



UNAH

UNIVERSIDAD NACIONAL
AUTÓNOMA DE HONDURAS

DEPARTAMENTO DE
INGENIERÍA EN SISTEMAS

IS-601 BASE DE DATOS II

Proyecto de clase

Ing. Emilson Acosta



UNAH

UNIVERSIDAD NACIONAL
AUTÓNOMA DE HONDURAS

IS-601 BASES DE DATOS II

Primer Periodo 2019

21 de abril
Proyecto Final

Elaborado por:

20151020842	Lester Alexander Flores
20151000308	Oscar E. Blandon
20151003416	German Rene Tejada
20151005232	Génesis Raquel Izaguirre
20141010053	Angie Pamela Argueta

Ing. Emilson Acosta



Contenido

DESCRIPCIÓN TÉCNICA.....	4
DataWarehouse	4
Northwind.....	4
Business Intelligence	4
REQUERIMIENTOS DE NEGOCIO.....	5
Base de datos OLTP	6
Base de datos OLAP	7
Sistema de Soporte a la Toma de Decisiones DSS.....	8
Propuesta para ejecutar proceso ETL más de una vez	15
CONCLUSIONES.....	23
RECOMENDACIONES	23
BIBLIOGRAFÍA	24



INTRODUCCIÓN

Hoy en día toda empresa necesita depositar mucha confianza en la toma de decisiones sobre los negocios por lo que se requiere de hechos y cifras que permitan una pronta respuesta a las demandas de quienes toman las decisiones en una empresa y estas deberán ser más acertadas.

En este informe se detallan los conceptos, así como la construcción de un Datawarehouse enfocado en una base de datos llamada Northwind. Un Datawarehouse es un repositorio unificado para todos los datos que se recogen de los diversos sistemas de una empresa. Haciendo uso de una base de datos OLTP y mediante ETL se cargarán los datos en una base de datos OLAP, para generar reportes se hará uso de una herramienta para el apoyo en la toma de decisiones DSS, además, se incluye la configuración de replicación en el motor de bases de datos SQL Server Enterprise para asegurar la alta disponibilidad de la información, todos los procesos descritos se abordarán a detalle a continuación.



DESCRIPCIÓN TÉCNICA

DataWarehouse

Una Datawarehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una empresa. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso.

Normalmente, un Datawarehouse se aloja en un servidor corporativo o cada vez más, en la nube. Los datos de diferentes aplicaciones de procesamiento de transacciones Online (OLTP) y otras fuentes se extraen selectivamente para su uso por aplicaciones analíticas y de consultas por usuarios.

Northwind

Es una base de datos proporcionada por Microsoft como ejemplo y puede ser descargada en la web oficial de Microsoft.

Su ventaja es que es una base de datos sencilla, pero al mismo tiempo contiene gran cantidad de información, relaciones, etc.... y simula las necesidades básicas de una empresa sencilla de importación de productos alimentarios. Contiene tablas para gestionar la información de productos, clientes, proveedores, personal, pedidos, transportistas, y más.

DataMart

Un data mart es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones. Los datos existentes en este contexto pueden ser agrupados, explorados y propagados de múltiples formas para que diversos grupos de usuarios realicen la explotación de los mismos de la forma más conveniente según sus necesidades.

Business Intelligence

Un Business Intelligence (BI) es una especie de “cuello de botella” de los datos recogidos del Datawarehouse, que llegan de forma exacta y útil para ayudar a la toma de decisiones. Business Intelligence transforma los datos en información útil para analizar no sólo los negocios, sino también las principales estrategias corporativas.



REQUERIMIENTOS DE NEGOCIO

Ya que la finalidad principal del Data Warehouse es almacenar la información ya depurada para su posterior empleo a nivel empresarial, es necesario conocer en que ámbitos se necesita la información para que facilite la toma de decisiones dentro de una compañía, así como también mejora la calidad de estas.

Esto se logra a través de las preguntas de negocio que pueden ayudar a ver claramente los requerimientos y cerrar un poco la brecha entre desarrollador-negocios.

Las preguntas que se plantearon para posteriormente basar nuestro modelo son las siguientes:

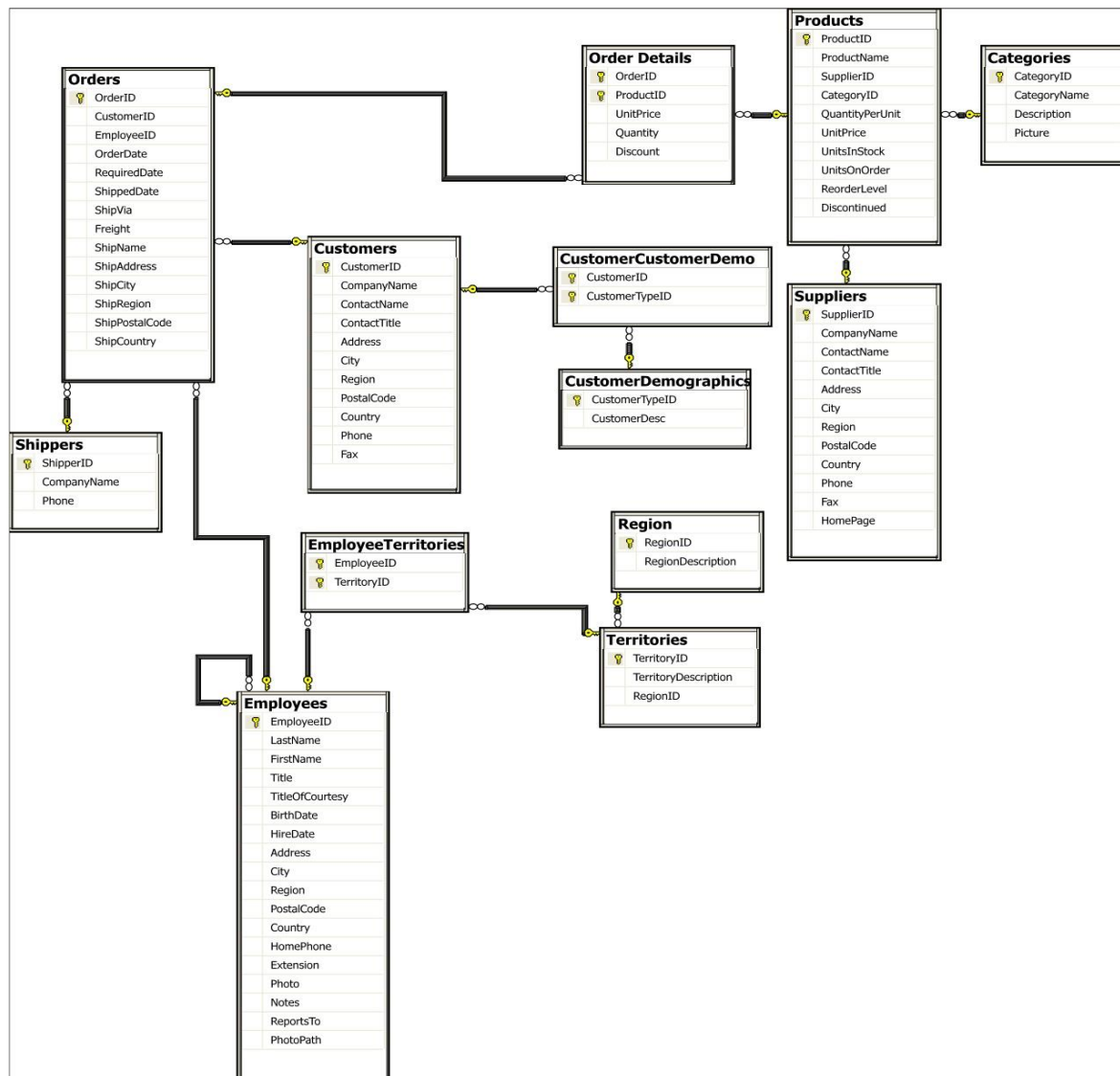
- ¿Qué empleado tiene un mayor número de ventas?
- ¿Qué empleado tiene un menor número de ventas?
- ¿Cuál es el producto que más se vende?
- ¿Cuál es la región o lugar que presenta mayor número de ventas?
- ¿Cuál es el producto que se vende más por lugar?
- ¿Quiénes son los clientes que más compran?
- ¿Cuál es el número o tendencia de compradores o ventas por fecha?

Teniendo consolidadas estas preguntas se procedió a determinar las dimensiones para el modelo; en sí el diseño lógico, siendo finalmente las siguientes:

- Dimensión Productos.
- Dimensión Tiempo.
- Dimensión Clientes.
- Dimensión Empleados.
- Dimensión Lugares.

Base de datos OLTP

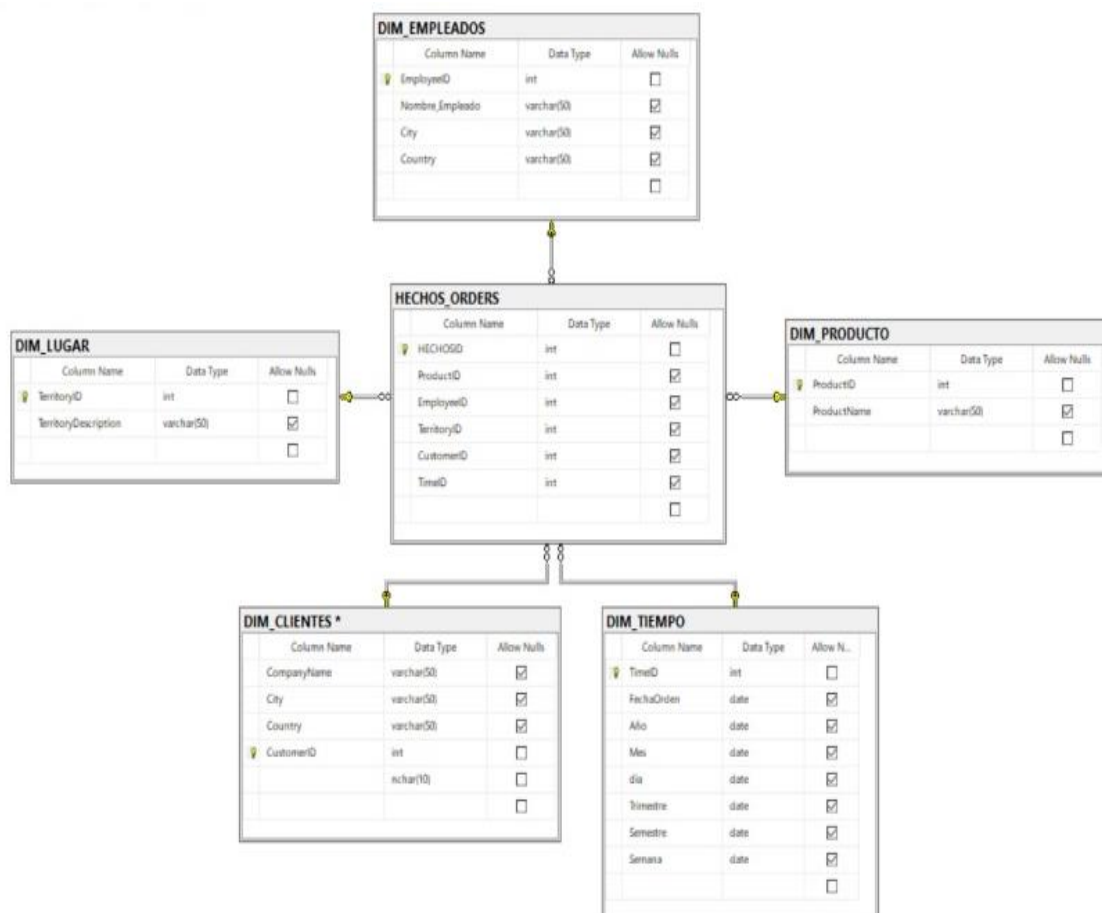
Los sistemas OLTP (On-Line Transactional Processing) son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico (que debe ser validado con un *commit*, o invalidado con un *rollback*), y que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales. A continuación, una pequeña parte del modelo OLTP del proyecto.



Northwind OLTP

Base de datos OLAP

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 datamars.



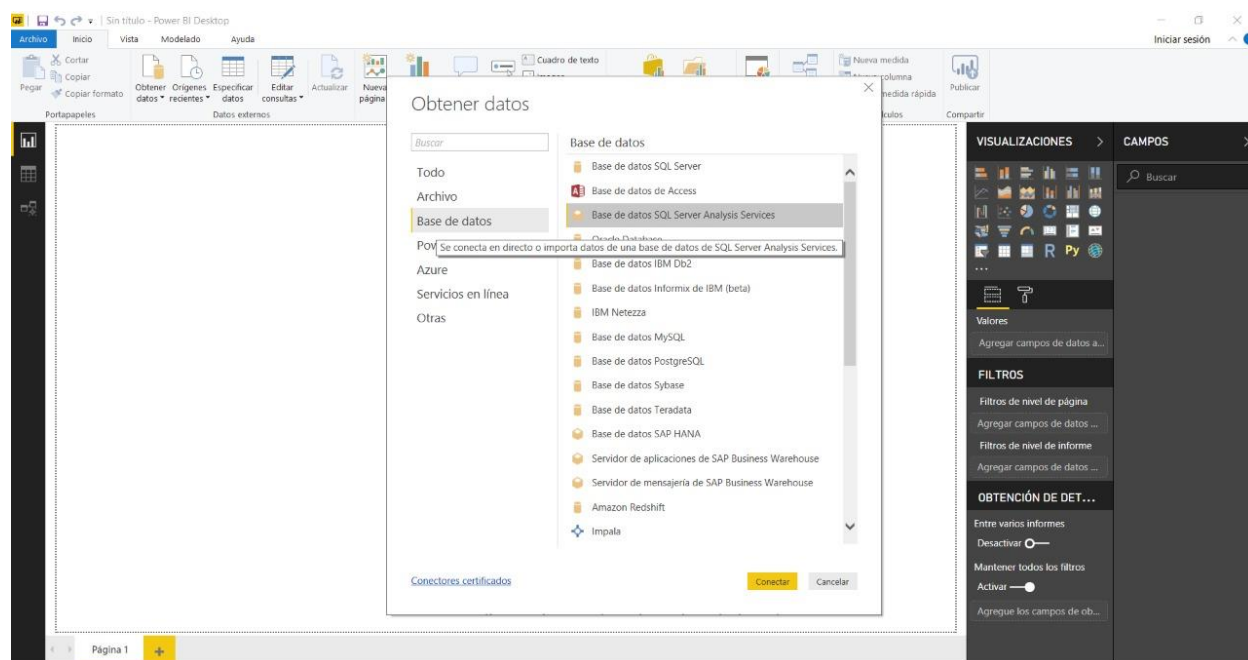
NorthWind DW

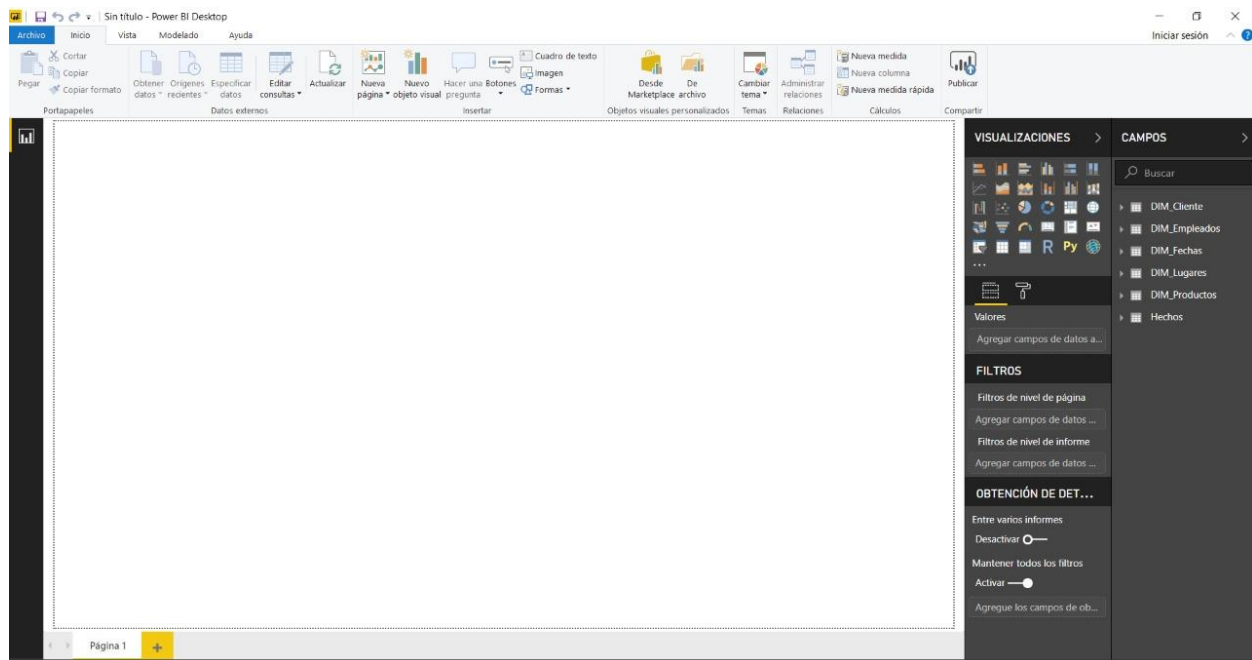
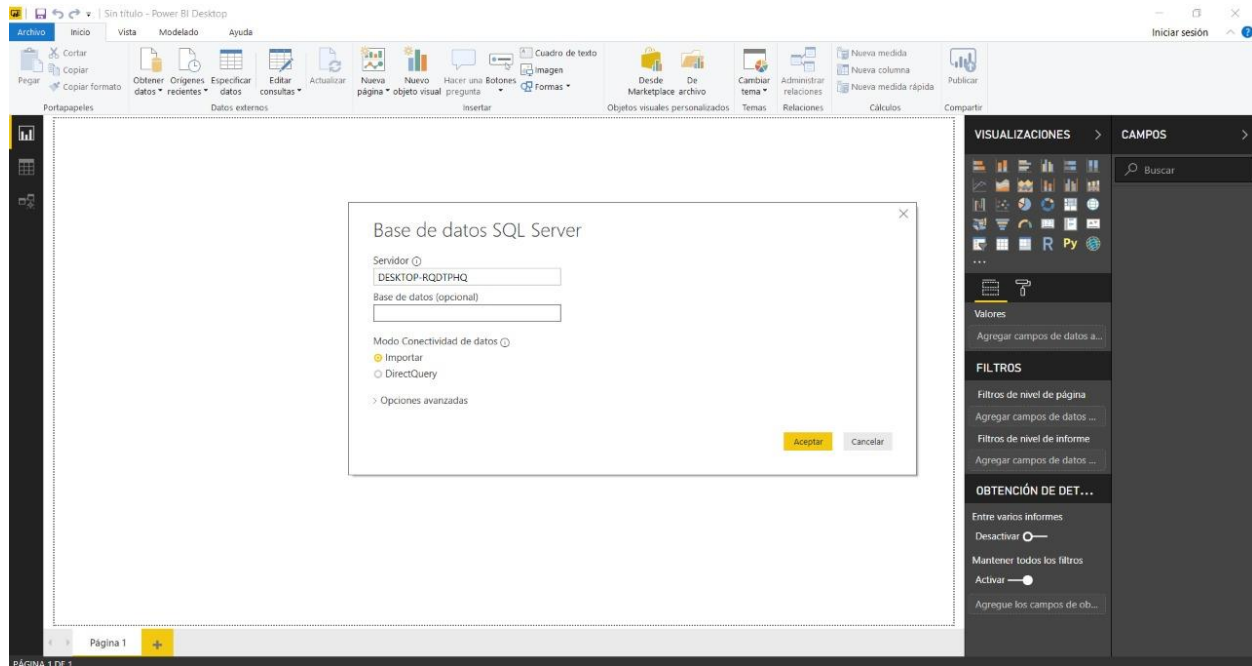
Sistema de Soporte a la Toma de Decisiones DSS

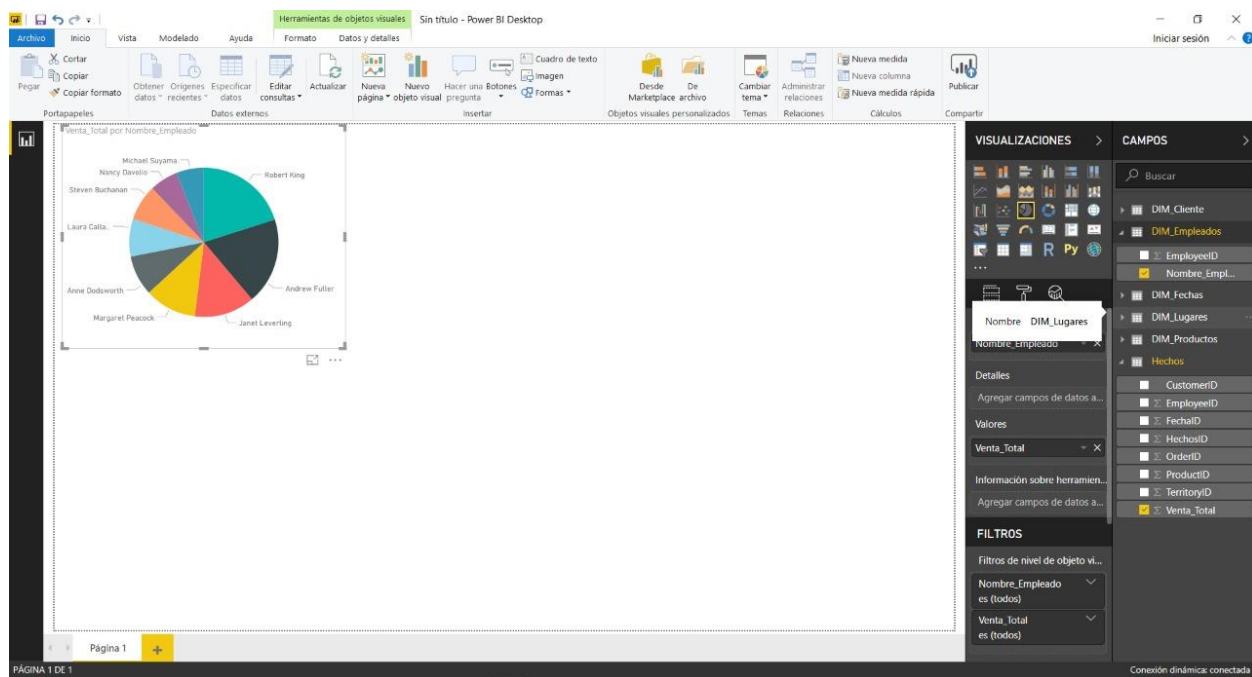
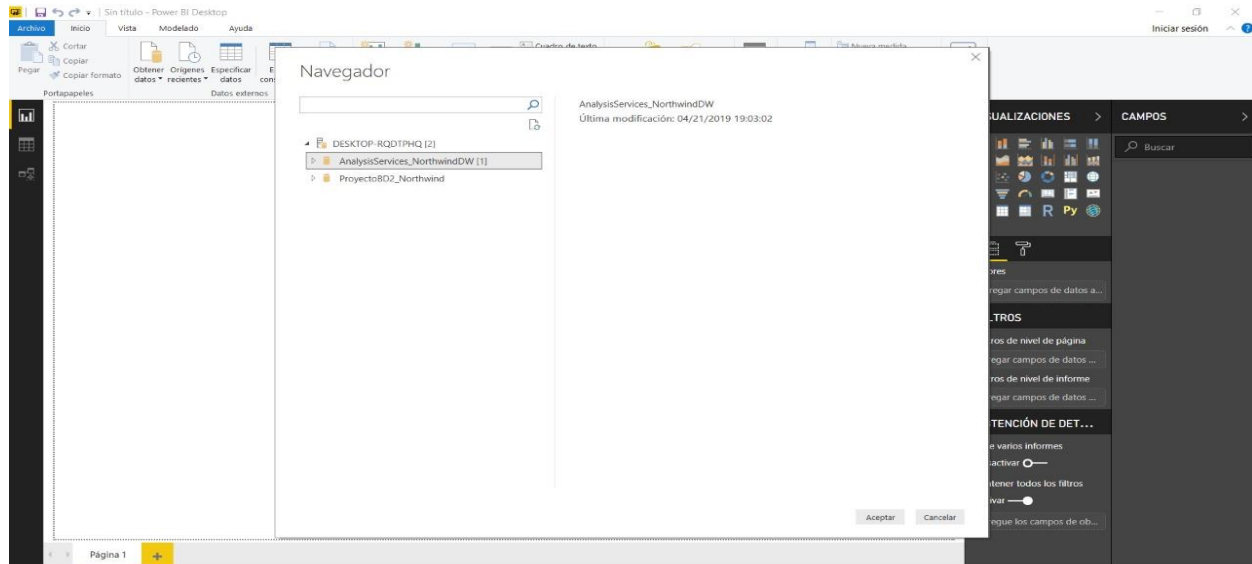
Para apoyar la solución de un problema de gestión no estructurado, mejorar la toma de decisiones, utilizar los datos anteriormente recolectados y responder a las preguntas de negocio; de forma que el usuario final pueda interpretar de manera correcta los datos, y proporcionar una interfaz amigable y permitir la toma de decisiones en el propio análisis de la situación, se decidió utilizar Power BI.

Power BI es una solución de análisis empresarial que permite visualizar los datos y compartir información con toda la organización, o insertarla en una aplicación o sitio web.

Generando reportes con PowerBI



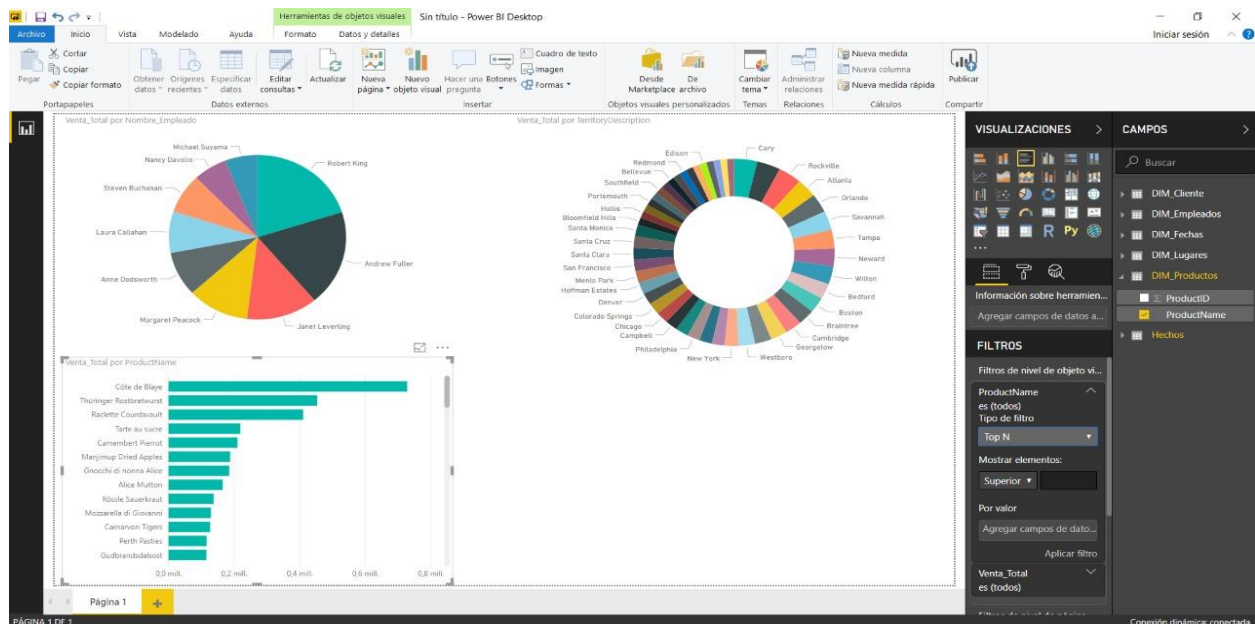






Proyecto de clase

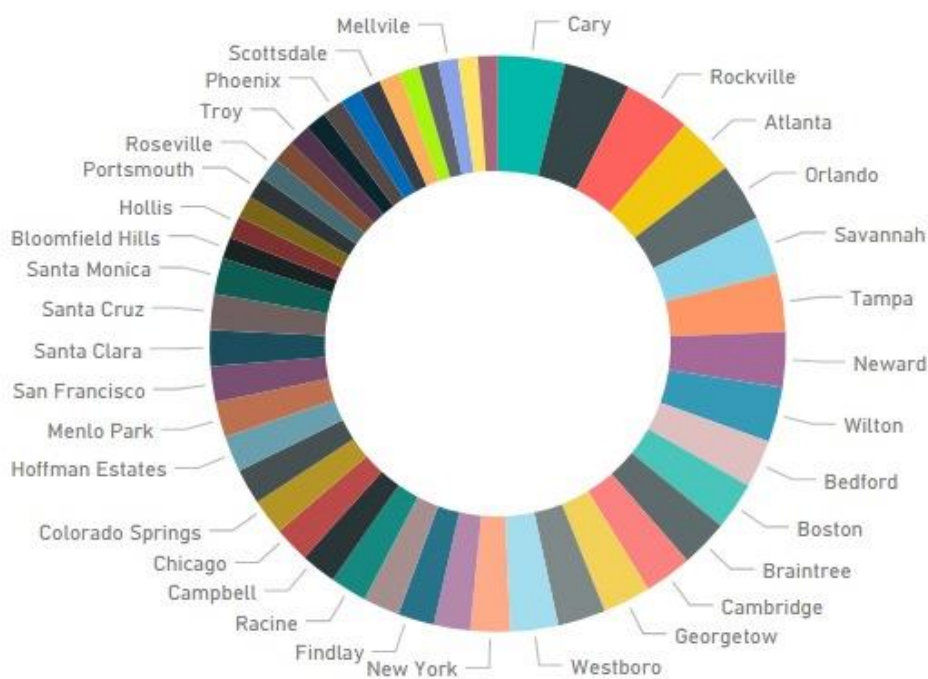
Ing. Emilson Acosta



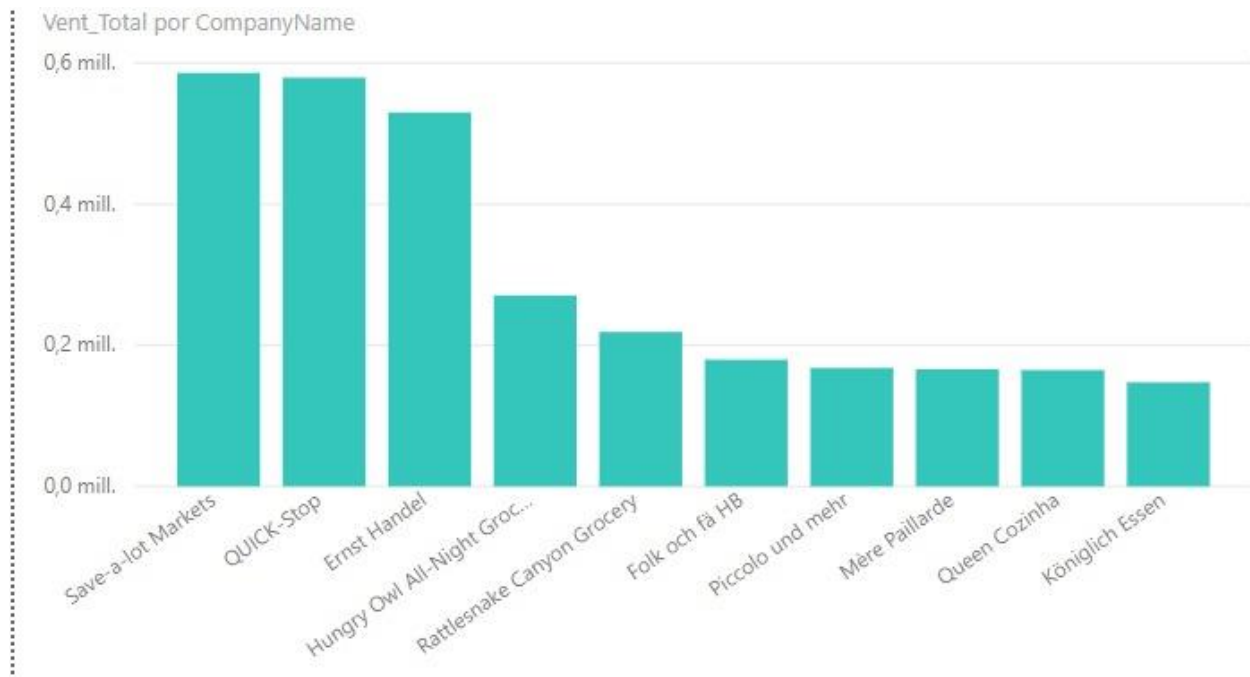


Reportes Generados

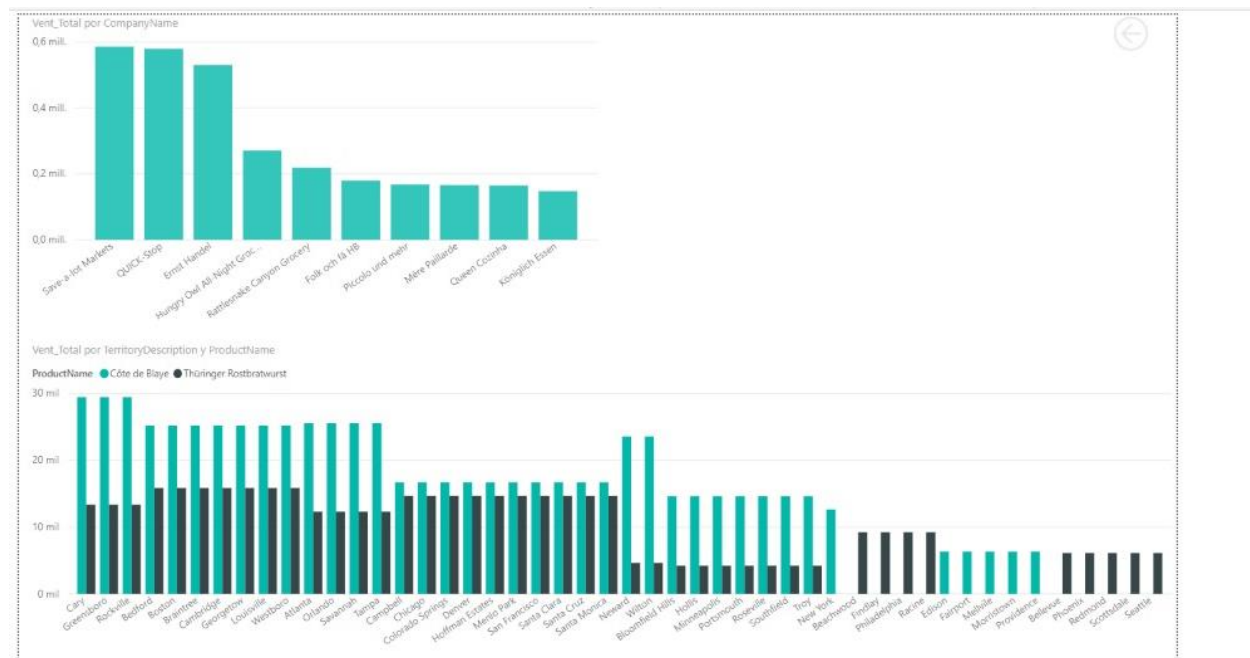
Vent_Total por TerritoryDescription



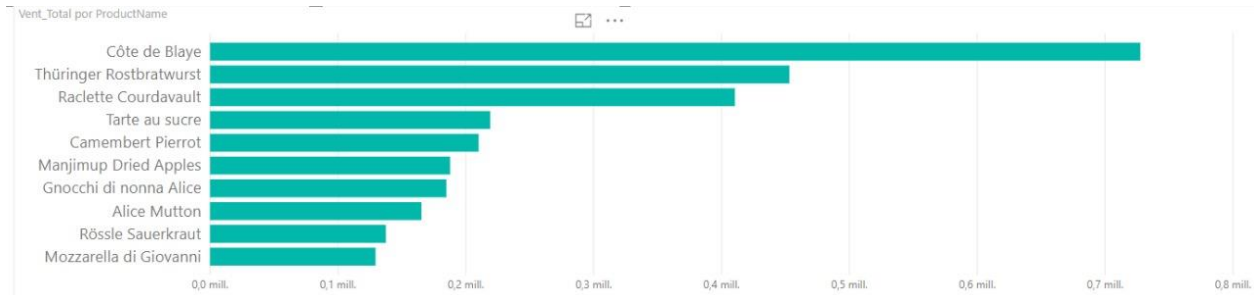
Cantidad de ventas por lugar



Empleados que más número de ventas tiene



Producto más vendido por lugar



Productos más vendidos en general



Tiempo o fecha con más ventas

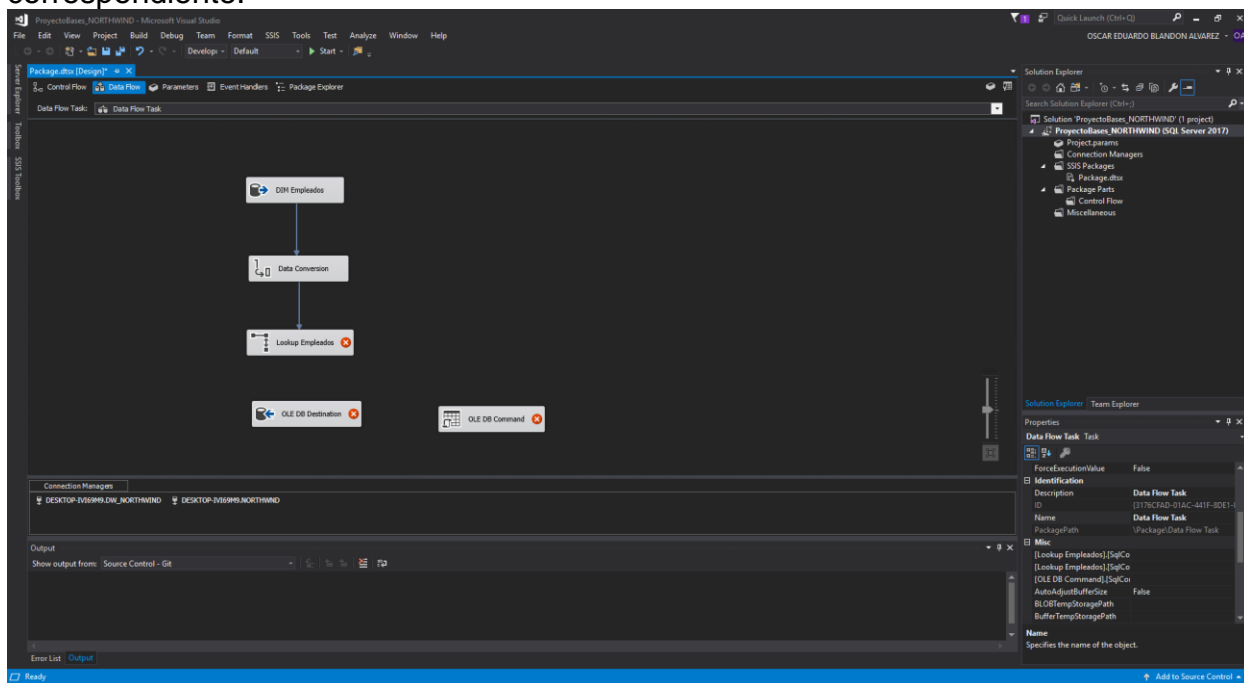


Propuesta para ejecutar proceso ETL más de una vez

El método escogido para solucionar este problema consiste en la utilización de un componente que se encuentra disponible en el set de herramientas de Integration Services, dicho componente es llamado Lookup, el cual especificando una condición nos permitirá desviar datos que ya se encuentren en cada una de las tablas de la base OLAP, para el caso, nuestra condición será corroborar la existencia de un registro mediante su llave primaria por lo que especificamos dicha restricción, para todos aquellos registros en donde no se encuentre coincidencia de los datos entrantes con los datos existentes se ingresaran en la tabla destino mientras que para aquellos que si cumplan la condición y por tanto presenten duplicidad se enviaran a un componente llamado SQL Command el cual hará la ejecución de un Update con los mismos datos que presentan problemas, en otras palabras se hará un update de los datos que ya existen con los datos nuevos duplicados de modo que no se genere el problema de duplicidad sino que se unifiquen dichos registros.

A continuación, se detalla el proceso de configuración:

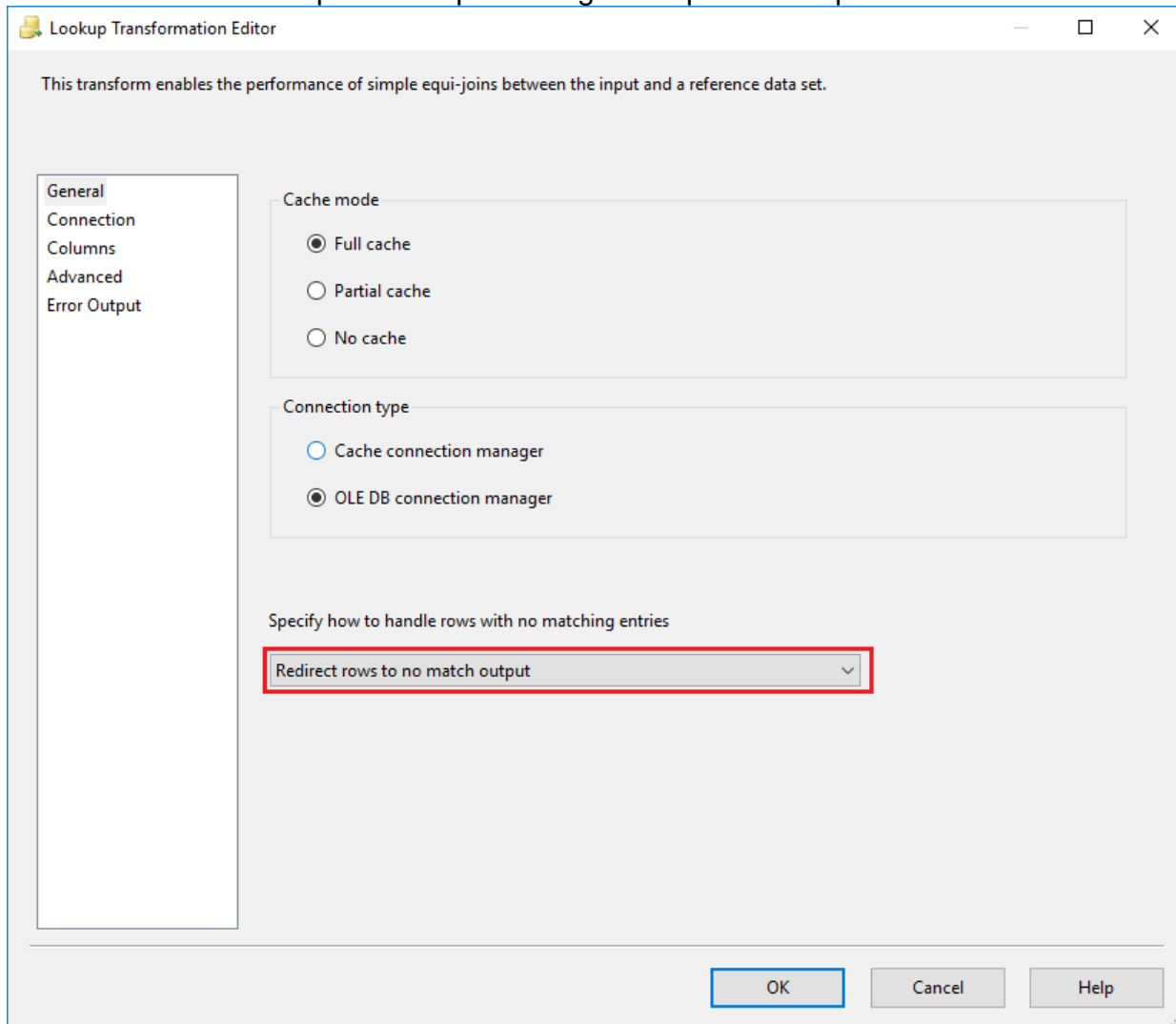
Como primer paso agregamos los componentes necesarios para realizar el ETL correspondiente.



Componentes ETL necesarios

Una vez configurado el origen de datos y la transformación de los datos se procede a configurar el Lookup.

Escogemos el modo de seleccionar los datos en base a la condición, para el caso nos interesa re direccionar primero aquellos registros que no cumplan con la condición.



Lookup – Modo de redirección de los datos



Especificamos la conexión y la tabla donde se analizará la condición

Lookup Transformation Editor

This transform enables the performance of simple equi-joins between the input and a reference data set.

General
Connection
Columns
Advanced
Error Output

Specify a data source to use. You can select a table in a data source view, a table in a database connection, or the results of an SQL query.

OLE DB connection manager:
DESKTOP-IVI69M9.DW_NORTHWIND New...

☒ Use a table or a view:
[dbo].[DIM_EMPLEADOS] New...

☐ Use results of an SQL query:
Build Query...
Browse...
Parse Query

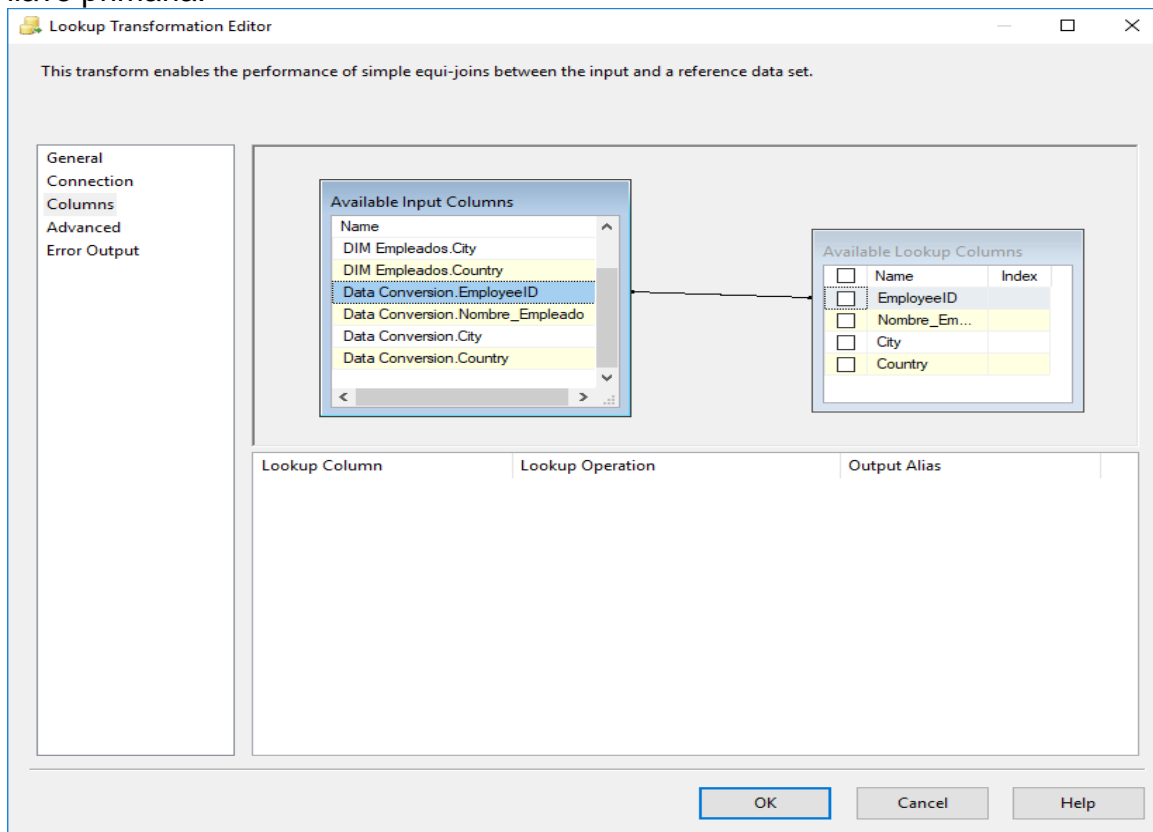
Preview...

OK Cancel Help

Lookup – Conexión base de datos



Especificamos la condición entre los campos que no se deben repetir, en este caso la llave primaria.

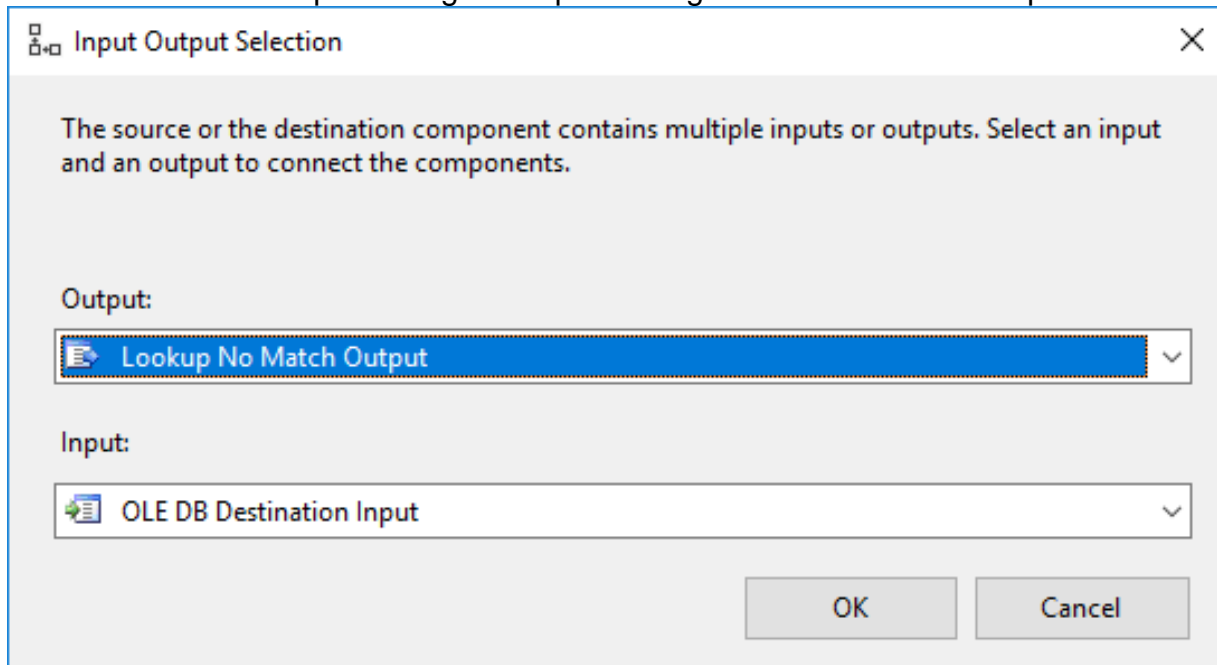


Lookup - Condicion

En el componente destino escogemos una de las salidas que ofrece el Lookup, en este

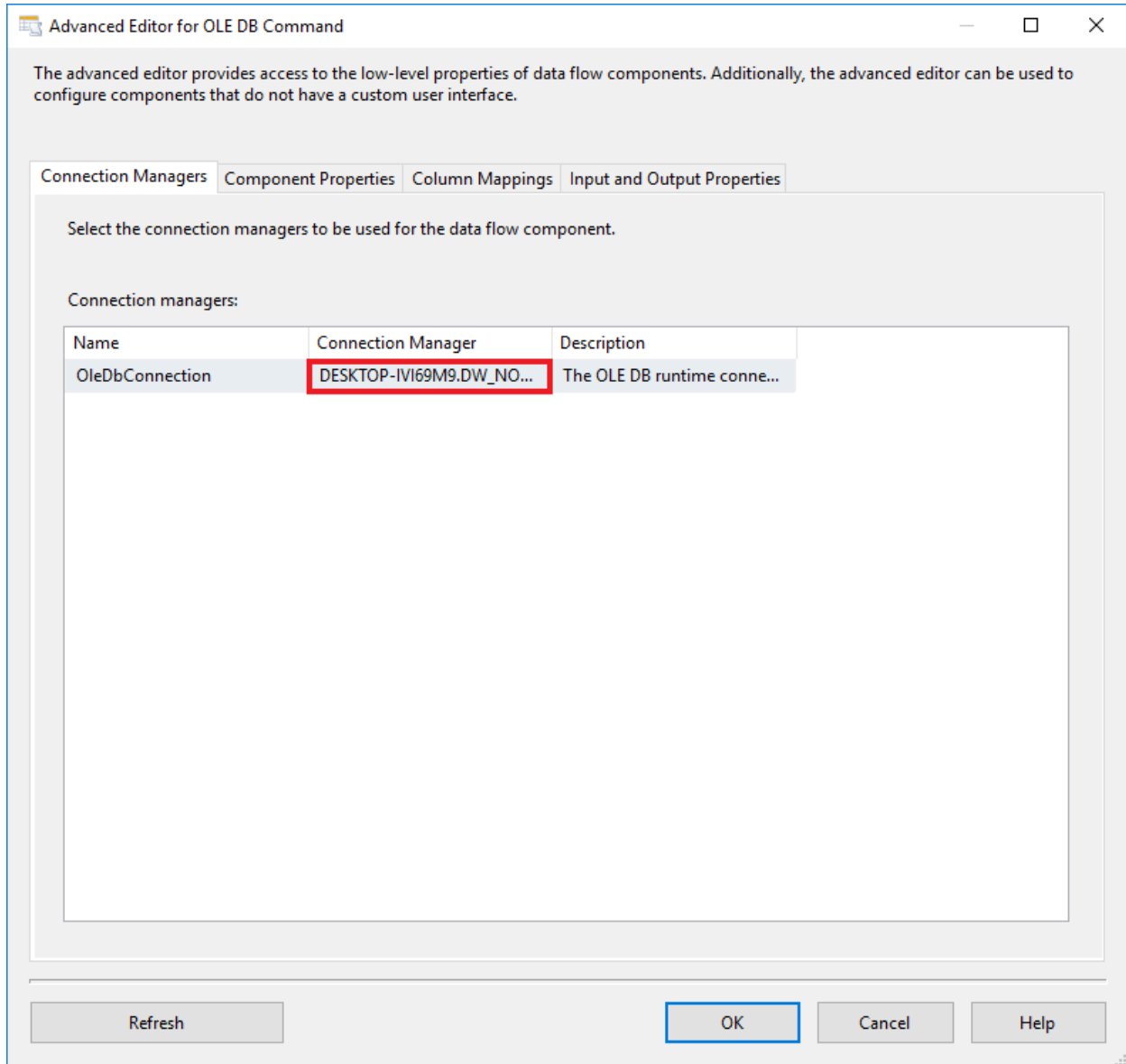


caso nos interesan aquellos registros que no hagan match con datos duplicados.



Lookup – Seleccionar salida de datos

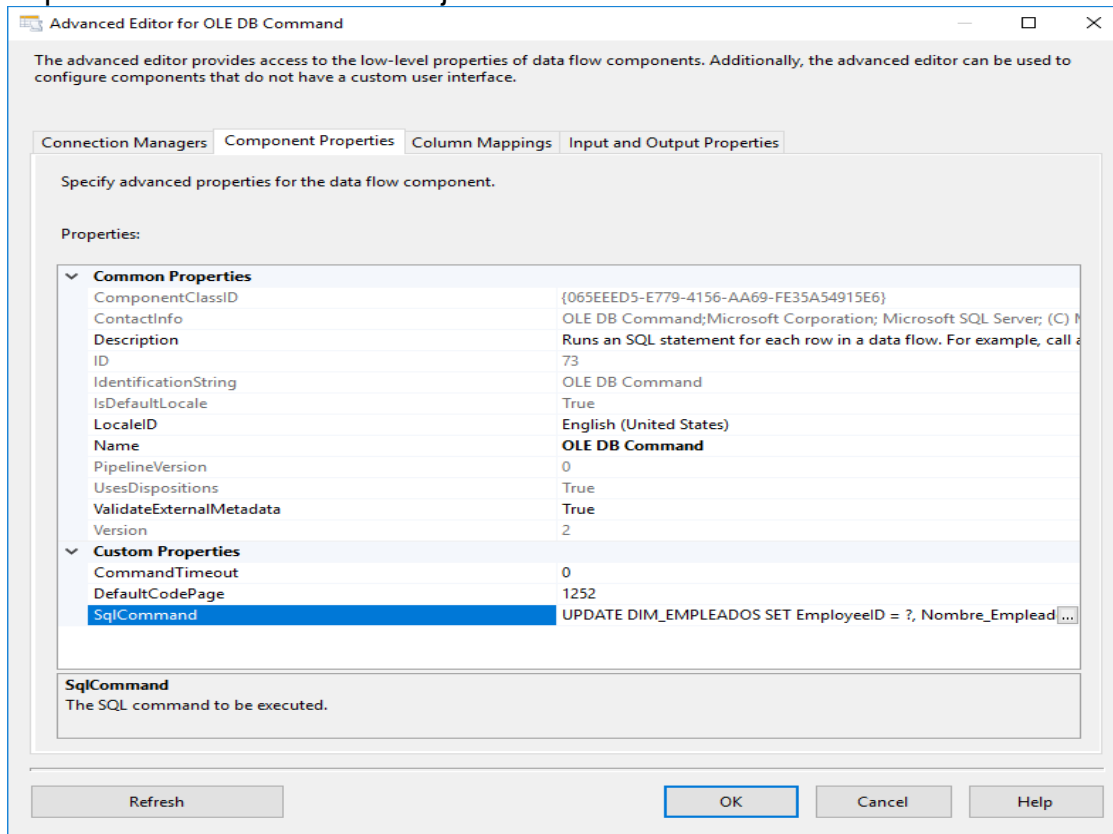
Configuramos el SQLCommand estableciendo la conexión con la base de datos.



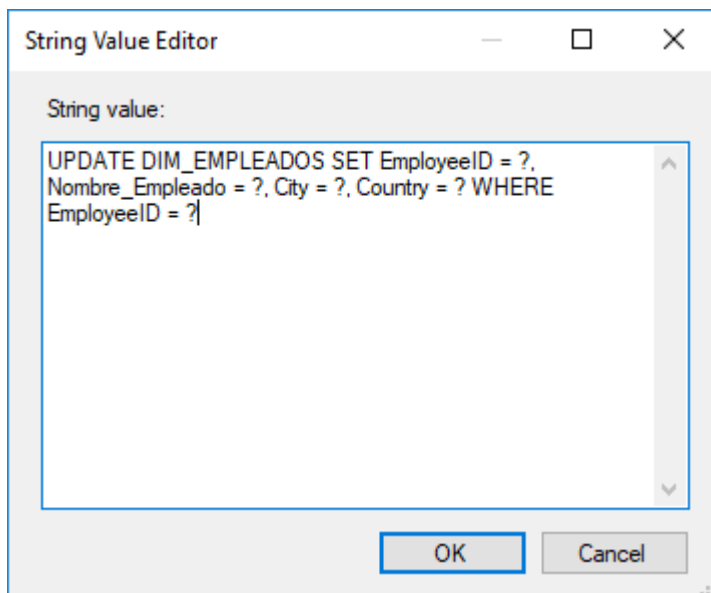
SqlCommand – Establecer conexion



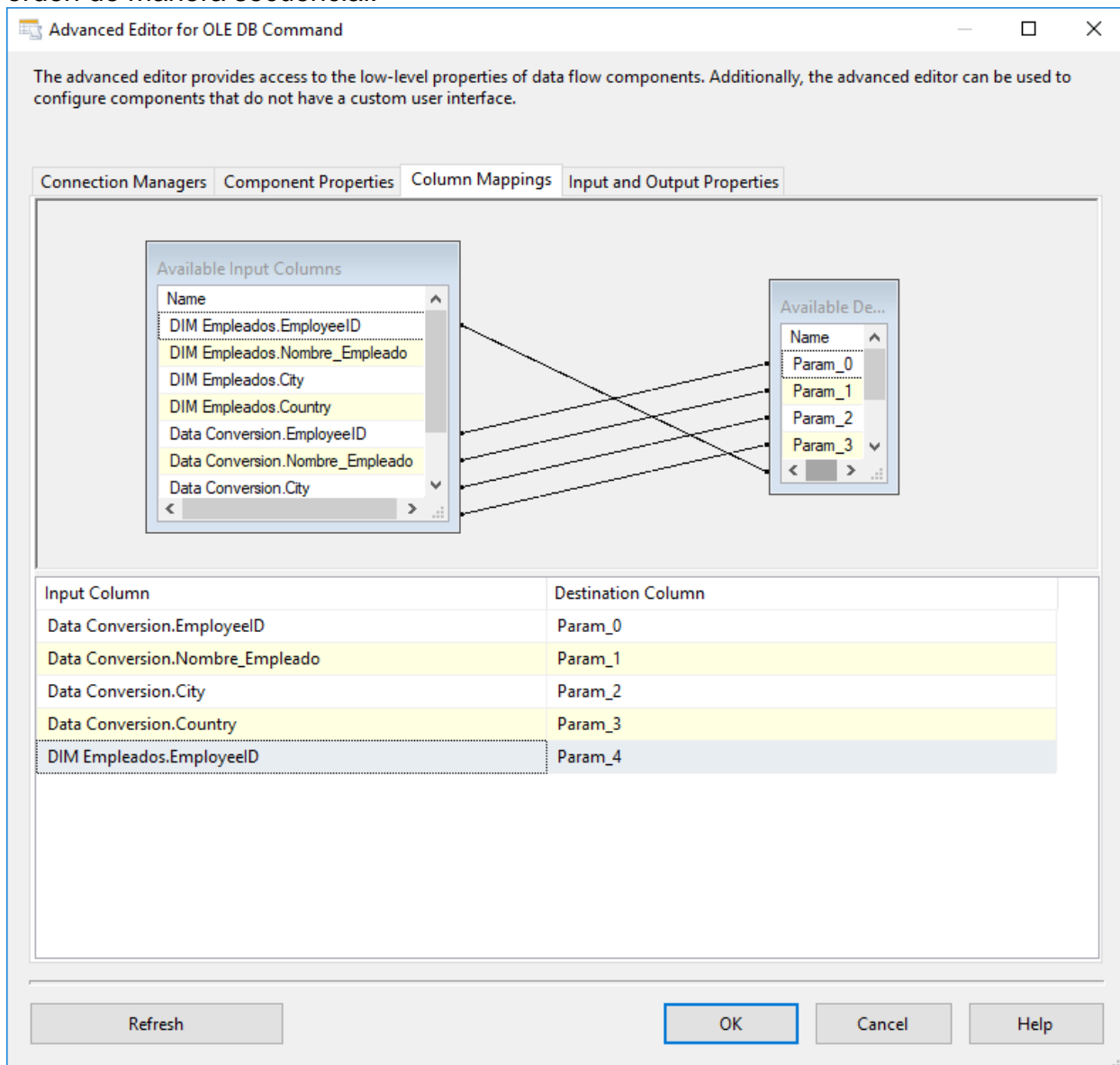
Especificamos el comando a ejecutar.



SqlCommand – Comando a ejecutar.



Hacemos un mapeo de los campos especificados en el comando update, siguiendo el orden de manera secuencial.



SQLCommand – Mapeo de parametros

Una vez realizadas las configuraciones anteriores se nos será posible ejecutar varias veces el ETL sin tener el error de la duplicidad de la información por las llaves primarias.



CONCLUSIONES

- El éxito de la implementación de un almacén de datos se puede medir cuando la empresa consultora y el cliente han ejecutado el proyecto de acuerdo con los requisitos acordados y el alcance del proyecto.
- Para diseñar una buena arquitectura de data warehouse es necesario como primer paso conocer bien los requerimientos del negocio y hacer un estudio profundo de las fuentes externas que nos van a suministrar los datos.
- El Data warehouse ha adquirido gran aplicación en lo que respecta al manejo de datos para el buen desarrollo de una empresa.
- Gracias al data warehouse se da una reducción de costes de almacenamiento y una mayor velocidad de respuesta frente a las consultas de los usuarios.
- El proceso de extracción, transformación y carga (ETL) es quizás el componente más crítico de un almacén de datos; sin embargo, este derecho no concluye que la implementación sea un éxito.
- ETL es un proceso continuo, ya que la base de datos de origen continuará cambiando y, cuando esto ocurra, las reglas de negocios o la lógica de transformación también deberán modificarse y, por lo tanto, es posible que se requieran cambios en el código.

RECOMENDACIONES

- Obtener una comprensión común de qué información es importante para el negocio será vital para el éxito del almacén de datos. A veces, las empresas mismas no conocen sus propias necesidades de datos o panorama. Utilizarán diferentes palabras para los mismos conjuntos de datos. Modelar la información de la empresa puede ser una verdadera revelación para todas las partes involucradas.
- Tener un diagrama de flujo de datos; saber dónde están los repositorios de datos de la empresa y cómo los datos viajan dentro de la compañía en un formato de diagrama permite a todos determinar los mejores pasos para avanzar.



BIBLIOGRAFÍA

Wikipedia. (4 de abril de 2018). Obtenido de

https://es.wikipedia.org/wiki/Cubo_OLAP#Un_ejemplo

Sinnexus. (2018). *B.I.* Obtenido de Sinnexus:

https://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx

Josep Curto, J. C. (2018). *El conocimiento imprescindible*. Obtenido de

<http://reader.digitalbooks.pro/content/preview/books/43005/book/OEBPS/chapter02.xhtml>

Wikipedia. (26 de oct de 2017). Obtenido de https://es.wikipedia.org/wiki/Data_mart

TABLA EVALUCION

Cuenta	Nombre	Puntaje
20151020842	Lester Alexander Flores	100%
20151000308	Oscar E. Blandon	100%
20151003416	German Rene Tejada	90%
20151005232	Génesis Raquel Izaguirre	100%
20141010053	Angie Pamela Argueta	90%