UNIVERSIDAD MAYOR DE SAN ANDRÉS

FACULTAD DE CIENCIAS PURAS Y NATURALES

CARRERA DE INFORMÁTICA



PROYECTO FINAL DATAWAREHOUSE en base a BD Northwind

ASIGNATURA:

SISTEMAS EN TIEMPO REAL Y DISTRIBUIDO INF-317

INTEGRANTES:

Richard Pomacosi Quispe

youtubrick@gmail.com

Yhamin Irvin Polo Apaza entnimras@gmail.com

PARALELO:

Α

DOCENTE:

Lic. Silva Choque Moises Martin

GESTION:

1/2020

INTRODUCCIÓN

La muestra utilizada consiste de la información de la Base de Datos ejemplo Northwind en Postgress, consistente de exportación de productos alimenticios a diversos países. Las tablas principales son la de productos con 77 tipos, categorías de los productos que son 8, empleados de la compañía que son 9, los clientes son 91, los pedidos son 830, los proveedores son 29 y los despachadores son 3.

De acuerdo con lo anterior, el presente proyecto se centrará en el diseo de un DataWere House, que responda una serie de requerimientos establecidos por el cliente, la ventaja ofrecida por el almacén de datos radica en disminución de los tiempos de consulta, gracias al análisis realizado por un equipo de ingenieros que garantiza la disposición y respuesta ágil de la estructura implementada, gracias a la sintesis de procesos habituales en bases de datos relacionales, para ser orientados como elementos de solo análisis.

OBJETIVO

Diseñar e Implementar un DataWareHouse de la base de datos Northwind alojada en PostgreSQL que permita obtener reportes de tablas clave para realizar decisiones administrativas eficientes.

MODELAMIENTO DIMENSIONAL

Gerencia quiere un reporte de datos dinámico sobre las ventas respecto a :

- Ventas anuales y mensuales
- Producto
- Categoría
- Clientes
- Empleados y su Territorio

TABLA DE DIMENSIONES

Las tablas de dimensiones se realizaron en función a los requerimientos de reportes del modelamiento dimensional, estas son:

- D CATEGORIA
- D CLIENTE
- D EMP
- D PRODUCTOS
- D TIEMPO
- D_TERRITORIO

TABLA DE HECHOS

En la tabla de hechos se especificaron las respectivas llaves externas que son de utilidad para la relación entre tablas dimensionales.

H_VENTA

DESARROLLO

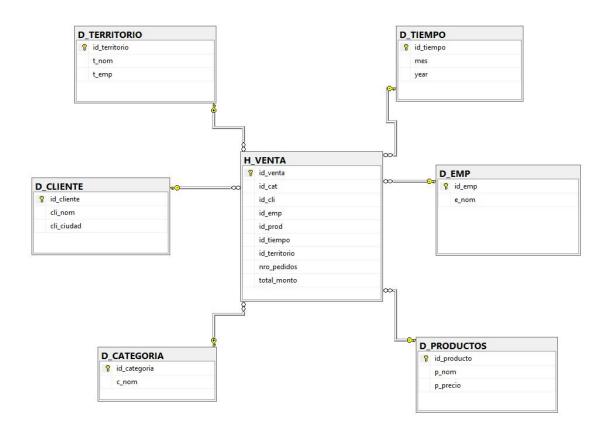
Para alcanzar el objetivo se siguieron los siguiente pasos:

- Creación de la base de datos del DataWareHouse
- Extraccion, Transformación y Carga (ETL)
 - Estandar de acceso
- Creación del cubo OLAP

Extracción, Transformación y Carga (ETL)

Creación de la base datos del DataWareHouse

Se creó la base de datos del DataWareHouse en SQL Server con un esquema en estrella de las tablas, necesarias para crear una instancia del modelo dimensional planteado según las necesidades administrativas.

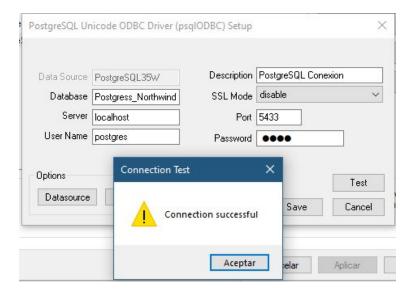


Extracción, Transformación y Carga (ETL)

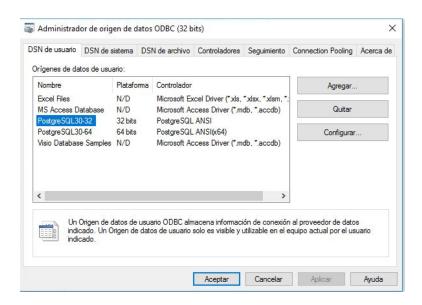
Como se Indicó al principio se tiene una DB llamada Postgress_Northwind en PostgresSQL del cual se migrarán los datos a la BD implementado en SQL Server, previamente a la migración se definió el estándar de acceso al base de datos en Postgress.

Estándar de acceso

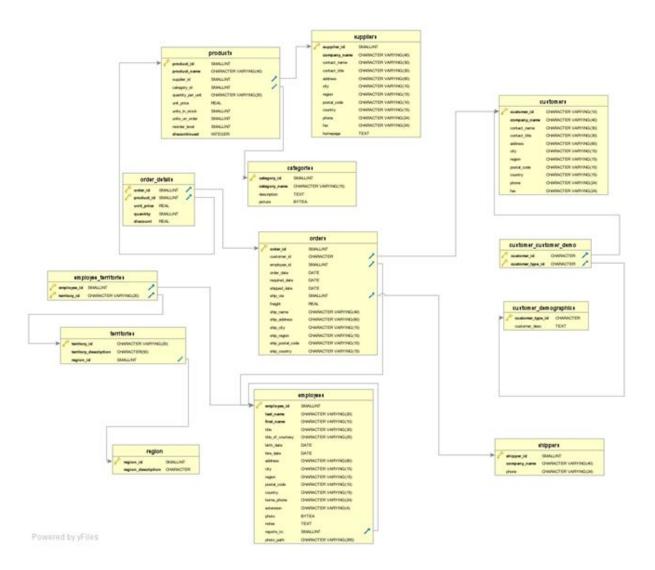
Se utilizó el estándar de acceso ODBC un controlador que tiene por objeto hacer posible el acceso a cualquier dato del BD Northwind en PostgreSQL.Previamente se descargo un controlador de ODBC para PostgreSQL.



Luego en la configuración del ODBC propiamente en Windows se definió el nombre de la base de dato, el puerto por el cual accedemos a los datos y por cuestiones del aplicativo se configuró el ODBC para 64 bits. Seguidamente se probó la conexión y finalizamos.

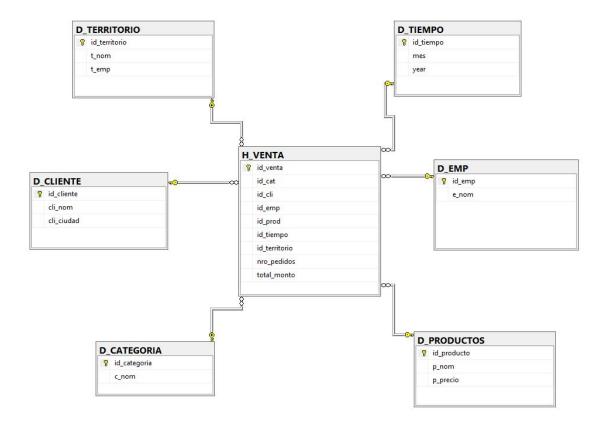


El diagrama ER de esta BD en PostgreSQL es la siguiente:



Con propósitos de hacer un Datawarehouse se hizo una base de datos en SQL Server con las tablas y datos necesarias, siguiendo el modelo de Ralph Kimball, en copo de nieve, esta BD será el destino de la migración de datos, el diagrama ER de la BD

DW_Northwind es el siguiente:



Donde cada tabla es una dimensión del cubo OLAP y la tabla de hechos será la tabla H_VENTAS.

Tablas en la base de datos destino DW_Northwind en SQL Server

D TERRITORIO

- Id_territorio recopilara el territory_id de la tabla "territories" en la BD Postgress Northwind
- t_nom recopilara el campo "territory_description" de la tabla "territories"
- t_emp recopilara el nombre del empleado asignado a ese territorio de la tabla "employees" cruzando datos con la tabla "territories".

D_TIEMPO

- · id_tiempo recopilara el "order_id" de la tabla "orders" en la BD Postgress_Northwind
- mes recopilara el mes del campo "order_date" de la tabla "orders"
- · year recopilara el año del campo "order_date" de la tabla "orders"

D_CLIENTE

- id_cliente recopilara el campo "customer_id" de la tabla "customers" en la BD Postgress Northwind
- · cli_nom recopilara el campo "company_name" de la tabla "customers"
- · cli_ciudad recopilara el campo "city" de la tabla "customers"

D_CATEGORIA

- id_categoria recopilara el campo "category_id" de la tabla "categories" en la BD Postgress Northwind
- · c_nom recopilara el campo "category_name" de la tabla "categories"

D PRODUCTOS

- · id_producto recopilara el campo "product_id" de la tabla "products" en la BD Postgress Northwind
- p_nom recopilara el campo "producto_name" de la tabla "products"
- p_precio recopilara el campo "unit_price" e la tabla "products"

D EMP

- · id_emp recopilara el campo "employee_id" de la tabla "employees" en la BD Postgress Northwind
- e_nom recopilara los campos "last_name" y "first_name" de la tabla "employees"
 H_VENTA
- · id venta es la llave primaria de la tabla y es incremental de a 1.
- Id_territorio recopilara el territory_id de la tabla "territories"
- · id_tiempo recopilara el "order_id" de la tabla "orders" en postgresql
- · id_cliente recopilara el campo "customer_id" de la tabla "customers"
- · id categoria recopilara el campo "category id" de la tabla "categories"
- · id_producto recopilara el campo "product_id" de la tabla "products"
- · id_emp recopilara el campo "employee_id" de la tabla "employees"
- · nro_pedidos recopilara el campo "quantity" de la tabla "order_details"

total_monto – recopilara los campos multiplicados "unit_price" y "quantity" de la tabla "order_details" para tener el monto total de la venta especifica

PROYECTO DE INTEGRACION

Base de datos origen

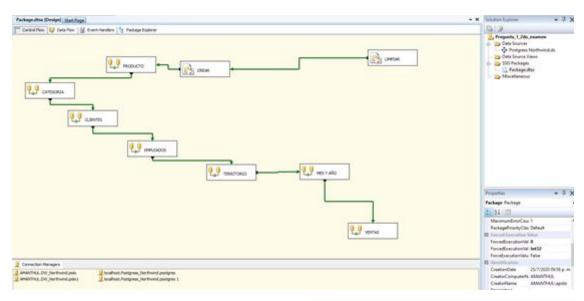
La base de datos de origen fue realizada en PostgreSQL 12.3 usando pgAdmin v4.23. La base de datos se llama: Postgress_Northwind y se adjuntara un backup de la misma en github.

Base de datos destino

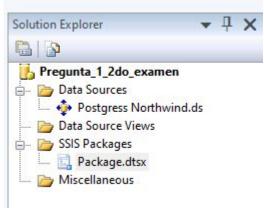
La base de datos destino fue realizada en SQL Server 10.50.16 (SQL Server 2008) utilizando SQL Server Management Studio v18.6 para su manejo.

La base de datos se llama: DW_Northwind y se adjuntara un backup de la misma en github. El proyecto de integración se hace en una herramienta de SQL Server 2008, esta se llama SQL Server Business Intelligence Development Studio que corre sobre Visual Studio 2008 v9.0.30729.4462 QFE.

Una imagen del proyecto terminado es la siguiente:



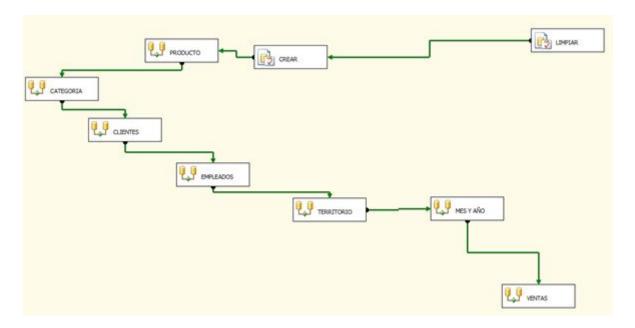
Donde el campo de conexiones y package es el que se aprecia en el gráfico.



EL administrador de conexiones es el siguiente:



Y el flujo de datos es el siguiente:



Donde:

LIMPIAR – es un SQL Task que borra la tabla de hechos, en nuestro modelo la tabla H_VENTA en la BD DW_Northwind que se aloja en SQL Server.

CREAR – es un SQL Task que crea la tabla de hechos, en nuestro caso crea la tabla H_VENTA con todas sus relaciones.

PRODUCTO – es un flujo de datos que extrae los datos de la tabla "products" alojado en la BD Postgress_Northwind

CATEGORIA - es un flujo de datos que extrae los datos de la tabla "categories" alojado en la BD Postgress_Northwind

CLIENTES - es un flujo de datos que extrae los datos de la tabla "customers" alojado en la BD Postgress_Northwind

EMPLEADOS – es un flujo de datos que extrae los datos de la tabla "employees" alojado en la BD Postgress_Northwind mediante un comando SQL.

TERRITORIO - es un flujo de datos que extrae los datos de la tabla "territories" y "employees" alojados en la BD Postgress_Northwind, extrae mediante un comando SQL.

MES Y AÑO - es un flujo de datos que extrae los datos de la tabla "orders" alojado en la BD Postgress_Northwind, mediante la herramienta de columna derivada se extrae el mes y año del campo "order_date" de la tabla "orders" y se crea dos nuevos campos "mes" y "año" propiamente que serán nuevos campos en la tabla "D_TIEMPO".

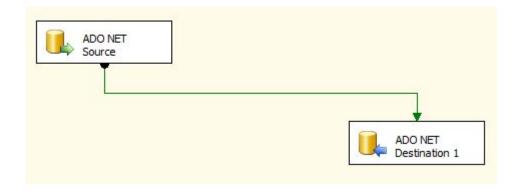
VENTAS - es un flujo de datos que extrae los datos de las tablas "employees", "order_datails", "categories", "customers", "products", "order" y "territories" alojados en la BD Postgress_Northwind mediante un comando SQL, multiplica también dos campos, el campo "unit_price" y "quantity" para tener un monto total a cobrar por la venta realizada, este nuevo campo se llama "Total" y será asignado a la tabla H_VENTA en el campo "total_monto".

Tambien en este flujo de datos hice una operación más, este nuevo campo "Total" que surge de la multiplicación de "quantity" y "unit_price", mutiplicados generan un tipo de dato entero de 8 bytes DT_I8, y en la base de datos destino DW_Northwind, el campo destino "total_monto" aloja solamente tipos de datos enteros de 4 bytes DT_I4.

Por lo tanto tuve que usar la herramienta de conversión de datos para hacer la respectiva transformación para que no haya perdida de datos o problemas.

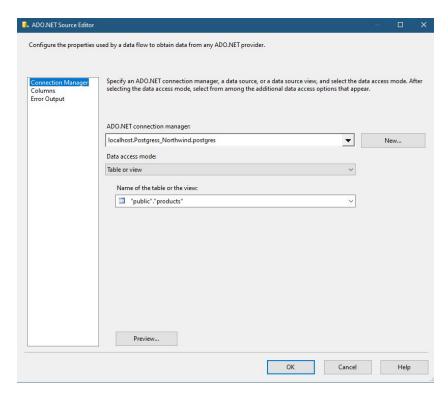
Breve explicacion de la configuración del flujo de datos:

Dentro de los flujos, en la mayoria, se siguió el siguiente tipo de conexion:

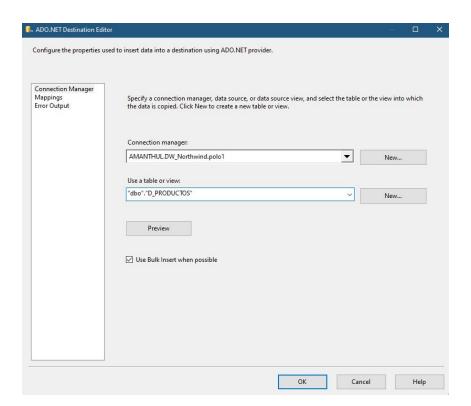


ADO.NET por que la conexion de origen es a través de PostgreSQL y usamos una interfaz ODBC, ADO NET para la conexión destino que es a traves de SQL Server.

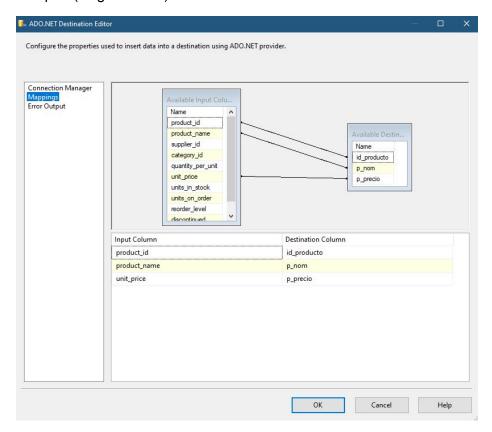
Definición del origen



Definición del destino

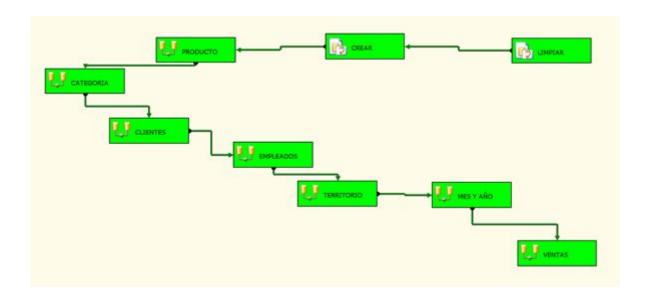


Proceso de mapeo (asignaciones).



PROBANDO LA INTEGRACIÓN

Luego de correr el flujo creado la imagen es la siguiente:



En la base de datos origen Postgress_Northwind, al ejecutar el comando SQL que utilizamos para el flujo de datos VENTAS, este es el resultado:

SELECT Categories.Category_ID, Customers.Customer_ID, Employees.Employee_ID, Orders.Order_ID, Products.Product_ID, Order_Details.Quantity, Order_Details.Unit_Price * Order_Details.Quantity AS Total, Territories.Territory_ID

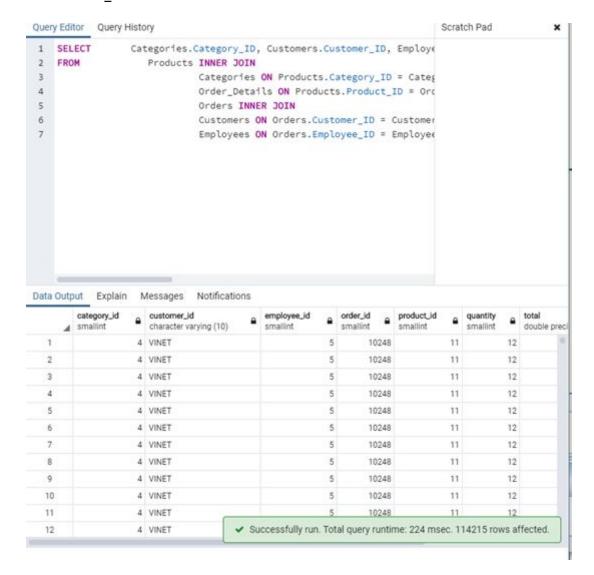
FROM Products INNER JOIN

Categories ON Products.Category_ID = Categories.Category_ID INNER JOIN
Order_Details ON Products.Product_ID = Order_Details.Product_ID INNER JOIN
Orders INNER JOIN

Customers ON Orders.Customer_ID = Customers.Customer_ID INNER JOIN

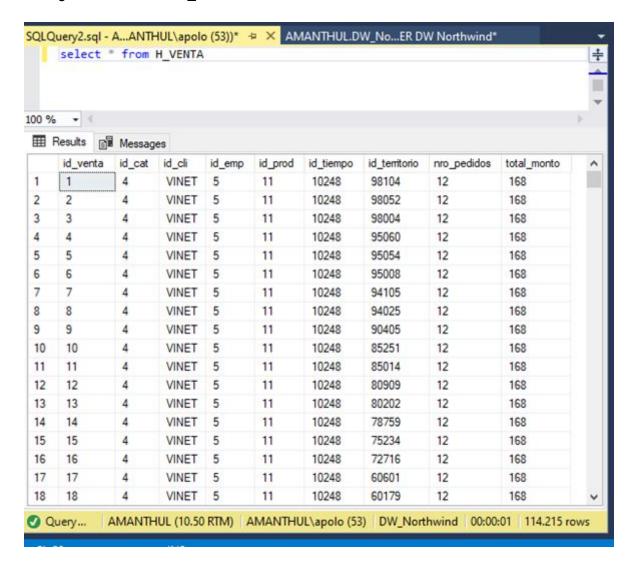
Employees ON Orders.Employee_ID = Employees.Employee_ID ON Order_Details.Order_ID

= Orders.Order_ID CROSS JOIN Territories



El total de registros es de 114215, eso indica que debería ser el mismo numero de registros en la BD destino DW_Northwind en la tabla "H_VENTA.

Al correr el flujo creado, procedo a verificar mediante el SQL Server Management Studio los registros de la tabla H_VENTAS en la base de datos destino.

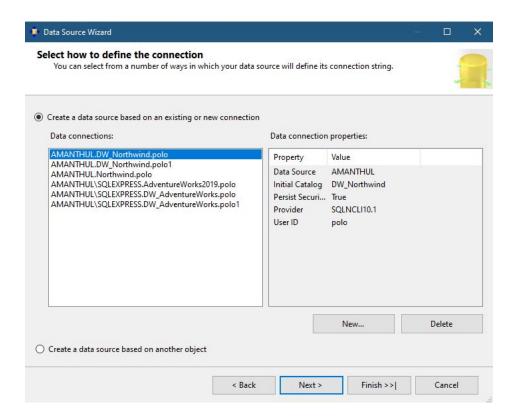


Se puede ver que la tabla H_VENTA contiene los 114215 registros de la consulta hehca en PostgreSQL con la BD origen.

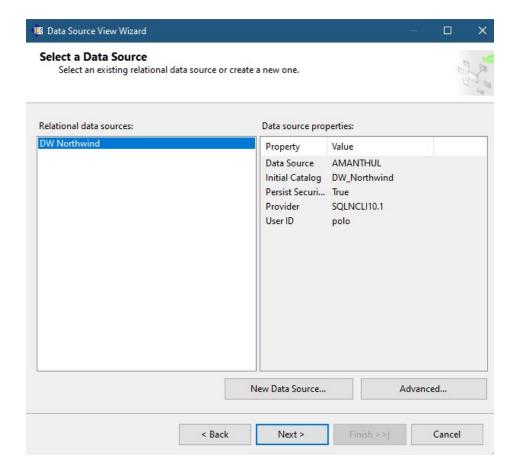
CREACIÓN DEL CUBO OLAP

Ya terminado y corriendo el proyecto de integración, entonces se puede proceder a realizar el análisis dimensional, esta vez se hara el analisis dimensional en SQL Server Business Intelligence Development Studio que corre sobre Visual Studio 2008 v9.0.30729.4462 QFE.

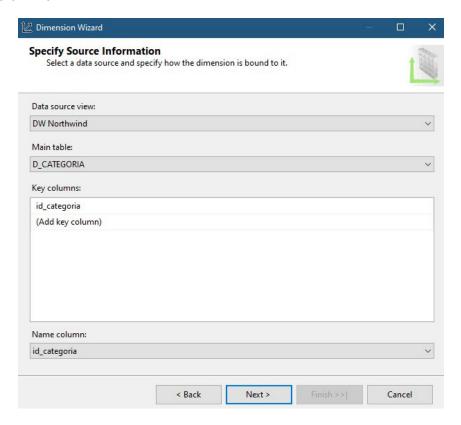
Una vez creado el proyecto, asignamos la fuente de datos como la BD DW_Northwind, ya que es la BD destino donde tenemos los datos que necesitamos.



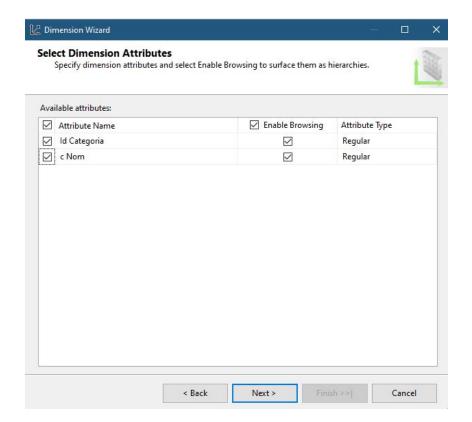
Luego procedemos a crear las vistas, y elegimos las mismas de nuestra BD DW_Northwind.



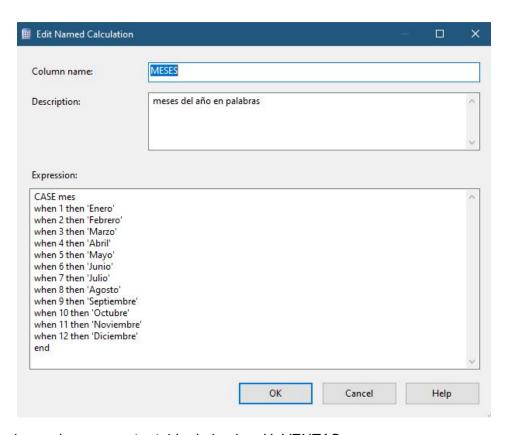
Luego procedemos a crear las dimensiones, que vendrian a ser cada tabla de nuestra DW_Northwind.



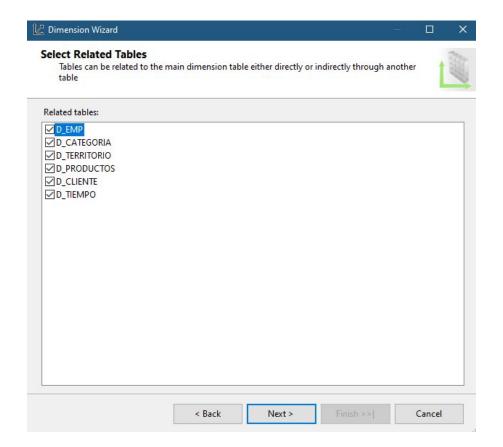
Elegimos cada campo de nuestras tablas.



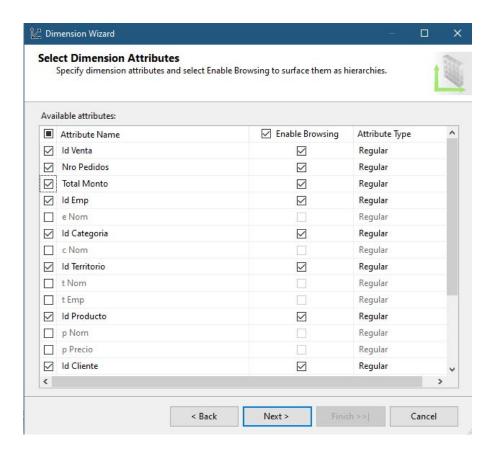
Para que los meses dentro del analisis sean faciles de entender, pusimos la siguiente regla dentro la dimension de tiempo.



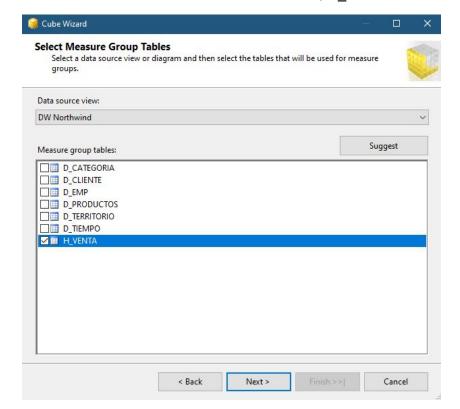
Tambien lo propio con nuestra tabla de hechos H_VENTAS

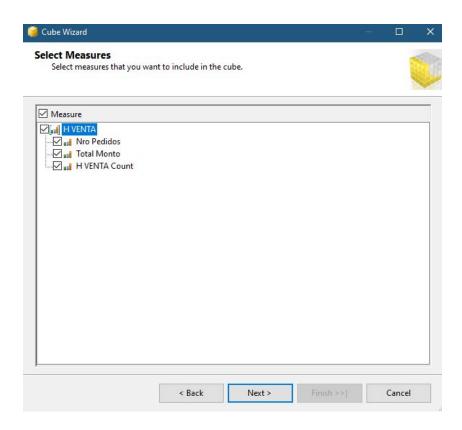


Elegimos los campos "total_monto" y "nro_pedidos" pra realizar los analisis de las ventas.

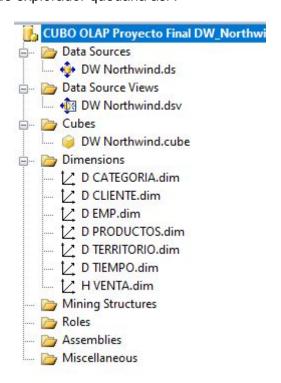


Ahora creamos el cubo orientado a nuestra tabla de hechos, H_VENTAS.

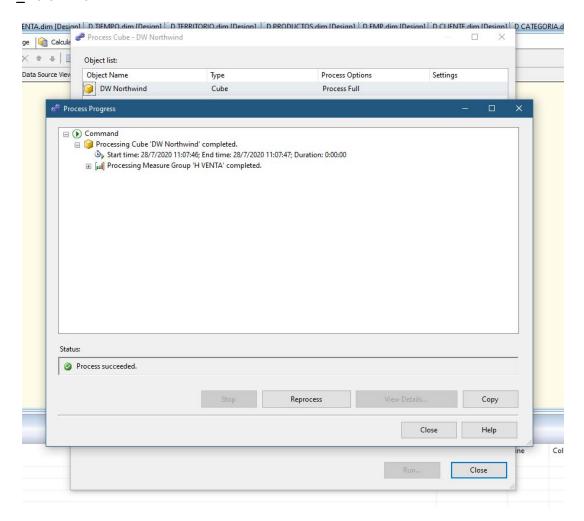




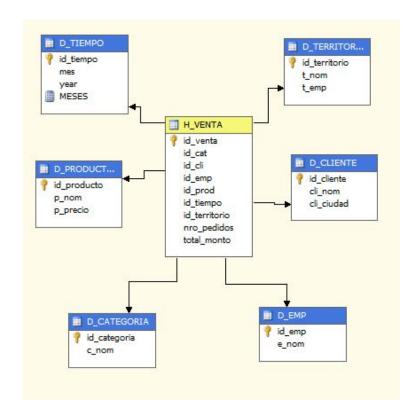
Entonces nuestra barra de explorador quedaria asi .



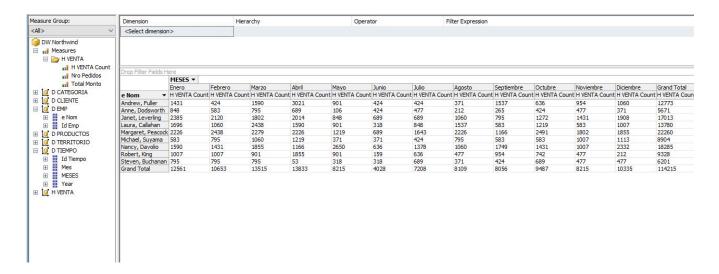
Procedemos al paso final, a procesar el cubo creado a partir de la base de datos DW_Northiwnd.



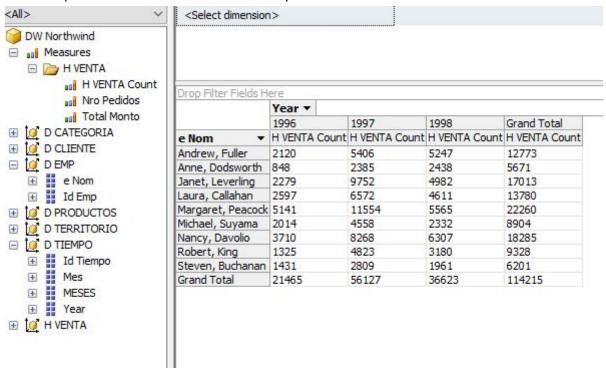
El esquema final queda de la siguiente manera:



Vamos a BROWSER para analizar, y analicemos el conteo de ventas que realizo cada empleado a lo largo de los meses en los ultimos años.



Tambien podemos hacer el mismo analisis pero a traves de los años.



De esta manera el analisis mediante el cubo dimensional es ejecutado exitosamente y el Datawarehouse fue construido siguiendo los lineamientos administrativos.

LINK AL PROYECTO

https://github.com/TharEntnimras/PROYECTO FINAL