

ANEXO C – Documento de Extracción

1. Objetivo

El objetivo del documento de extracción es presentar aquellas características que se mencionan de manera general en el documento de tesis. Aquí se enfoca directamente la extracción, transformación y carga de cada componente de los datamarts que se ejecutarán dentro del proyecto.

2. Alcance

Se presentará el diseño dimensional final y el proceso de construcción dentro del proceso de ETL ejecutado.

El asistente para el diseño de cubos basado en la herramienta SQL Integration Services 2008 (SSIS) ayudará a la creación de los datamarts facilitando la definición de las dimensiones, los niveles de jerarquía, la tabla de hechos y la conexión a la base de datos relacional en un fichero XML. Dichos paquetes se accederán a las fuentes descritas en el punto anterior y, dependiendo de la tarea, transformarán, cargarán o harán ambas cosas.

3. Arquitectura de la Extracción

La arquitectura de esta solución constituye el diseño de una aplicación de alto nivel que se divide en los distintos orígenes de datos existentes y las capas donde extraerán, transformarán y cargarán dicha data.

A. Origen

Corresponde a las múltiples fuentes (hojas de texto, sistemas transaccionales, base de datos externa), las fuentes provienen de las siguientes necesidades:

Base de datos del sistema Core

Los datos son extraídos desde base de datos relacionales OLTP ubicados en el motor de base de datos SQL Server 2008 que son captados por sistemas transaccionales como SAP Business One 8.8; este sistema principal de la organización contiene la base de datos operacional fuente para la

construcción de la base de datos OLAP. La base de datos se muestra en la Figura 1.

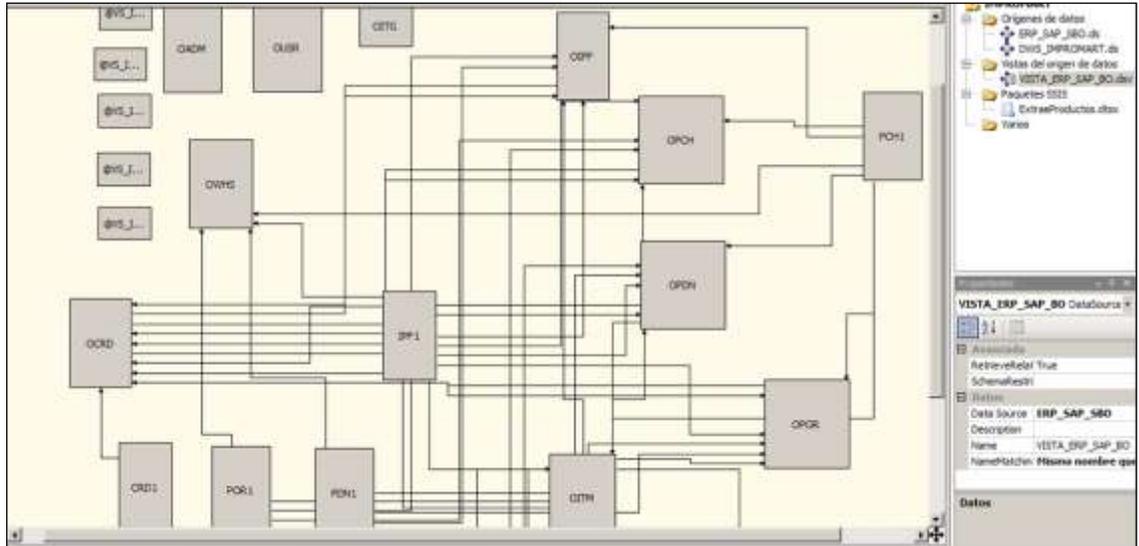


Figura 1 - Diagrama de Base de Datos de ERP fuente

Archivo de almacenaje y control de gastos en almacén intermedio

Estos archivos son provistos por el operador logístico encargado de los procesos de almacenaje intermedio de productos. Contiene información de los movimientos, estado y saldo de los productos dentro de sus almacenes a una fecha determinada.

Archivo de control de envío en aduanas

Estos archivos son provistos por la agencia de aduanas encargado de la entrada y salida de productos a territorio nacional. Contiene información del estado y los montos o gastos asignados a los productos que entran a nuestro país.

B. Capas

Capa Operacional

Esta capa servirá para la integración de los datos, es la que seguirá el modelo de almacenamiento MOLAP, ya que tendrá la mayor atomicidad en detalle de la información. Asimismo, guardará información histórica y será la

que se actualice cada cierto tiempo por cada carga del proceso ETL extraído de la bases de datos fuente.

Capa de Almacenamiento Temporal Operacional

Esta capa servirá de almacenamiento intermedio entre la base de datos fuente y la base de análisis operacional.

Teniendo presente que las soluciones se encontrarán en un servidor distinto al de la base de datos del sistema principal, será necesario tener un stage o almacén de datos donde se consolide la información de las distintas fuentes. Aquí se realiza la transformación de datos fuente como cambios de tipo de dato, truncamiento de datos, cambios de formato para tener nuevos datos fuente sumariados para la capa de explotación; para estandarizar y realizar la limpieza de la base de datos OLAP Y permitir el mapeo adecuado.

Capa de Explotación

Esta capa servirá para almacenar información resumida y agregada de lo que existe en la capa de datos operacional, de acuerdo a las necesidades del negocio.

Primero la información se importa de la capa operacional mediante herramientas ODBC, luego se crea el diseño físico, lógico y de presentación para el usuario final.

La naturaleza de la estructura que se tiene en este ambiente sirve para procesar consultas con un performance mayor al de un modelo relacional. En tal sentido, permitirá una explotación analítica.

4. Proceso de Extracción, Transformación y Carga

Este punto se encargará de describir los procesos de extracción, transformación y carga que se necesitarán para la población del datamart. Las tablas a usarse tendrán las fuentes presentadas en la Figura 1.

Carga de Tablas de Almacenamiento Temporal Operacional:

Para la realización de la carga de esta capa, es necesario extraer los datos de las fuentes, estos serán los que realicen la carga a una fecha de corte

determinada. Este proceso puede basarse usando la herramienta jobs de SQL Server 2008 o basado en alertas de Windows.

La Figura 2 presenta la extracción de los distintos orígenes de datos hacia la capa intermedia.

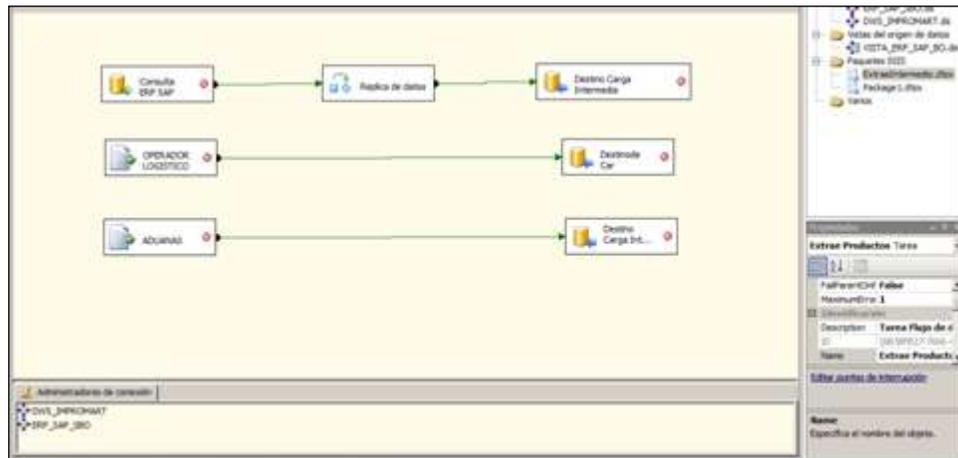


Figura 2 - Diagrama de extracción de archivos fuente

Transformación y Carga de Datos hacia base operacional

Una vez que se extraen los datos de las diversas fuentes lo que hace este paquete es ejecutar la réplica hacia la capa de Operación, siempre y cuando todos los archivos contengan registros nuevos. A continuación, en la Figura 7 se presenta consultas para realizar las extracciones y transformaciones estándar de la herramienta.



Administrador de conexiones de ADO.NET:

ERP_SAP_SBO Nueva...

Modo de acceso a datos:
Comando SQL

Texto de comando SQL:

```
select P.ItemCode AS CodigoProducto ,P.ItemName AS NombreProducto,  
B.ItmsGrpNam AS NombreGrupo,P.LastPurPrc AS UltimoPrecio,  
P.OnOrder AS Solicitado,P.Onhand AS EnStock,P.IsCommitted AS Comprometido  
from  
OITM P  
INNER JOIN OITB B ON B.ItmsGrpCod=P.ItmsGrpCod  
where  
P.validFor='N'
```

Generar consulta...
Examinar...
Vista previa...

Figura 3 - Consulta de extracción

Para cargar dicha información en la base de datos de intermedia, se tiene que realizar paquetes para la creación de tablas y que este la plataforma conforme para el ingreso de datos. En la presente Figura se realiza la extracción, la transformación de la tabla Producto.

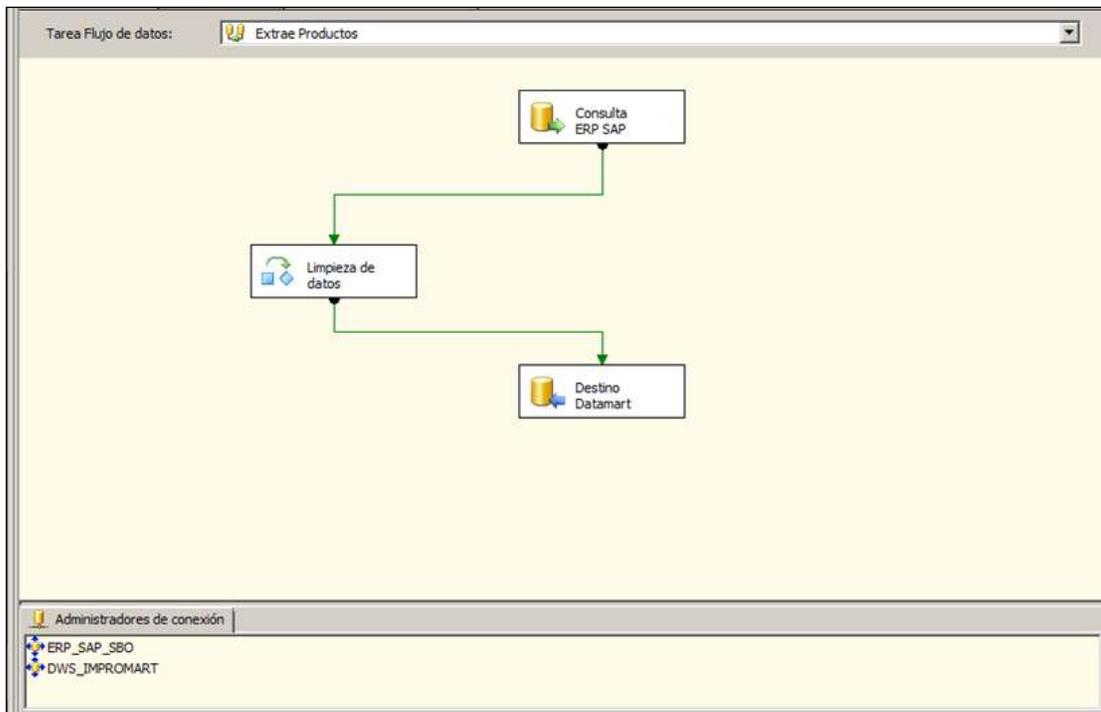


Figura 4 - Diagrama de extracción de tabla Producto

En el presente anexo se detallan las tablas fuente de extracción para todas las dimensiones y tabla de hechos involucradas en la solución.

La extracción de las tablas origen dentro de la ETL para la tabla de Dimensiones de Productos son las siguientes:

- OITM –Tabla que representa información de Productos
- OITB -Tabla que representa información de la familia de productos.

La extracción de las tablas origen dentro de la ETL para la tabla de dimensiones de Proveedores son las siguientes:

- OCRD –Tabla que representa la información de Socios de Negocios tanto Clientes como Proveedores.
- OSLP – Tabla que representa la información de Empleados de Compra.
- OSTG – Tabla que representa los grupos que representan dichos Socios o Proveedores.
- CRD1 – Tabla que representa las direcciones del Socio o Proveedor.

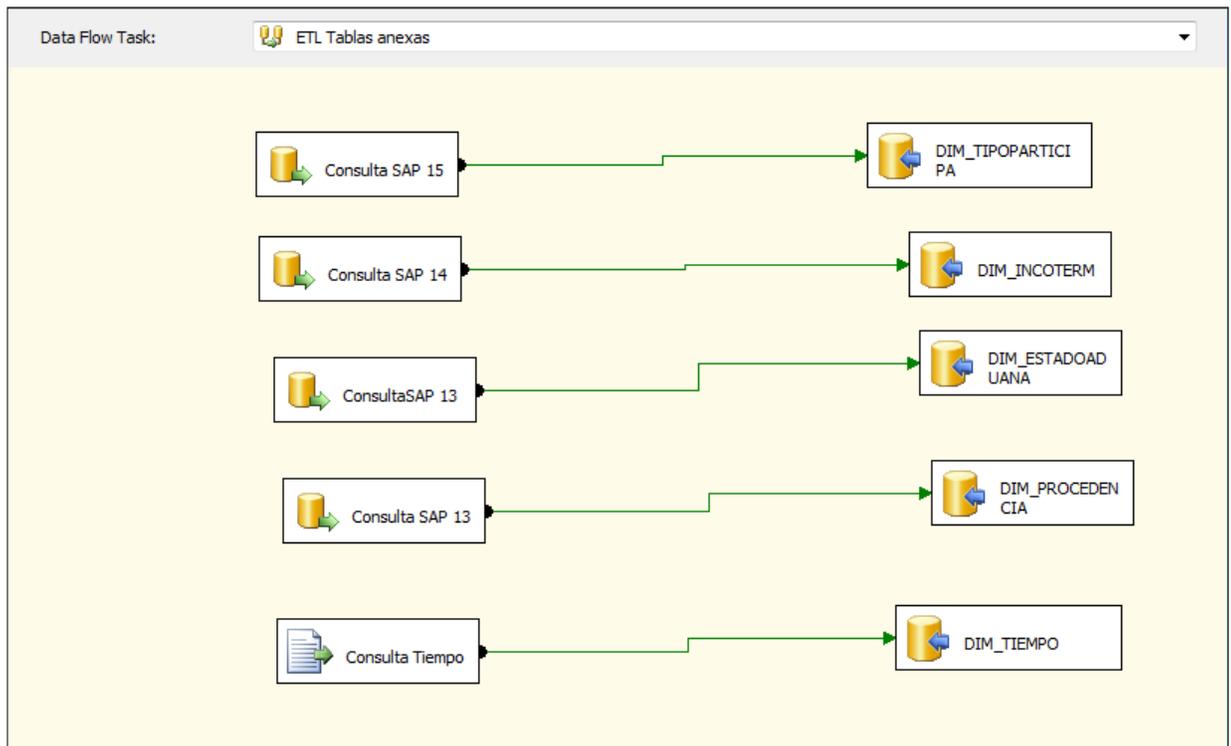


Figura 5 - Diagrama de mapeo de tablas Anexas

La extracción de las tablas origen dentro de la ETL para la tabla de Dimensiones de Procedencia son las siguientes:

- OCRD –Tabla que representa la información de Socios de Negocios tanto Clientes como Proveedores
- CRD1 – Tabla que representa las direcciones del Socio o Proveedor.
- OCST - Tabla de códigos de Países.
- OCRY – Tabla de códigos de Departamentos.

Tiene como objetivo tener toda la base de datos de las direcciones de almacén de los proveedores internacionales para referenciar todos los orígenes posibles.

La extracción de las tablas origen dentro de la ETL para la tabla de Dimensiones son las siguientes:

- OALC –Tabla que representa la información de la condición o tipo de gasto representado en la importación.

Tiene como objetivo tener toda la base de datos de las condiciones para el envío para los operadores logísticos para referenciar todos los orígenes posibles.

La extracción de las tablas origen dentro de la ETL para la tabla de Dimensiones de Estado_Aduana son las siguientes:

- @TE_ESTADOADUANA–Tabla que representa la información del estado de la importación dentro de la oficina de aduana.

Tiene como objetivo tener toda la base de datos de los estados para el envío dentro de la oficina de aduanas para referenciar todos los orígenes posibles.

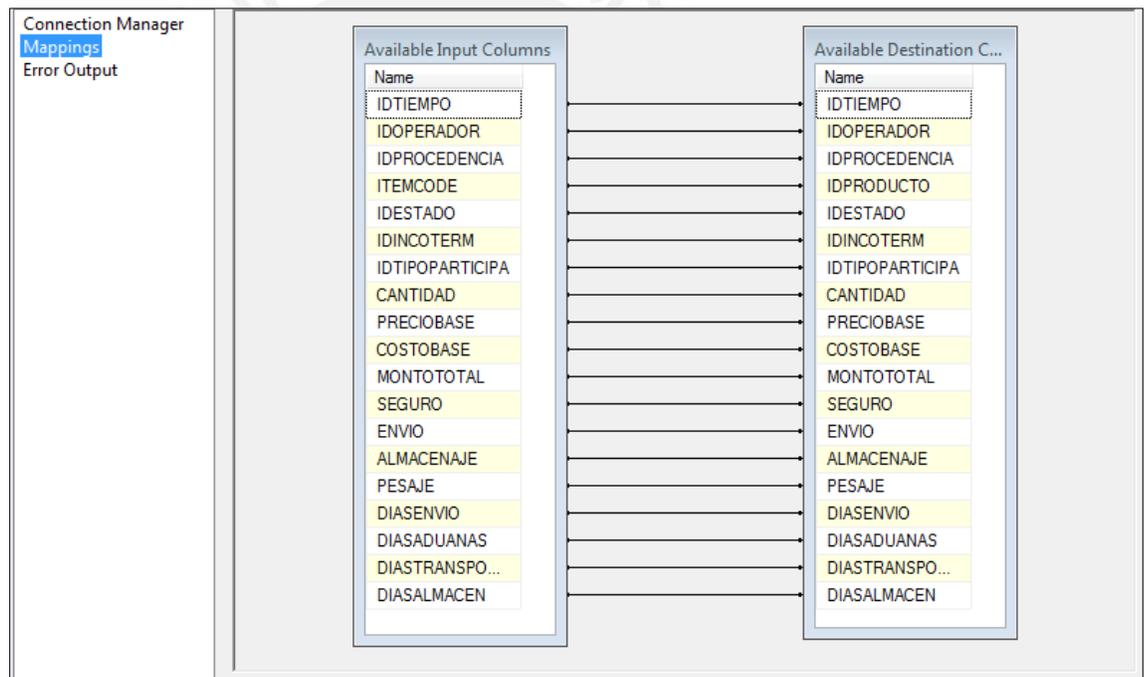


Figura 6 – Mapeo de Tabla de Hechos de Importación

La extracción de las tablas origen dentro de la ETL para la tabla de Dimensiones de Incoterm son las siguientes:

- @TE_INCOTERMINOS–Tabla que representa la información de los términos internacionales de comercio

Tiene como objetivo tener toda la base de datos de los términos o condiciones internacionales para el comercio exterior referenciando todos los orígenes posibles.

La extracción de las tablas origen dentro de la ETL para la tabla de Dimensiones de Tipo de participación son las siguientes:

- @TE_TIPOPARTICIPA –Tabla que representa la información de la condición o tipo de gasto representado en la importación.

Tiene como objetivo tener toda la base de datos de las condiciones para el envío para los operadores logísticos para referenciar todos los orígenes posibles.

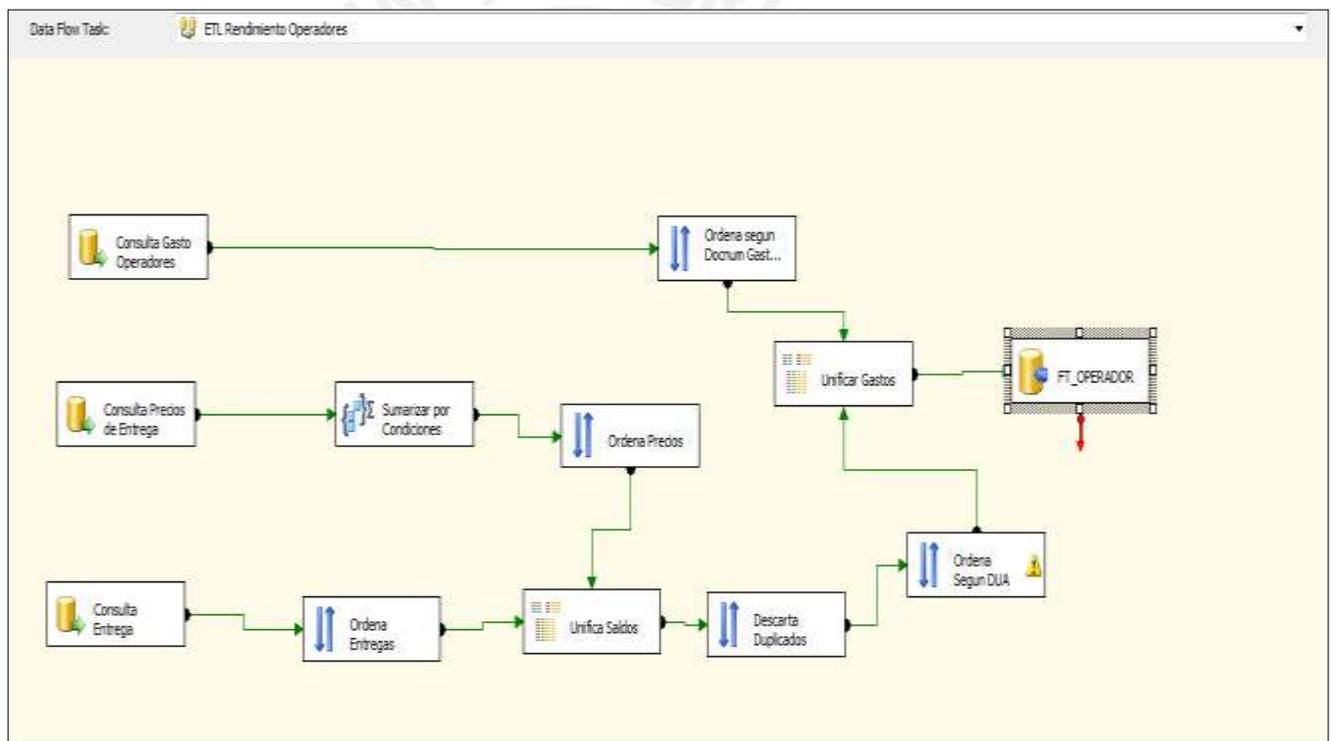


Figura 7 – Diagrama de extracción de tabla de Hechos Rendimiento Operadores

La extracción de las tablas origen dentro de la ETL para la tabla de Hechos FT_Importacion y el FT_OperadorLogistico son las siguientes:

- OIPF–Tabla que representa la información de la cabecera que representan gastos en una importación.

- IPF1 - Tabla que representa la información del detalle de los productos que representan gastos en la importación
- IPF2 - Porcentaje de gasto en las importaciones sobre cada condición sea flete, almacenaje, documentación, aduanaje.
- OPDN – Tabla que representa la cabecera de las guías de entrega.
- PDN1 – Tabla que representa el detalle de las guías de entrega.
- PDN12 - Tabla que representa las direcciones de envío origen de las importaciones.
- OPCH – Tabla que representa las facturas de gastos de importación.
- OITM –Tabla que representa información de Productos
- OITB -Tabla que representa información de la familia de productos.

Tiene como objetivo tener toda la base de datos de las combinaciones dentro de las compras de importación para referenciar todos los orígenes posibles.

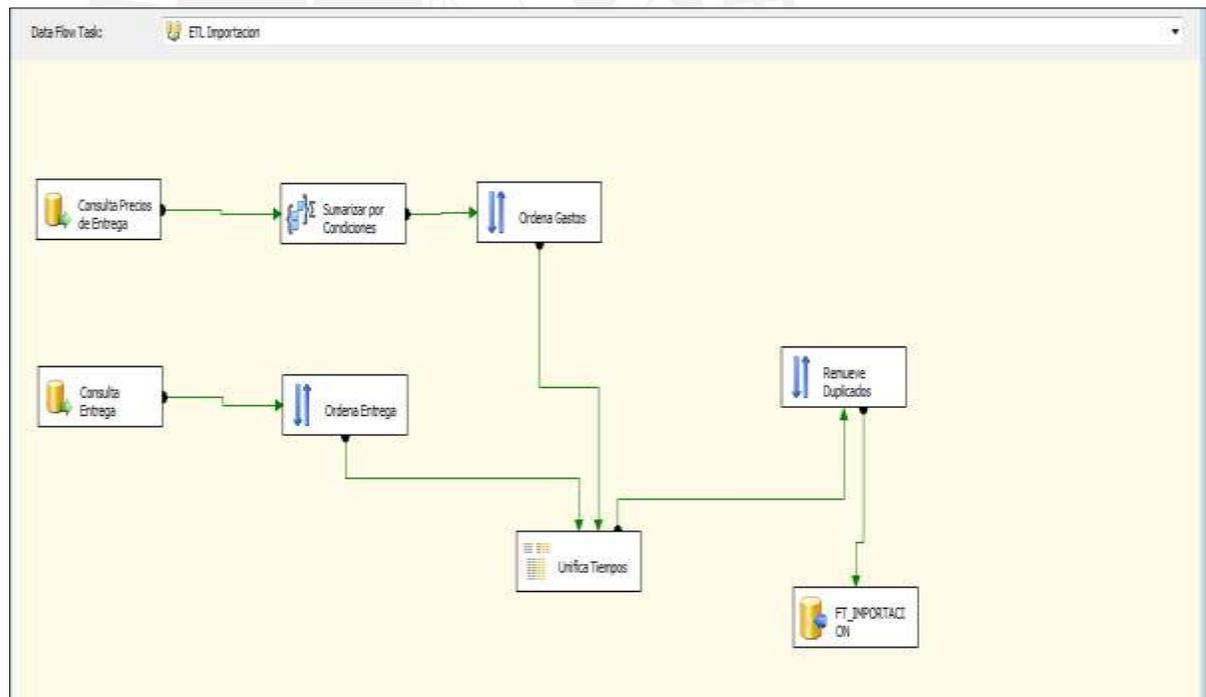


Figura 8 – Diagrama de extracción de tabla de Hechos Importación