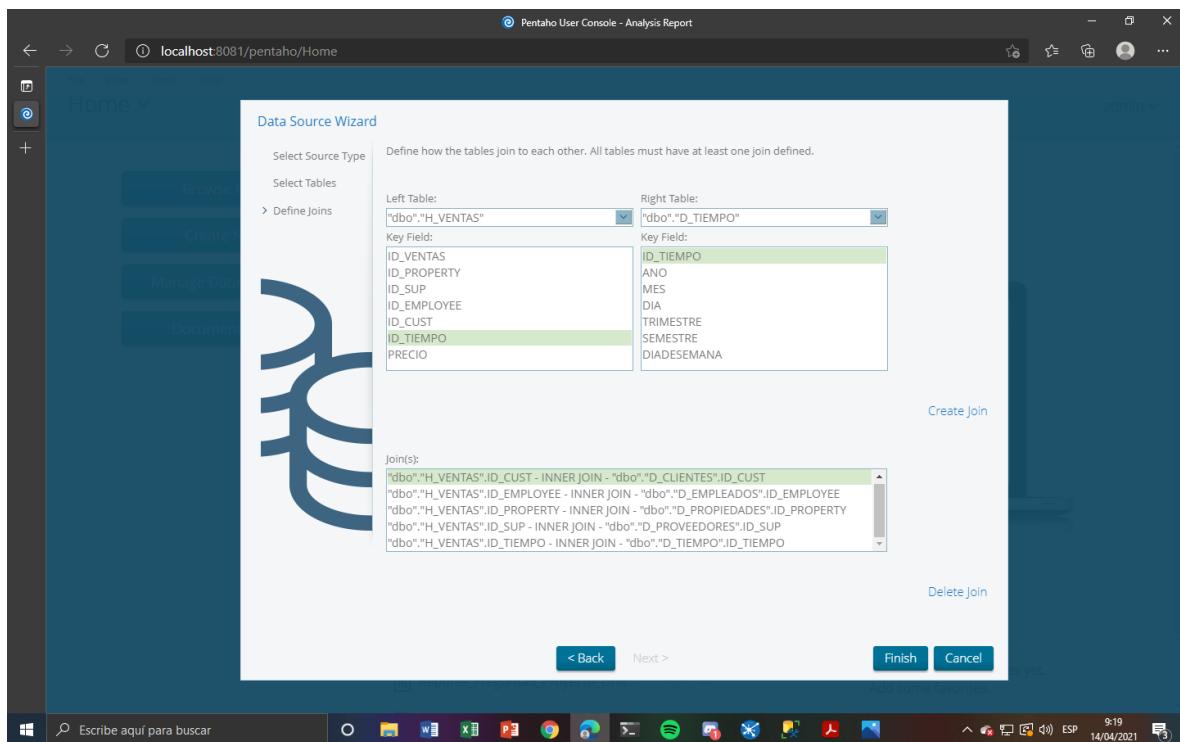
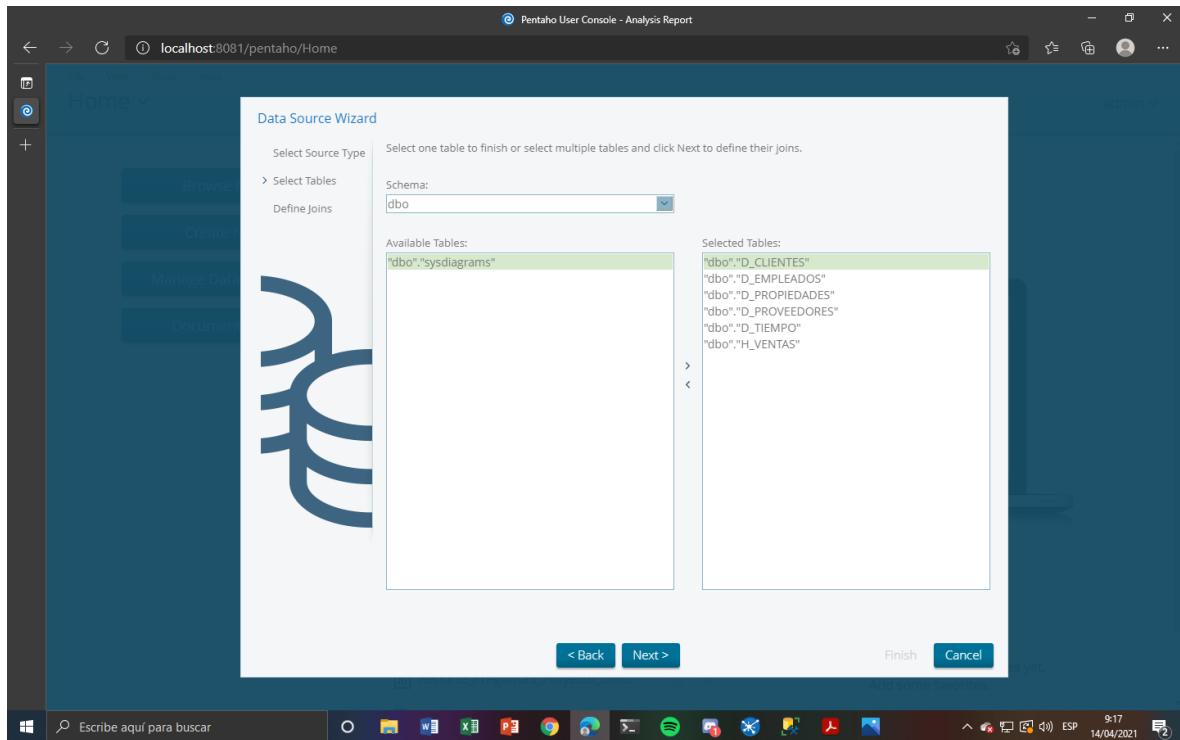




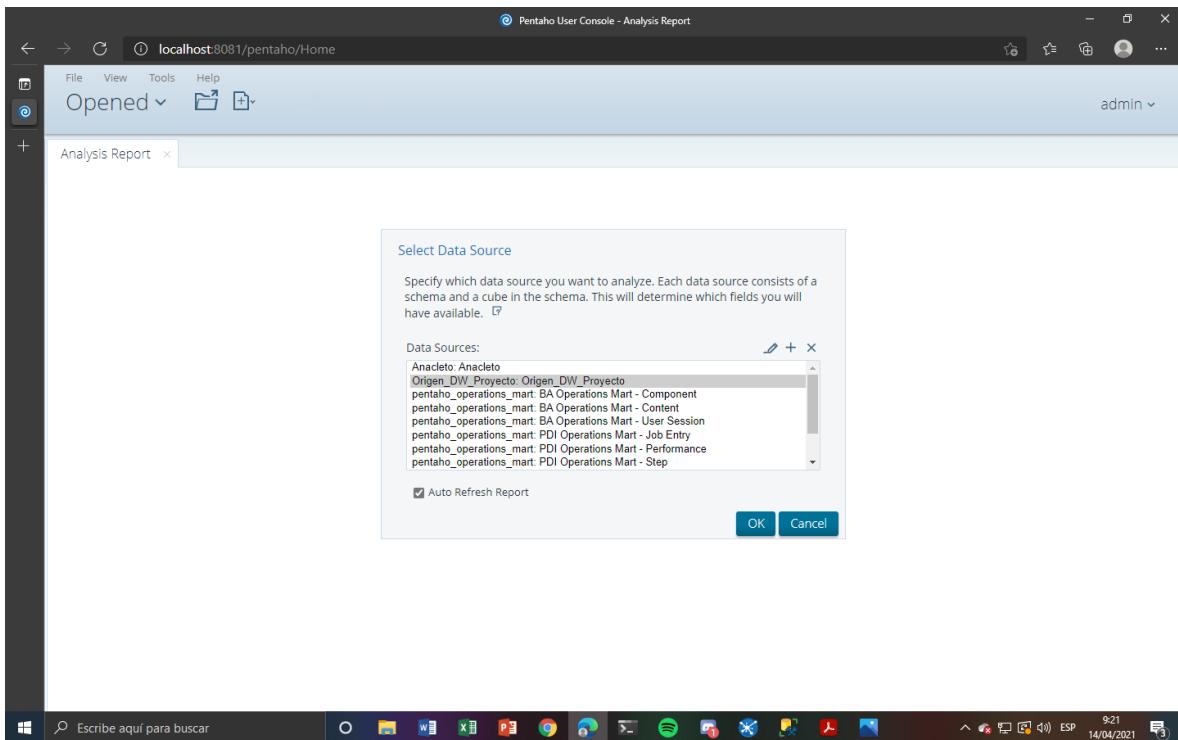
7.2 Data Source Model / Cubo OLAP

7.2.1 Creación

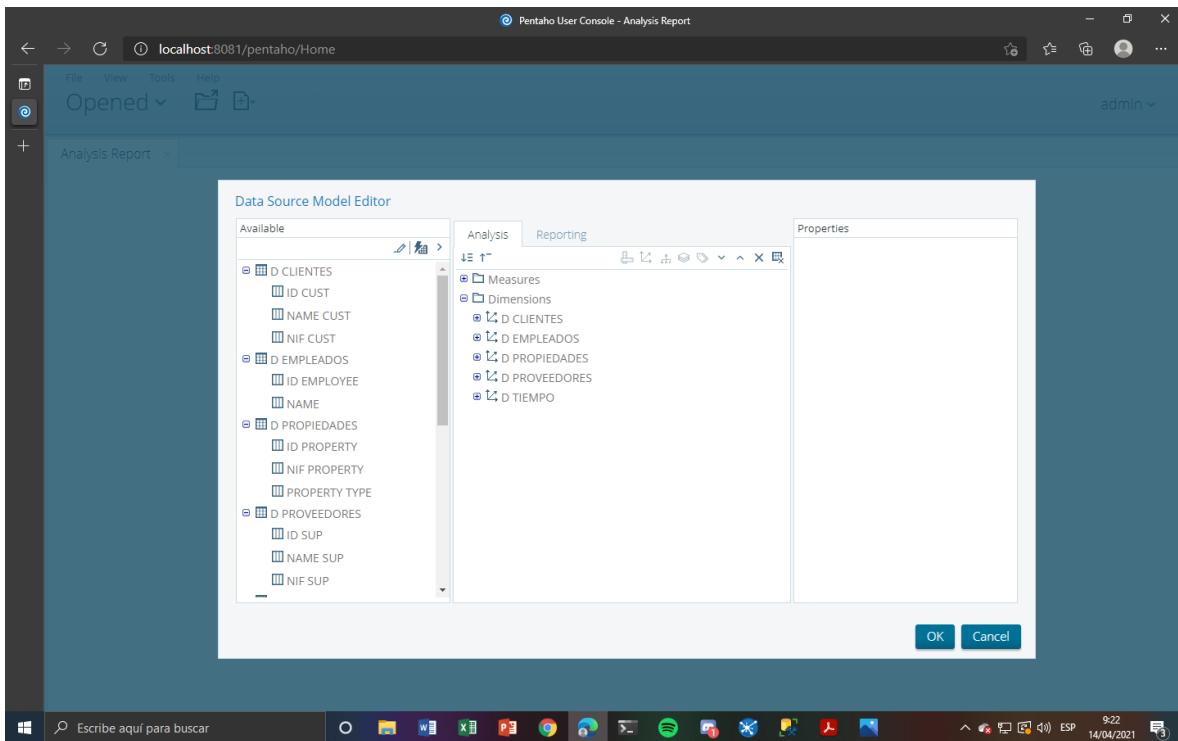




7.2.2 Selección

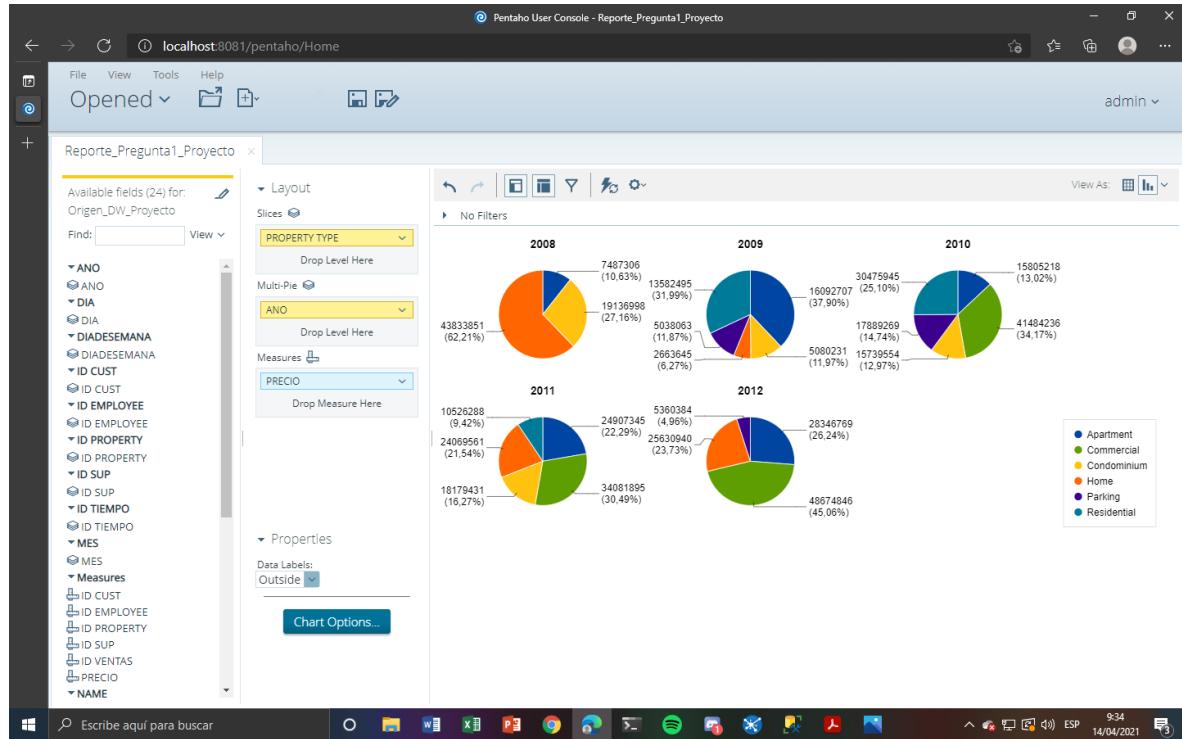


7.2.3 Edición

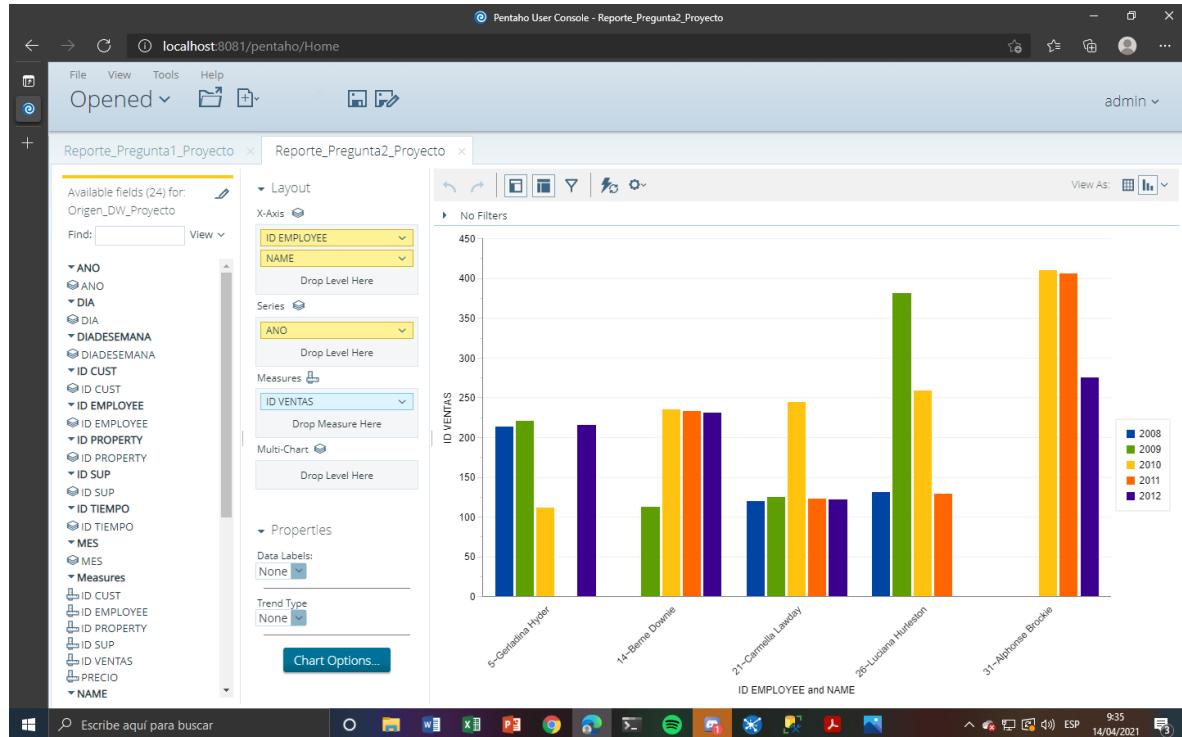


7.3 Generación de Reportes

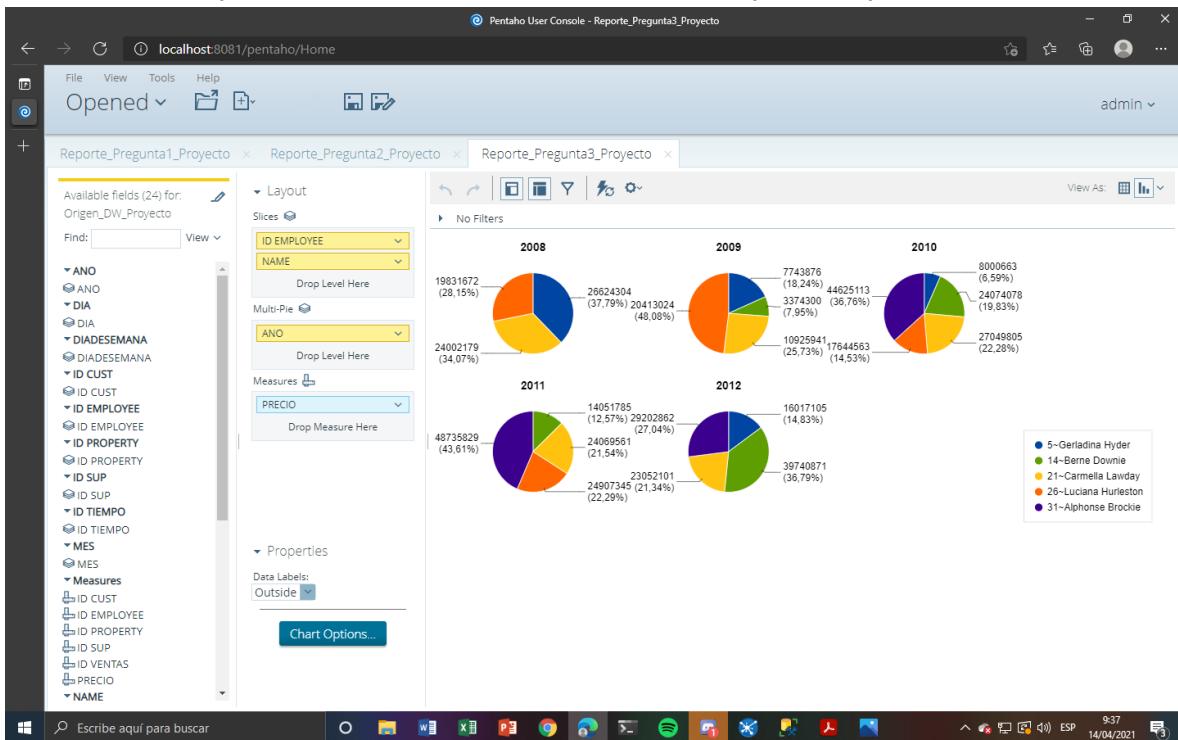
7.3.1 Reporte – Total de ventas en base a tipo de producto



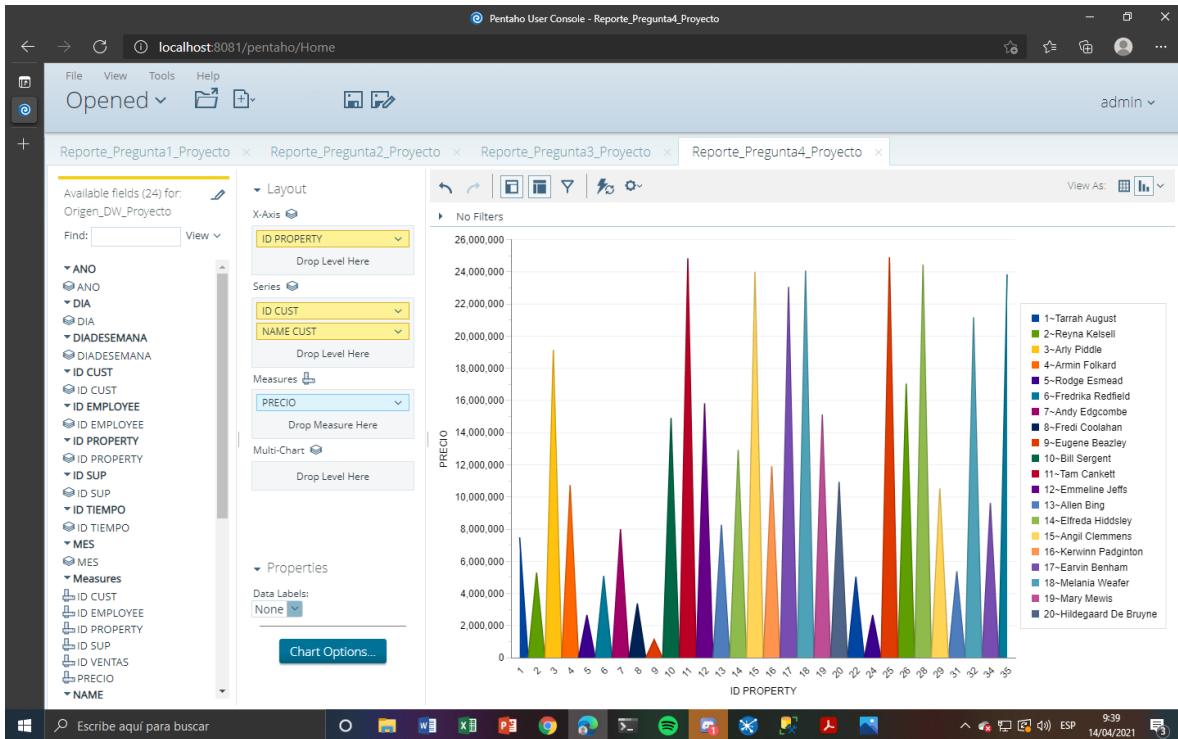
7.3.2 Reporte – Nombre y Código de empleado que realiza más ventas



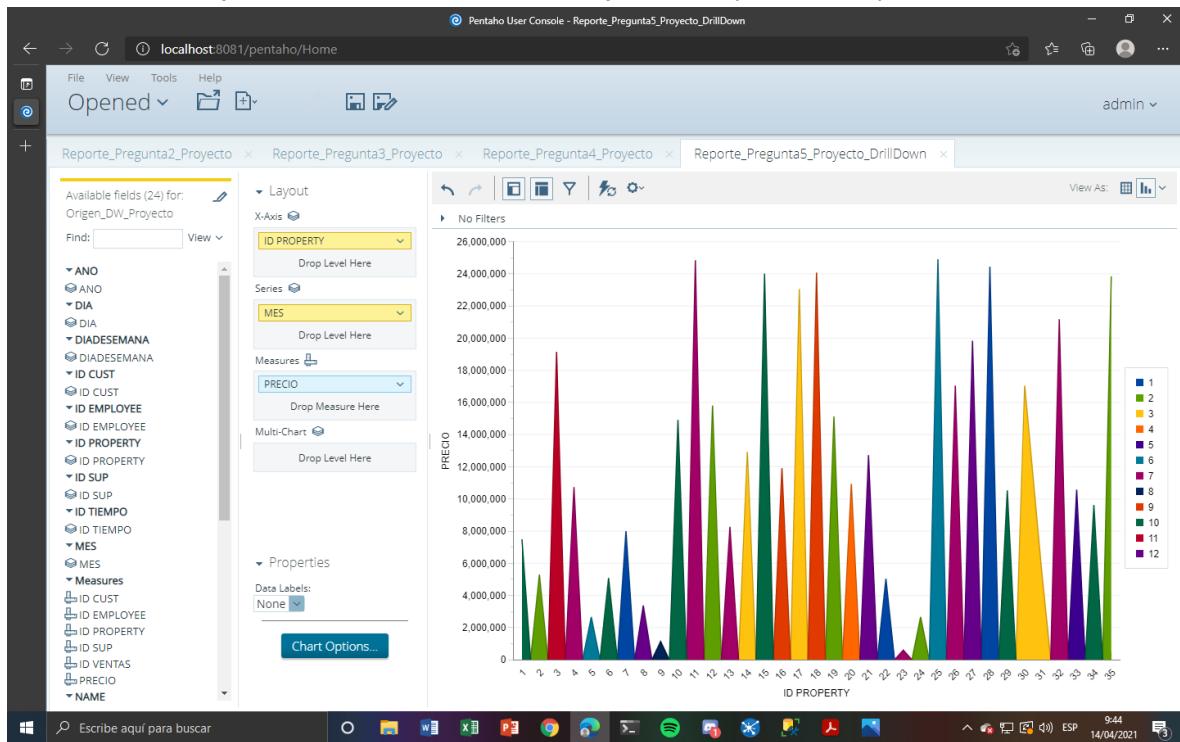
7.3.3 Reporte – Total de ventas en base al empleado que las realiza



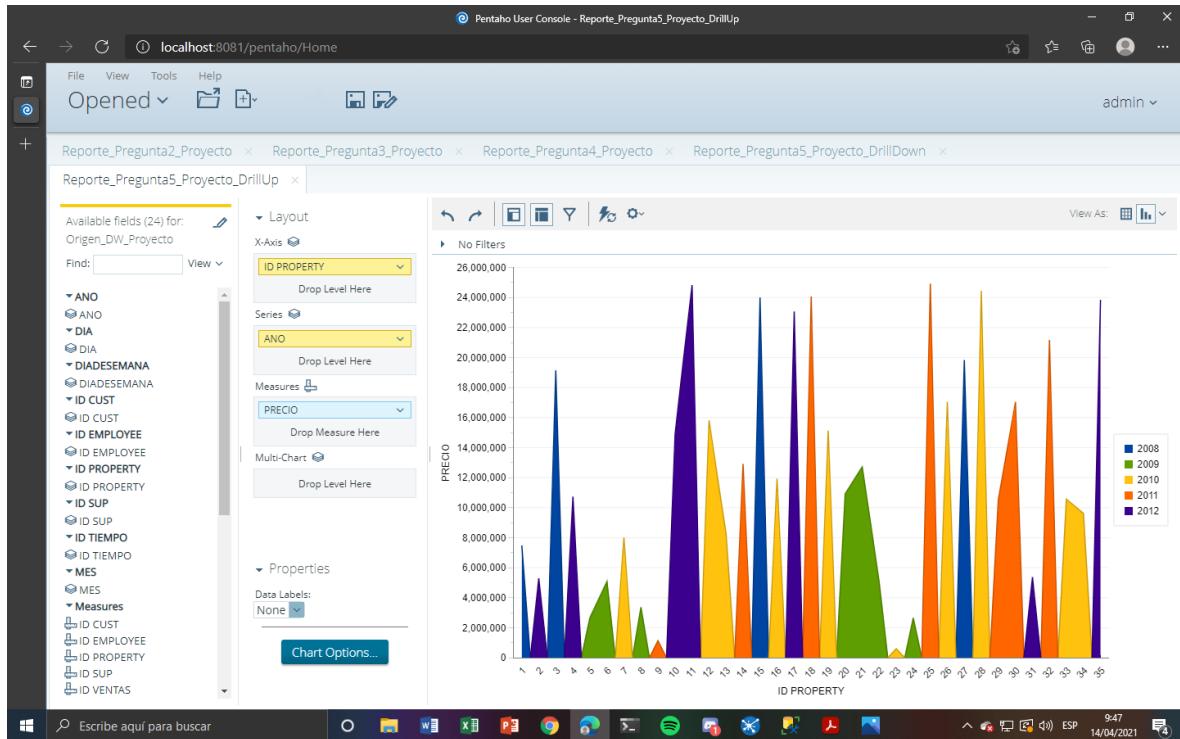
7.3.4 Reporte – Total de cada propiedad vendida en base al cliente



7.3.5 Reporte – Analizar las ventas por mes (DrillDown)



7.3.6 Reporte – Analizar las ventas por año(DrillUp)



7.4 Asignación de Reportes

7.4.1 Usuario Admin

The screenshot shows the Pentaho User Console interface for the 'admin' user. The left sidebar displays a tree view of folders: 'home' (containing 'admin', 'pat', 'suzy', 'tiffany'), 'Public', and 'Trash'. The 'admin' folder is selected and highlighted in green. The main area, titled 'Browse Files', shows a list of files under the 'Files' tab. The files listed are:

- FillAuditMartTime.html
- Reporte_Pregunta1_Proyecto
- Reporte_Pregunta2_Proyecto
- Reporte_Pregunta3_Proyecto
- Reporte_Pregunta4_Proyecto
- Reporte_Pregunta5_Proyecto_DrillDown
- Reporte_Pregunta5_Proyecto_DrillUp
- UpdateAuditData.html
- Ventas_empleado_mes

The 'Folder Actions' menu on the right includes options like 'New Folder...', 'Move to Trash', 'Rename...', 'Paste', 'Upload...', 'Download...', and 'Properties...'.

7.4.2 Usuario Suzy

The screenshot shows the Pentaho User Console interface for the 'suzy' user. The left sidebar displays a tree view of folders: 'home' (containing 'admin', 'pat', 'suzy', 'tiffany'), 'Public', and 'Trash'. The 'suzy' folder is selected and highlighted in green. The main area, titled 'Browse Files', shows a list of files under the 'Files' tab. The files listed are:

- Reporte_Pregunta2_Proyecto
- Reporte_Pregunta4_Proyecto
- tablero_ventas
- Ventas_producto_trimestre

The 'Folder Actions' menu on the right includes options like 'New Folder...', 'Move to Trash', 'Rename...', 'Paste', 'Upload...', 'Download...', and 'Properties...'.

8. Preguntas de Negocio Utilizadas

Tomando en cuenta las siguientes interrogantes sobre el negocio es que se crea el modelo estrella con el cual se hace referencia a las entidades y atributos que contiene la base de datos OLAP; esto para obtener un análisis más detallado que nos permite generar los reportes necesarios para la toma de decisiones:

- Analizar cuál es el total de ventas en cuanto a los tipos de propiedades.
- ¿Cuál es el código y nombre del empleado que realiza más ventas?
- Conocer el total de ventas en base al empleado que realiza la venta
- Conocer el total de cada propiedad vendida en base al cliente
- Analizar las ventas por año, mes, día, trimestre, semestre y día de la semana

9. Explicación de Métricas Utilizadas

- ✓ **Precio de las Propiedades (PRECIO):** Con esta métrica se pretende analizar, cuáles fueron las ventas totales ya sea por producto o tipo de producto. Mediante la herramienta para generar reportes en este caso Pentaho se nos permite visualizar cuáles fueron los totales de ventas con respecto a las fechas en que se realizaron.
- ✓ **Identificador de Ventas(ID_Ventas):** Nos permite acceder a toda la información acerca de la venta, como, por ejemplo, el cliente que compró el terreno, el proveedor que vendía el terreno, el empleado que realizó la gestión, la fecha en la que se realizó y la información existente acerca de la propiedad.

10. Conclusiones

Una vez realizado el proyecto y contando con una ejecución exitosa en cada uno de los procesos para el desarrollo del mismo, es posible concluir que se lograron los objetivos del proyecto, como ser el desarrollo de la base de datos OLAP a través de un modelo en estrella para la elaboración del datamart, al cual a través de la herramienta de Pentaho se le aplicaron las técnicas para la elaboración de ETLs que posteriormente nos permitieron generar una serie de reportes en base a la información que se necesitaba reconocer para poder ondear en los resultados que está presentando la inmobiliaria.

En general esto nos permite reconocer que los sistemas de inteligencia empresarial son de mucha ayuda para brindar acceso a la información necesaria a la hora de la toma de decisiones. Y así las empresas pueden conocer y manejar la información con la que cuentan de una forma más interactiva y fácil permitiéndole tener control de más áreas de las que podría sin contar con una herramienta de este tipo.

En la actualidad al contar con múltiples empresas que desean introducirse al mundo de la digitalización de la información el contar con este tipo de softwares que bien serían de terceros o con uno creado específicamente para las necesidades que esta debe solventarle sería de infinita ayuda puesto que no solo es fácil de utilizar, sino que también es muy preciso a la hora de brindar estadísticas y cuenta con entornos de desarrollo básicos, en todas las herramientas de la aplicación.

11. Recomendaciones

- Crear o utilizar una base de datos OLTP con la suficiente información para poder generar los reportes necesarios.
- Instalar los drivers necesarios para poder utilizar Pentaho.
- Tener habilitado el usuario SA en el sistema gestor de bases de datos.
- Verificar que el puerto dedicado a escuchar las peticiones que se deben hacer esté habilitado.
- Verificar que el proveedor de servicios de la base de datos tenga habilitado el protocolo TCO/IP.
- Recordar que se deben hacer siempre las conexiones con las bases de datos OLAP y OLTP cuando se van a hacer las transformaciones con las bases de datos.

12. Bibliografía

- Acosta, E. (24 de febrero de 2021). Clase 1 - Introducción al DW
<https://drive.google.com/file/d/1lr5buzFm7IEx9U4Aq3jDDenRgo7Q6-uE/view?usp=sharing>
- Acosta, E. (24 de febrero de 2021). Clase 2 - Construcción del data mart
https://drive.google.com/file/d/1U4LKXiPQ6SS1cTUBBiH5BP_Z5zT9UAYN/view?usp=sharing
- Acosta, E. (24 de febrero de 2021). Clase 3 - Creación de ETL (parte A)
https://drive.google.com/file/d/1ei5MLWJrphV_qPKbetYVS3R6gWIQQiX1/view?usp=sharing
- Acosta, E. (24 de febrero de 2021). Clase 3 - Creación de ETL (parte B)
<https://drive.google.com/file/d/1DlxuOPnBokGwJIO80dnylaWukyXkRflz/view?usp=sharing>
- Acosta, E. (24 de febrero de 2021). Clase 3 - Creación de ETL (parte C)
<https://drive.google.com/file/d/1fvap-ZnDCOcM70lbfU8XvqJQC2drk8b6/view?usp=sharing>
- Acosta, E. (24 de febrero de 2021). Clase 4 - Creación de reportes en Pentaho (parte A)
<https://drive.google.com/file/d/1s78GOdGLsgX32OIPoV-Loa9gaqkMxhs/view?usp=sharing>
- Acosta, E. (24 de febrero de 2021). Clase 4 - Creación de reportes en Pentaho (parteB)
<https://drive.google.com/file/d/1DMVgoIGy8uu3C6mhf1vvn3fkNc2tbN9/view?usp=sharing>
- ¿Qué es un Data Warehouse?. (2021). Recuperado 13 Abril 2021, de <https://www.ibm.com/mx-es/analytics/data-warehouse>
- Sistemas OLTP: procesamiento, administración y mantenimiento de transacciones - Evaluando Software. (2021). Recuperado 13 Abril 2021, de <https://www.evaluandosoftware.com/sistemas-oltp-procesamiento-administr>

acion-mantenimiento-transacciones/#:~:text=Diferencia%20entre%20sistemas%20OLTP%20y,proceso%20de%20soporte%20de%20decisiones.

- ¿Qué es Business Intelligence?. (2021). Recuperado 13 Abril 2021, de https://www.sinnexus.com/business_intelligence/