

IS601, BASE DE DATOS II

Primer Periodo 2021

Informe Proyecto de Clase Elaborado por:

**20151030886**

**Víctor Miguel Pineda Alonzo**

Catedrático: **Emilson Acosta**

# Índice

Contenido

[Índice 2](#_Toc69337220)

[Introducción 4](#_Toc69337221)

[Marco Teórico 5](#_Toc69337222)

[Base de Datos OLTP 10](#_Toc69337223)

[Diagrama Relacional 10](#_Toc69337224)

[Compras 10](#_Toc69337225)

[Persona 11](#_Toc69337226)

[Producción 12](#_Toc69337227)

[Recursos Humanos 13](#_Toc69337228)

[Ventas 14](#_Toc69337229)

[Base de Datos OLAP 15](#_Toc69337230)

[Diagrama Estrella 15](#_Toc69337231)

[Métricas 16](#_Toc69337232)

[ETL 17](#_Toc69337233)

[**Transformaciones para las tablas de dimensión** 17](#_Toc69337234)

[Transformación Orden Ventas 18](#_Toc69337235)

[Transformación Territorio 21](#_Toc69337236)

[Transformación Productos 24](#_Toc69337237)

[Transformación Tiempo 27](#_Toc69337238)

[**Transformaciones para las tablas de hechos** 30](#_Toc69337239)

[Transformación Tiempo 30](#_Toc69337240)

[**JOB** 33](#_Toc69337241)

[Creación de Reportes 35](#_Toc69337242)

[**Reporte 1** 37](#_Toc69337243)

[**Reporte 2** 41](#_Toc69337244)

[**Reporte 3** 45](#_Toc69337245)

[**Reporte 4** 48](#_Toc69337246)

[Creación de Dashboard 51](#_Toc69337247)

[Conclusiones 56](#_Toc69337248)

[Recomendaciones 57](#_Toc69337249)

[Referencias 58](#_Toc69337250)

# Introducción

Los Datawarehouse son almacenes de grandes cantidades de datos que tiene como objetivo el análisis de esta información para mejorar la toma de decisiones dentro de las empresas.

La gerencia comercial de Adventure Works nos ha solicitado un dashboard de ventas por producto y su territorio donde se han realizado sus compras y establecer filtros para saber cuando se han realizado sus ventas y el total de las ventas por producto.

Como ingenieros en sistemas con conocimientos en business inteligence somos capaces de solventas este tipo de problemas aplicando el conocimiento adquirido sobre bases de datos OLAP y bases de datos OLTP, así como también desarrollaremos ETLs para poder llenar esta base de datos OLAP de manera automática y periódica.

# Marco Teórico

Para comenzar a explicar que es un **Datawarehouse** podemos encontrar que se llegó a originar en 1988 de la mano de investigadores de la empresa IBM, Barry Devlin y Paul Murphy, pero en realidad quien acuño el termino de Datawarehouse fue William H. Inmon, más conocido como el padre de Datawarehousing. William Inmon manifestó que el Datawarehouse sería como una colección de datos orientados sobre un tema específico, para que pueda llegar a soportar el proceso de toma de decisiones.

Entrando en materia específica sobre el Datawarehouse, lo podemos describir como una tecnología que unifica los datos entre diversos sistemas, esta tecnología hace hincapié en la recopilación o captura de los datos, todo esto para tener un fin analítico y de fácil acceso. Podemos usarla para contener los datos necesarios para la empresa y poder utilizarlos en un futuro para poder organizar la información valiosa para la empresa y los mismos clientes, permite a los ejecutivos u empleados de negocios organizar, comprender y utilizar los datos con mayor eficiencia para poder tomar decisiones ventajosas.

Los datos que alimentan a un datawarehouse son originados de las distintas bases de datos que puedan tener los departamentos de la empresa, estas bases de datos tendrán distintas estructuras por lo que para facilitar la lectura y el análisis de los datos es necesario hacer el proceso de integración. Para que no existan conflictos entre los datos se suele normalizar los datos antes de combinarlos en el almacén todo esto es posible hacerlo mediante herramientas que extraen, transforman y carga más conocido como ETL, por sus siglas en inglés (Extract, Transform and Load). Esto lo que realiza es que leen los datos, a menudo de una base de datos OLTP, transforma al almacén o sea que filtra, adapta o cambia el formato y luego de esto lo vuelve a escribir ya conjunto.

Se asegura que el datawarehouse debe cumplir la función de integración, como se mencionó antes, todos los datos recolectados deberían integrase en una estructura consistente y estructurase según las necesidades. Debe de ser temático, los datos al estar organizados por temas facilitan así el acceso a estos y hace que sean fácilmente digeribles, aunque también debe de ser no volátil la información guardada ya no puede ser modificada y es permanente. La última característica de un datawarehouse es que esta estrictamente vinculada al tiempo, así los datos pueden ser servir para marcar tendencias para el análisis a lo largo del tiempo.

Los datawarehouse contienen una estructura dividida en tres: básica, básica con un área de ensayos y básica con área de ensayo más data marts.

Estructura básica. Los sistemas operativos y archivos planos proporcionan datos en bruto que almacenan metadatos esto con el fin que el usuario pueda acceder a ellos para realizarse el respectivo análisis o hacer minería de datos.

Estructura básica y área de ensayo. Se pueden colocar las fuentes de datos y el almacén, proporciona un espacio para que los datos sean integrados antes de ingresar al almacén.

Estructura básica y área de ensayo más Data Marts. Son sistemas diseñados para una la línea de negocio especifica.

Podríamos incluir una cuarta parte que serían los Sandboxes, son privados, seguros y contienen áreas seguras que permiten a las compañías explorar nuevos datasets o formas de analizar los datos de manera más rápida y formal, todo sin tener que usar las reglas y protocolos de un data warehouses.

A pesar de que los data warehouses son muy distintos a los data marts y a los ODSs (por sus siglas en ingles Operation Data Stores). Un data mart tiene las mismas funcionalidades de un data warehouse pero con un espectro más estrecho, usualmente enfocado a un solo departamento o una sola línea de negocio. Esto hace que se acople perfectamente a un data warehouse. Aunque suelen tender a introducir inconsistencias gracias a la dificultad de integrar y manejar los datos a través de muchos data marts.

**OLAP - (On-Line Analytical Processing)**

Es una Técnica de Análisis Multidimensional Diseñado, para lograr un buen rendimiento en consultas ad-hoc, podremos hacer vista multidimensional de los datos y mecanismos para almacenar un cubo. Sirve como puente entre como los datos están almacenados en la bodega y en cómo son presentados al usuario.

Fácil de usar por los analistas del negocio

–Navegar en los Datos

–Velocidad de las consultas

–Esconde complejidad

–Riqueza analítica

Así como permite más fácilmente:

–Analizar datos

–Generar reportes

–Acceder los datos por navegadores de web

–Visualizar datos

–Importar datos

Mientras que el **Cubo** Contiene datos de primer interés para los usuarios y es un subconjunto de los datos que están en la bodega. Contiene valores agregados a todos los niveles de las dimensiones usados para organizar los datos en dimensiones y medidas. Así como estos mejoran la velocidad de consulta.

El cubo puede responder preguntas que incluyan tres dimensiones y una medida:

–Dimensión producto: contiene categorías del producto.

–Dimensión almacén: contiene almacenes.

–Dimensión tiempo: contiene periodos del año.

–Medida ventas: cantidad numérica que puede ser sumarizada.

**Operaciones OLAP**

Roll Up: dimensiones generales.

Drill Down: dimensiones específicas.

Slicing: es la selección de un grupo de celdas de la matriz multidimensional especificando un valor para una o más dimensiones.

Dicing: rotar el cubo para mirar otras dimensiones.

**MOLAP y ROLAP**

Las arquitecturas de los Bodegas han evolucionado en base al crecimiento, en funcionalidad y rendimiento, de los DBMS. Es por ello, que la arquitectura reinante de los ’80 fue el modelo MOLAP (Multidimensional OLAP). En los ’90, surge una nueva arquitectura de DW denominada ROLAP (Relational OLAP).

Utilizando **MOLAP** nos ofrece como capacidad de análisis:

–Ofrece vistas de objetos multidimensionales

–Tiempo de respuesta cero, pues tiene todo precalculado.

–Si no se precalcula todo (en general todo el precálculo tiene volúmenes inaceptables) la capacidad de análisis se limita a aquellas porciones del cubo que fueron precalculadas.

Sistema de diseño propietario:

–Generalmente el cubo se trata de una “caja negra” de datos encriptados que pueden residir de forma local o en un servidor MOLAP.

–Flexibilidad y escalabilidad limitados.

–Cambios en el modelo dimensional del negocio implican la generación de todos los cubos nuevamente.

**ROLAP** nos ofrece en cuanto capacidad de análisis:

–Ofrece vistas de objetos multidimensionales.

–Tiempos de respuestas que rondan entre los segundos y los minutos.

–Existen técnicas de tuning, caching, materialización de vistas, indexación y esquema de diseño que mejoran la performance de respuesta de los ROLAP.

Sistema de diseño abierto:

–El cliente interactúa directamente contra el RDBMS vía SQL en distintos motores.

–Provee flexibilidad y escalabilidad.

–Los cambios en el modelo dimensional del negocio son trasladados al DW e inmediatamente se encuentra disponible para las consultas pertinentes.

–La ventana de carga del datawarehouse es menor pues no existe el tiempo de generación de los multi-cubos.

# Base de Datos OLTP

## Diagrama Relacional

Se utilizará la base de datos de pruebas de Microsoft Adventure Works en su versión 2014.

Al ser una base de datos realmente grande se dividirá el diagrama en sus categorías internas.

Compras

Persona



### Producción

### Recursos Humanos





### Ventas

# Base de Datos OLAP

Se creo la base de datos DW\_AdventureWorks que servirá como base de datos OLAP y tendrá esta estructura en estrella.

## Diagrama Estrella

# Métricas

Comenzamos con una identificación de sustantivos y verbos distribuyéndose así:

Ventas:

Cantidad

Precio de venta

Dimensiones:

Productos

Tiempo

Lugar de venta

Es de ahí de donde salen nuestras dimensiones producto, tiempo, ventas y territorio.

# ETL

Se comenzará con el desarrollo del ETL utilizando la herramienta Spoon.bat de Pentaho.

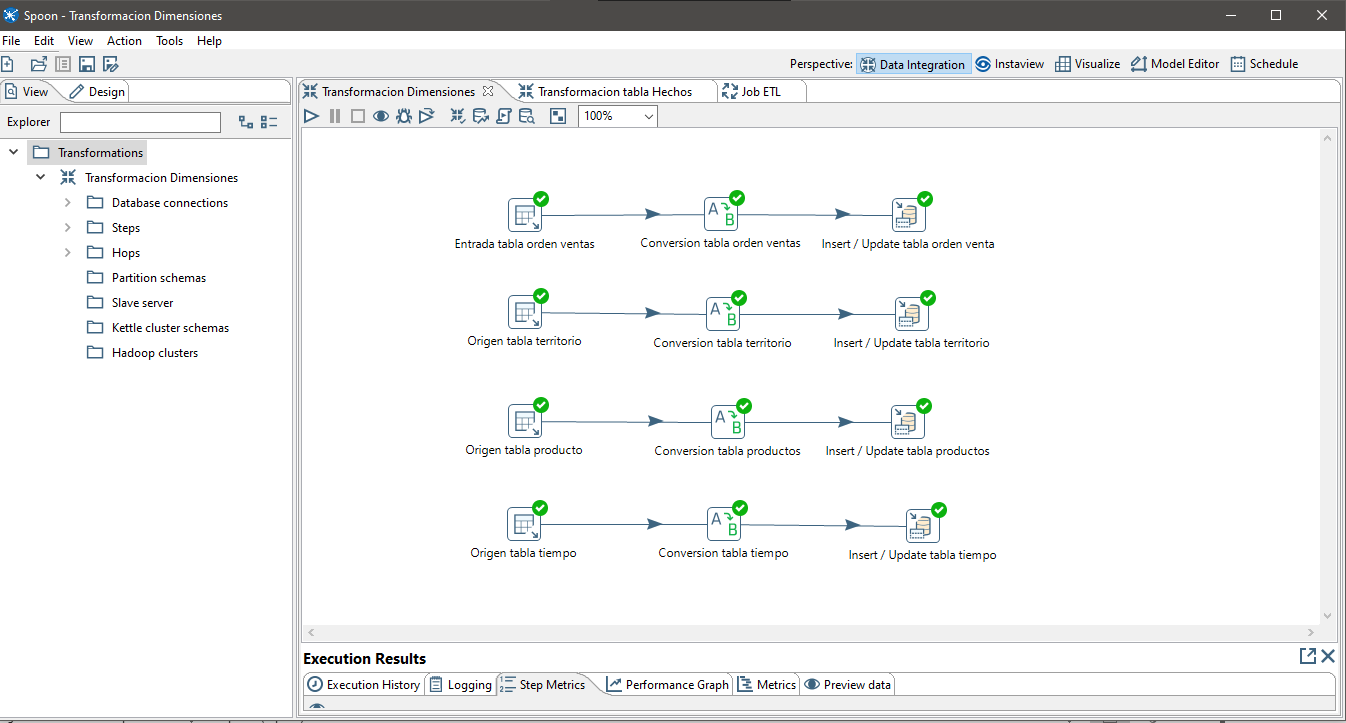
## **Transformaciones para las tablas de dimensión**

Se utilizarán tres componentes en todo el desarrollo de las trasformaciones:

Input - Table Input

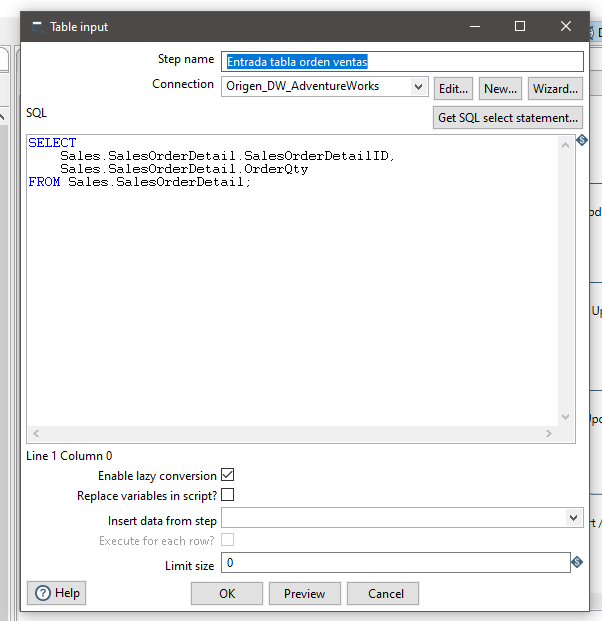
Transform – Value Mapper

Output – Insert/Update

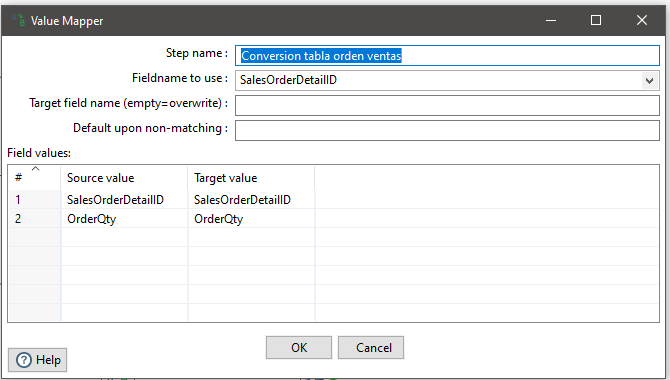


### Transformación Orden Ventas

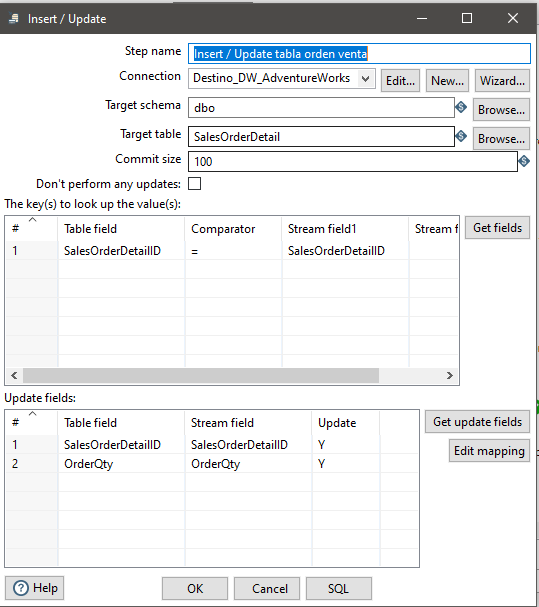
#### Configuración Input Table Input



#### Configuración Transform Value Mapper

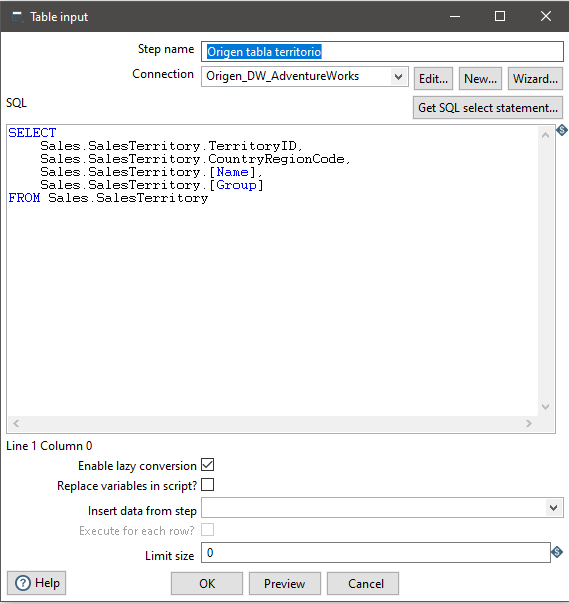


#### Configuración Output Insert/Update

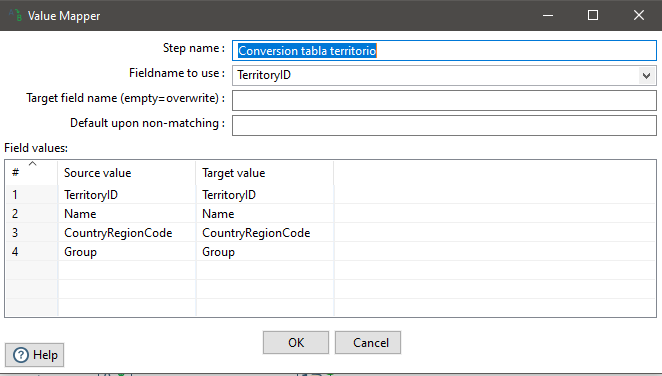


### Transformación Territorio

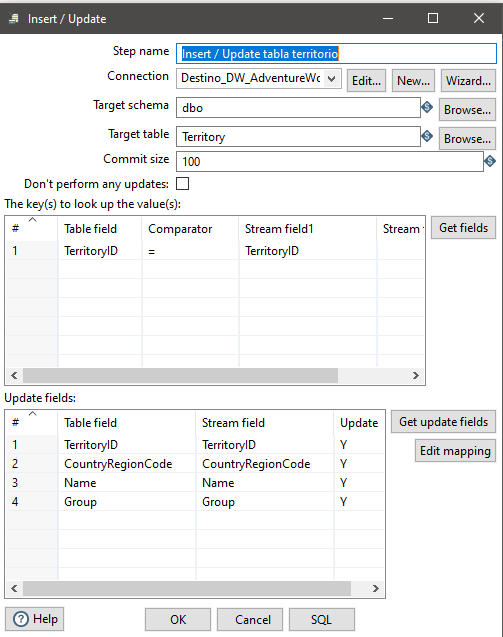
#### Configuración Input Table Input



#### Configuración Transform Value Mapper

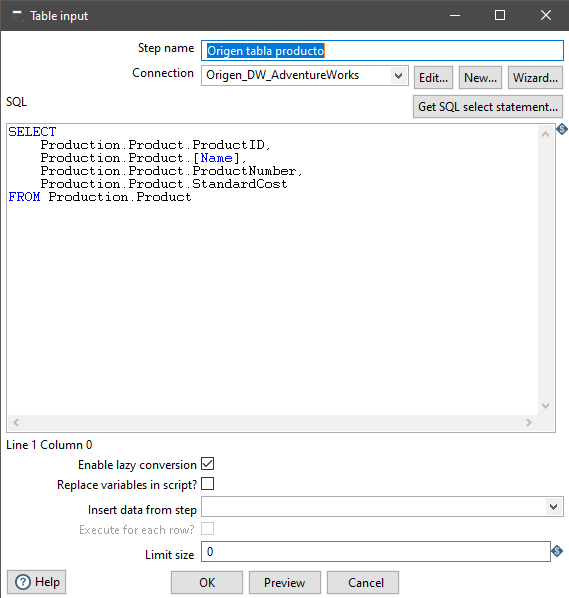


#### Configuración Output Insert/Update

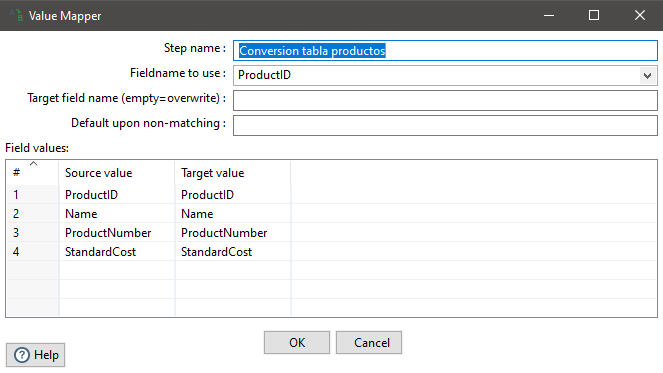


### Transformación Productos

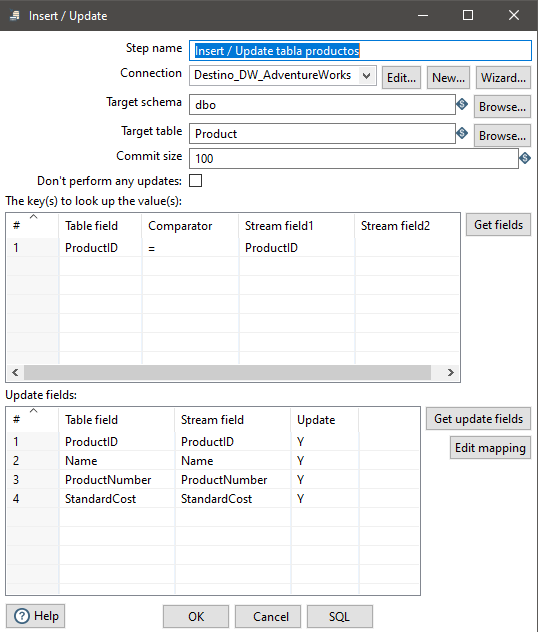
#### Configuración Input Table Input



#### Configuración Transform Value Mapper

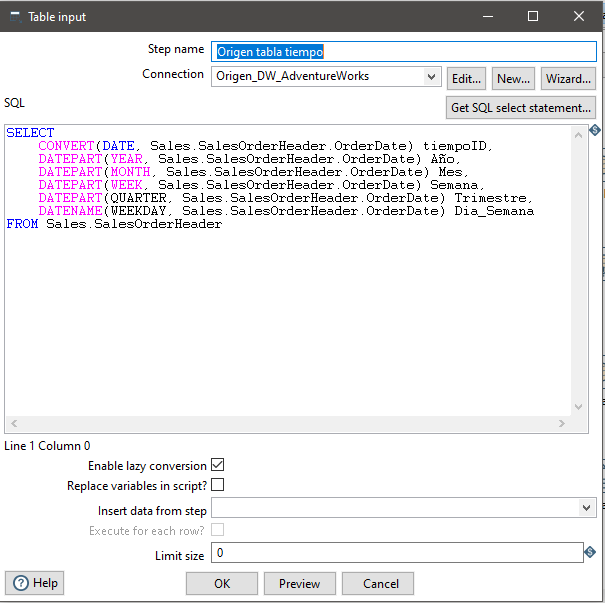


#### Configuración Output Insert/Update

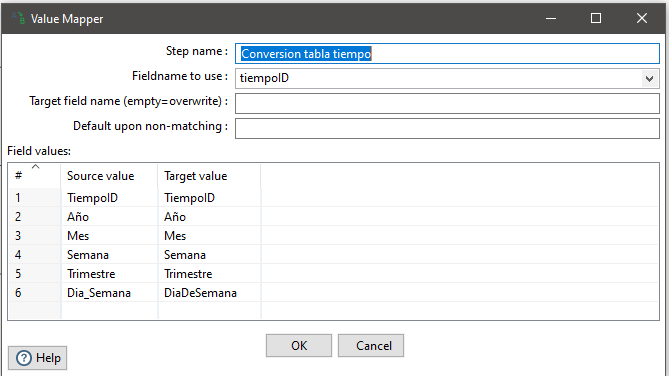


### Transformación Tiempo

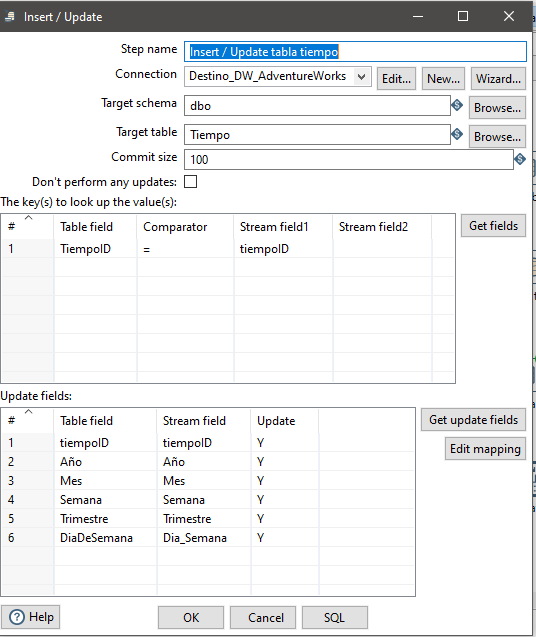
#### Configuración Input Table Input



#### Configuración Transform Value Mapper



#### Configuración Output Insert/Update

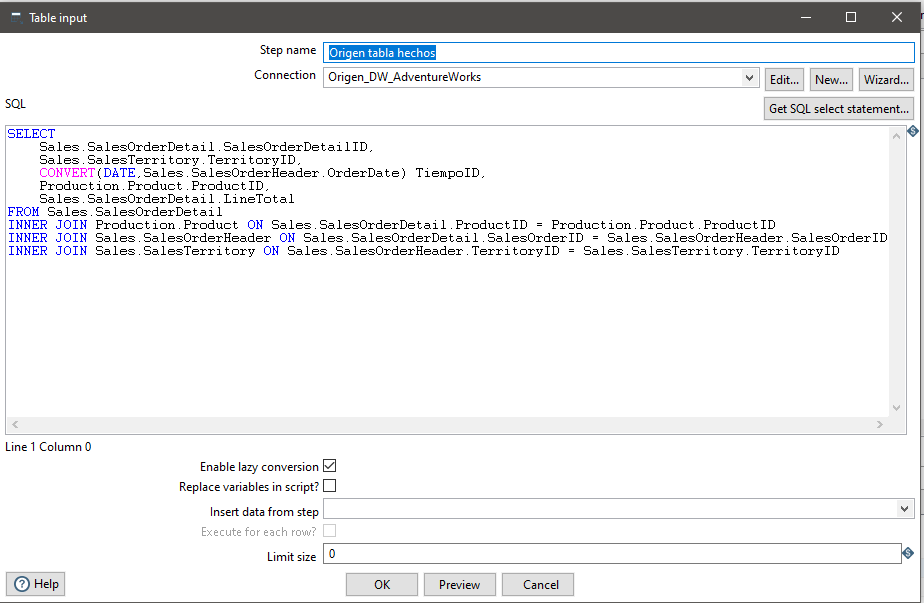


## **Transformaciones para las tablas de hechos**

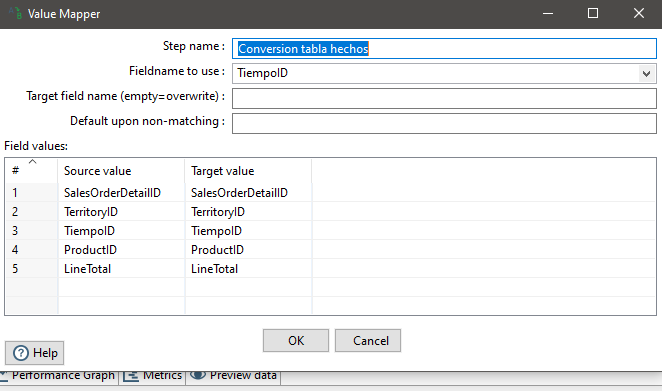
Para esta se crea otra nueva pestaña de transformaciones

### Transformación Tiempo

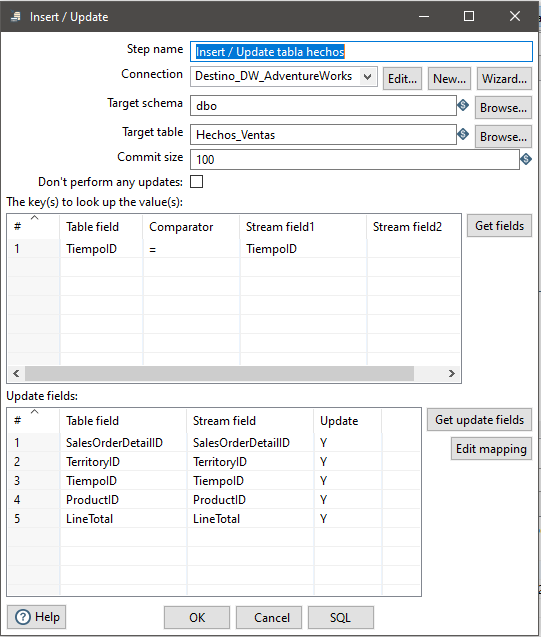
#### Configuración Input Table Input



#### Configuración Transform Value Mapper



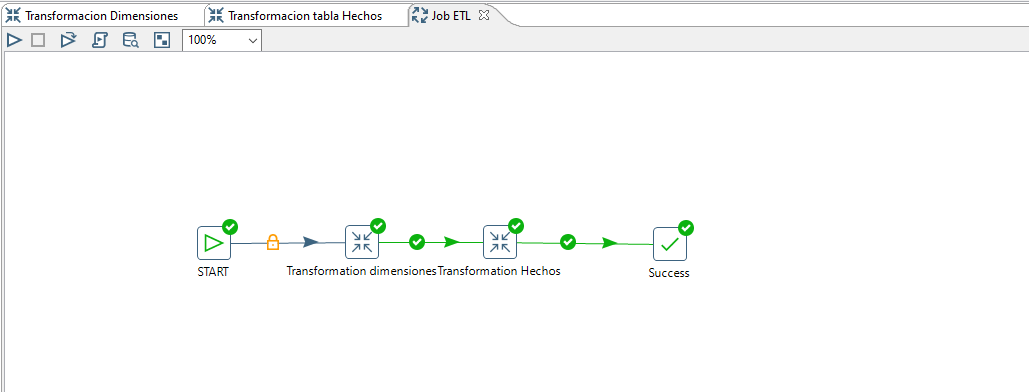
#### Configuración Output Insert/Update



## **JOB**

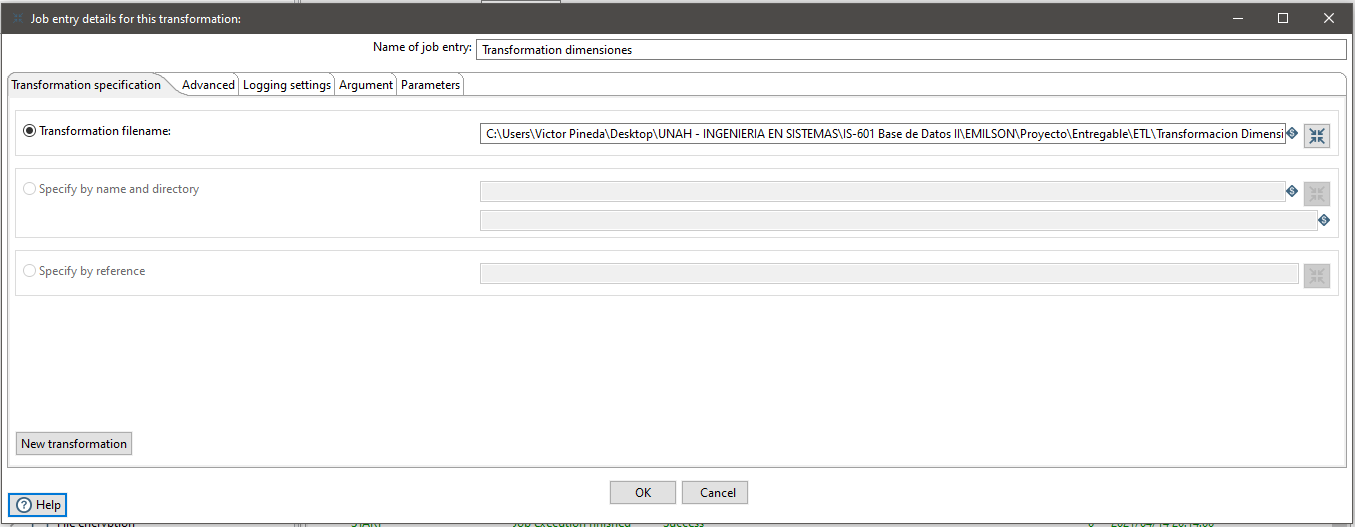
Este Job servirá para llenar la base de datos OLAP.

Para este Job se utilizarán solamente cuatro componentes todos de la pestaña general utilizaremos Start, dos Transformaciones y por último un Succes.



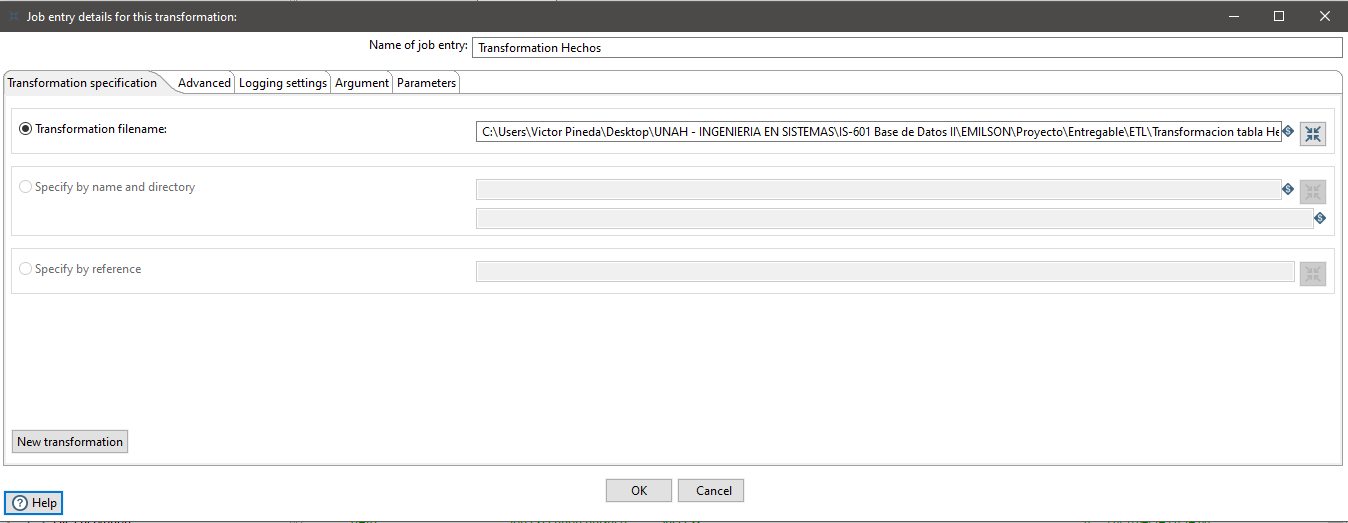
#### Configuración Transformación de dimensiones

Solamente buscamos donde guardamos las transformaciones de dimensiones y ponemos la dirección.



#### Configuración Transformación de dimensiones

Solamente buscamos donde guardamos las transformaciones de hechos y ponemos la dirección.

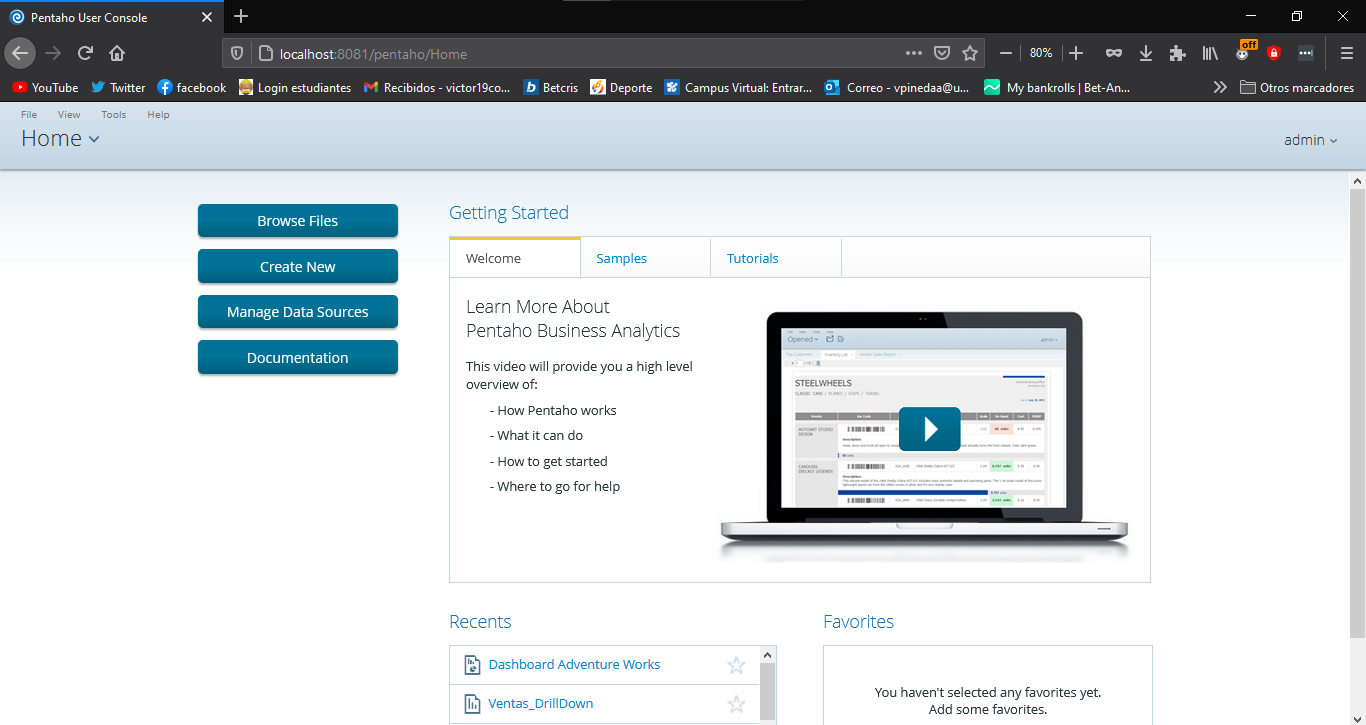


Los demás componentes no necesitan configuración alguna.

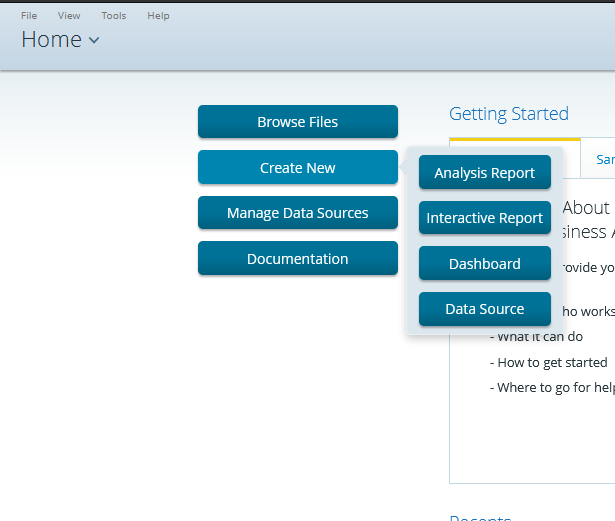
# Creación de Reportes

Entramos en **localHost:8081/pentaho/login** e iniciamos sesión como administradores.



Creamos un nuevo archivo en “Create New”

Damos click en “Analysis Report”

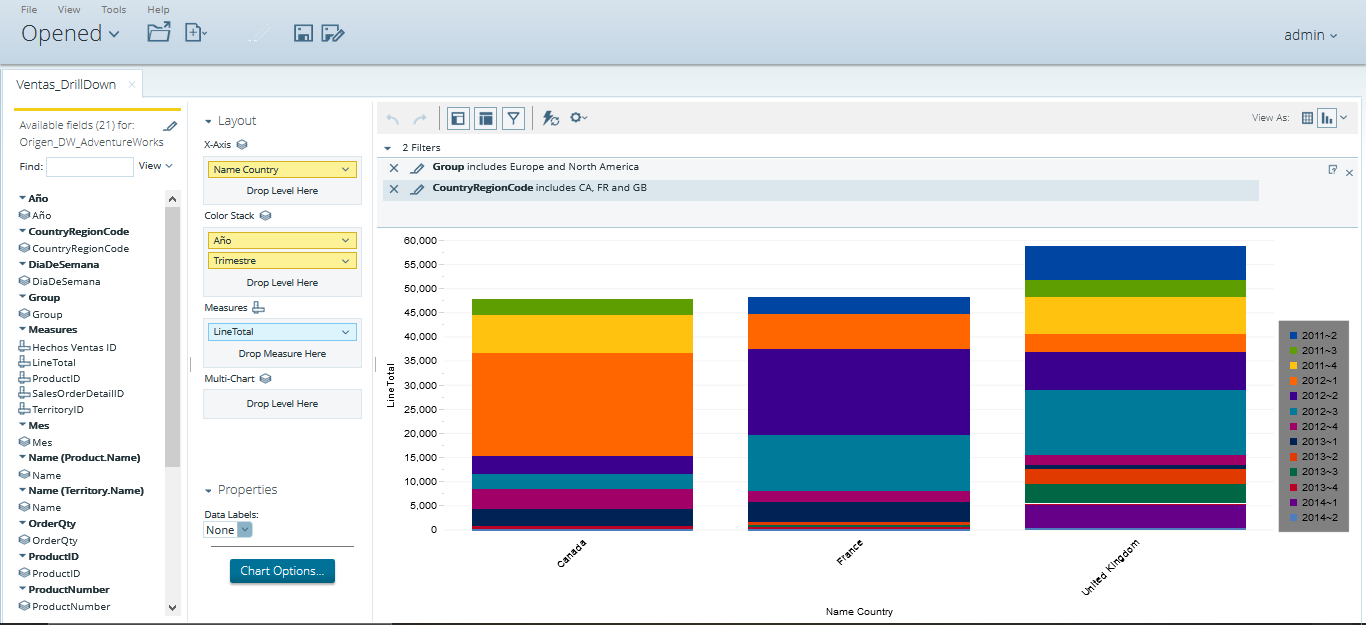


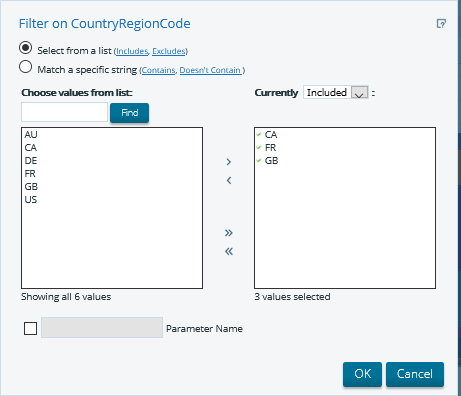
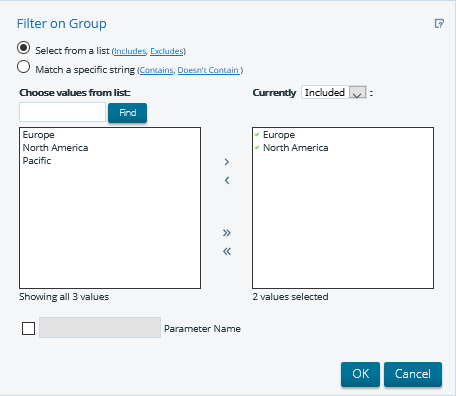
## **Reporte 1**

Lo configuramos de manera que en “Measures” ira nuestra métrica “LineTotal” que representa al total de ventas, en el eje X ira el nombre de país.

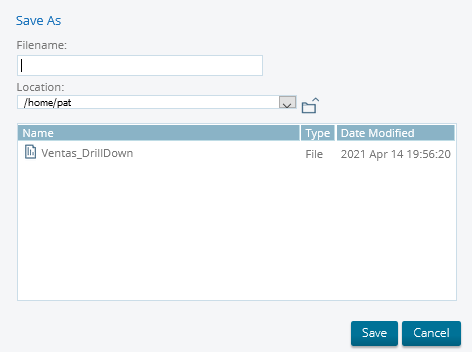
En el “Color Stack” pondremos como medidas el “trimestre” y “año”.

Lo filtramos por su grupo de territorio y el código de región.





Lo guardaremos en el directorio /home/pat para que se le asigne a este usuario.

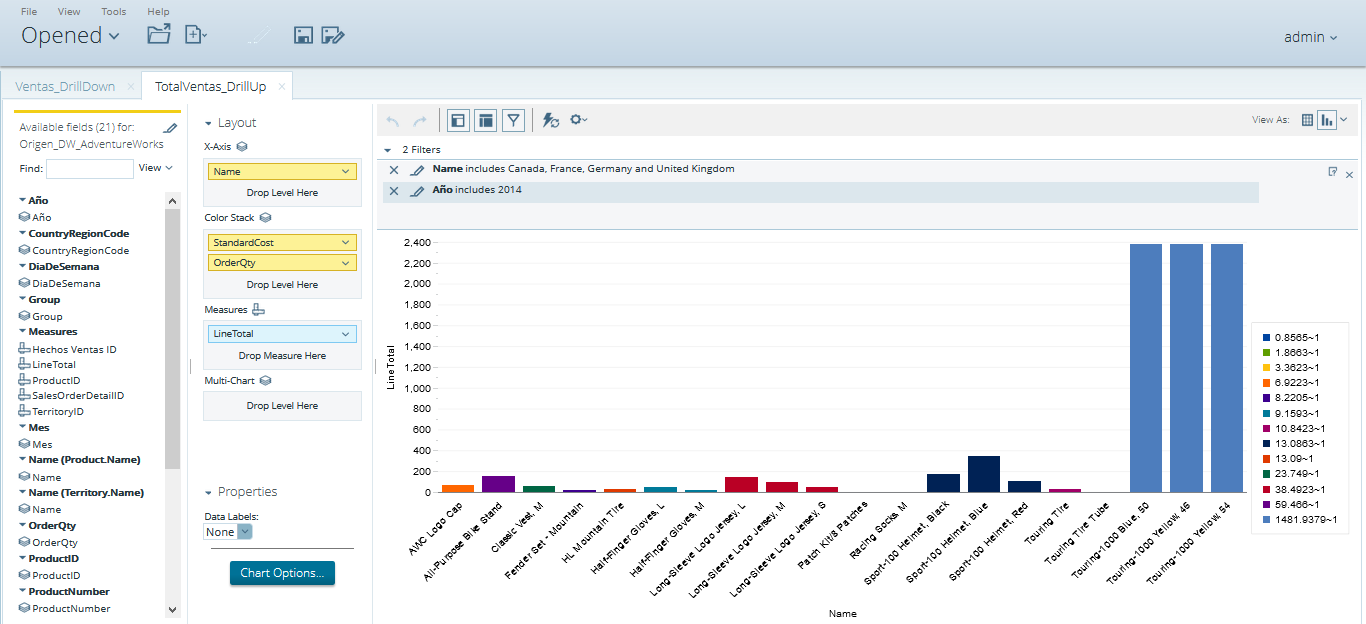
****

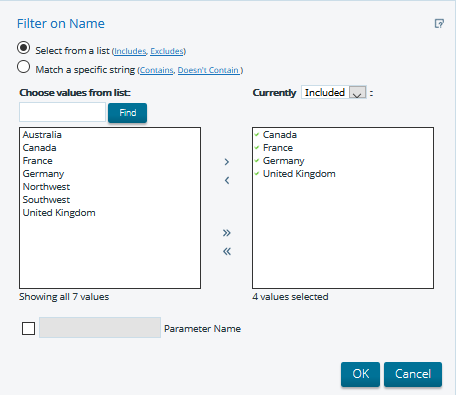
## **Reporte 2**

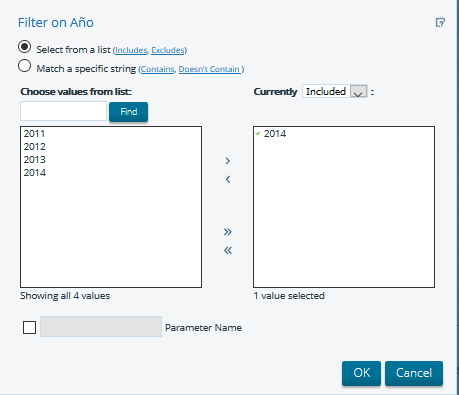
Lo configuramos de manera que en “Measures” ira nuestra métrica “LineTotal” que representa al total de ventas, en el eje X ira el nombre de producto.

En el “Color Stack” pondremos como medidas el “Standard Cost” y “OrderQty”.

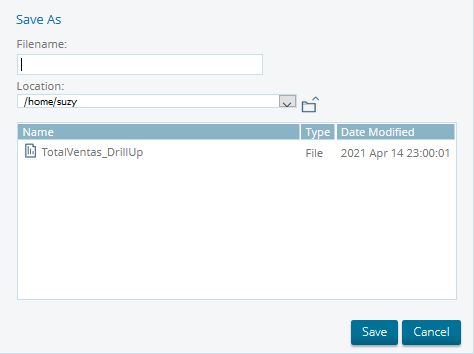
Lo filtramos por su nombre de país y año.







Lo guardaremos en el directorio /home/suzy para que se le asigne a este usuario.

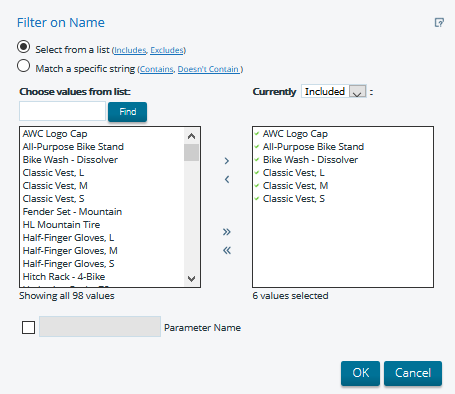
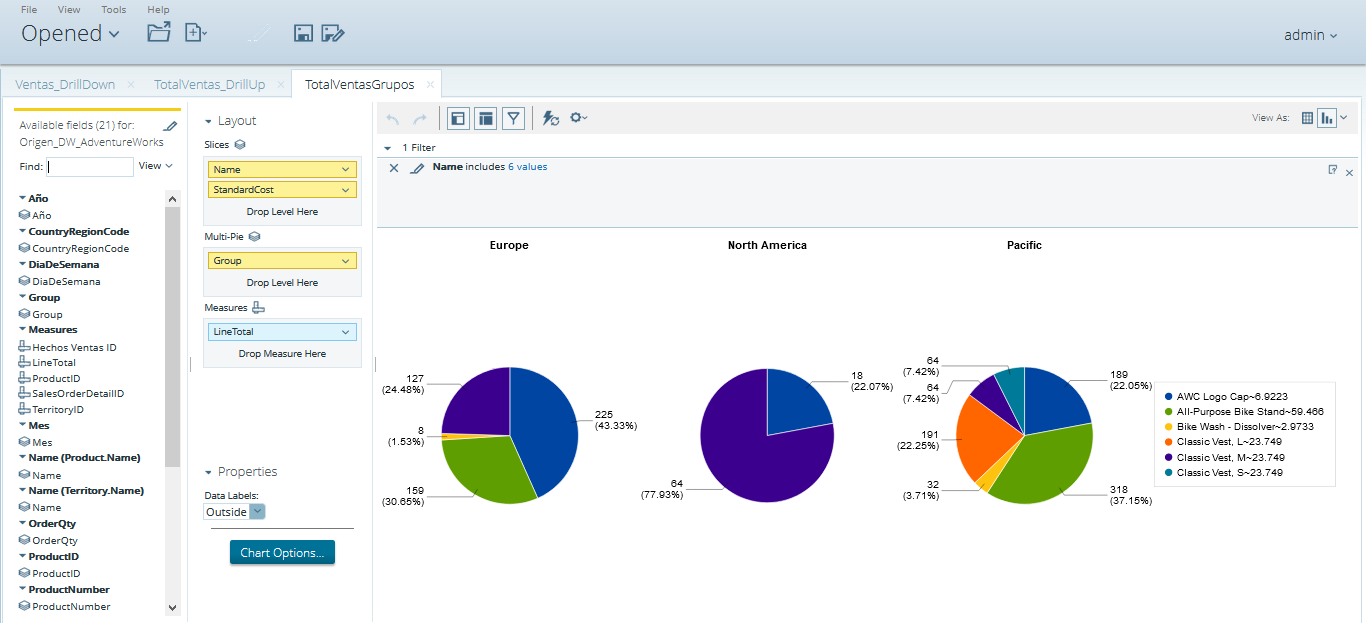
****

## **Reporte 3**

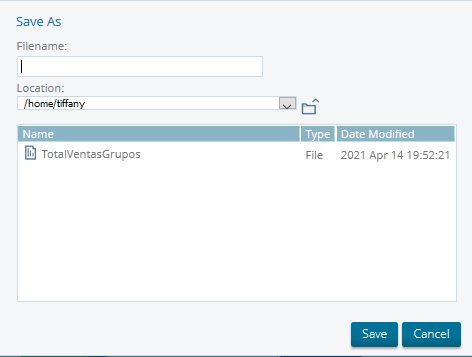
Lo configuramos de manera que en “Measures” ira nuestra métrica “LineTotal” que representa al total de ventas, en el eje X ira el nombre de producto y standard cost.

En el “Color Stack” pondremos como medidas el “Group”.

Lo filtramos por 6 nombres de productos.



Lo guardaremos en el directorio /home/tiffany para que se le asigne a este usuario.

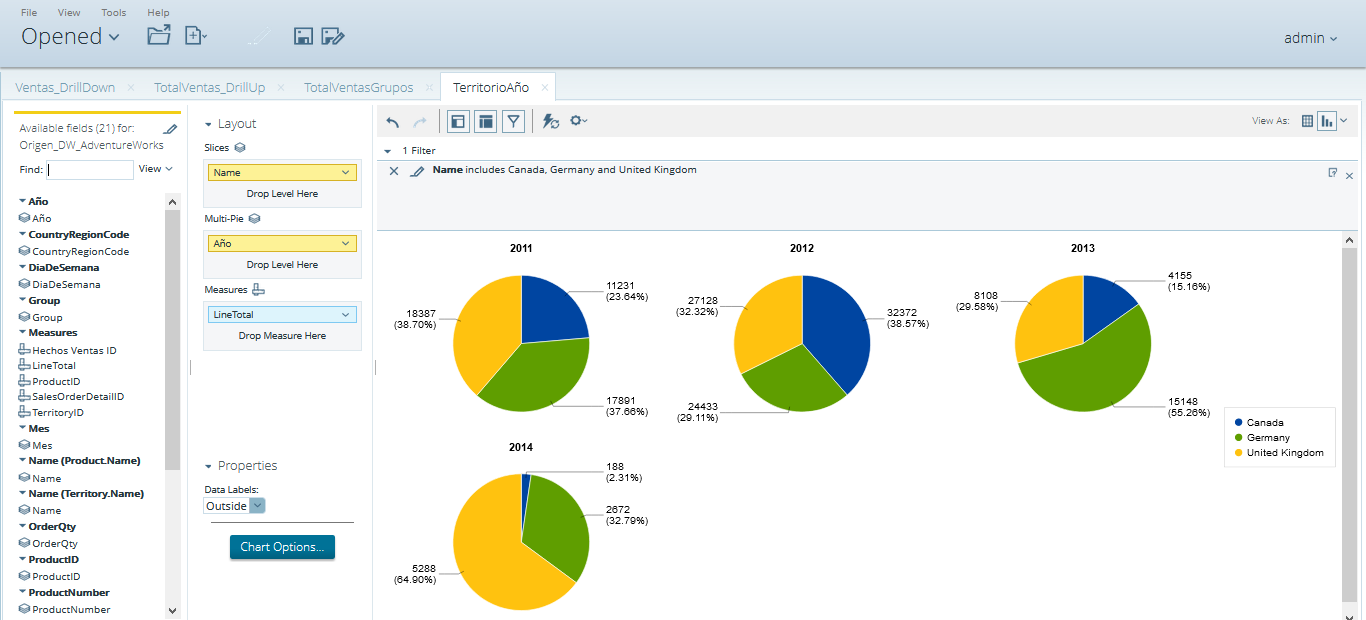


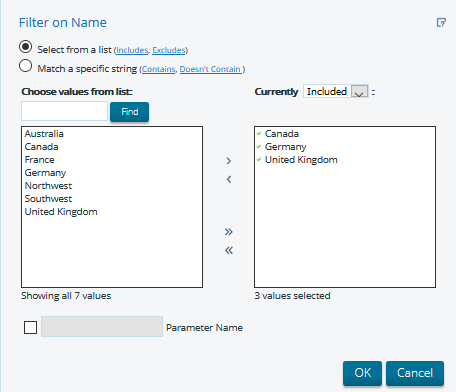
## **Reporte 4**

Lo configuramos de manera que en “Measures” ira nuestra métrica “LineTotal” que representa al total de ventas, en el eje X ira el nombre de país.

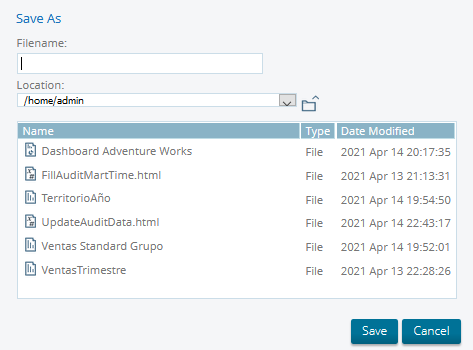
En el “Multi Pie” pondremos como medidas el “año”.

Lo filtramos por países.



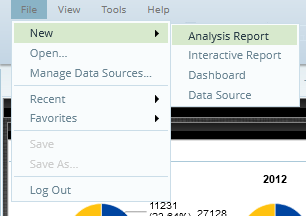


Este último será guardado en la ubicación de administración /home/admin.



# Creación de Dashboard

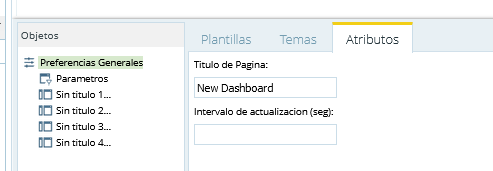
Para crear el dashboard daremos click en nuevo y en dashboard.



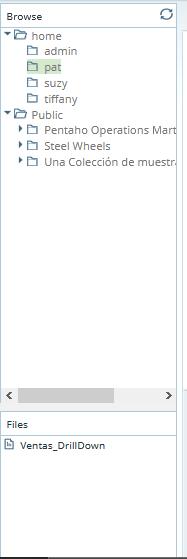
Escogemos la plantilla 2x2



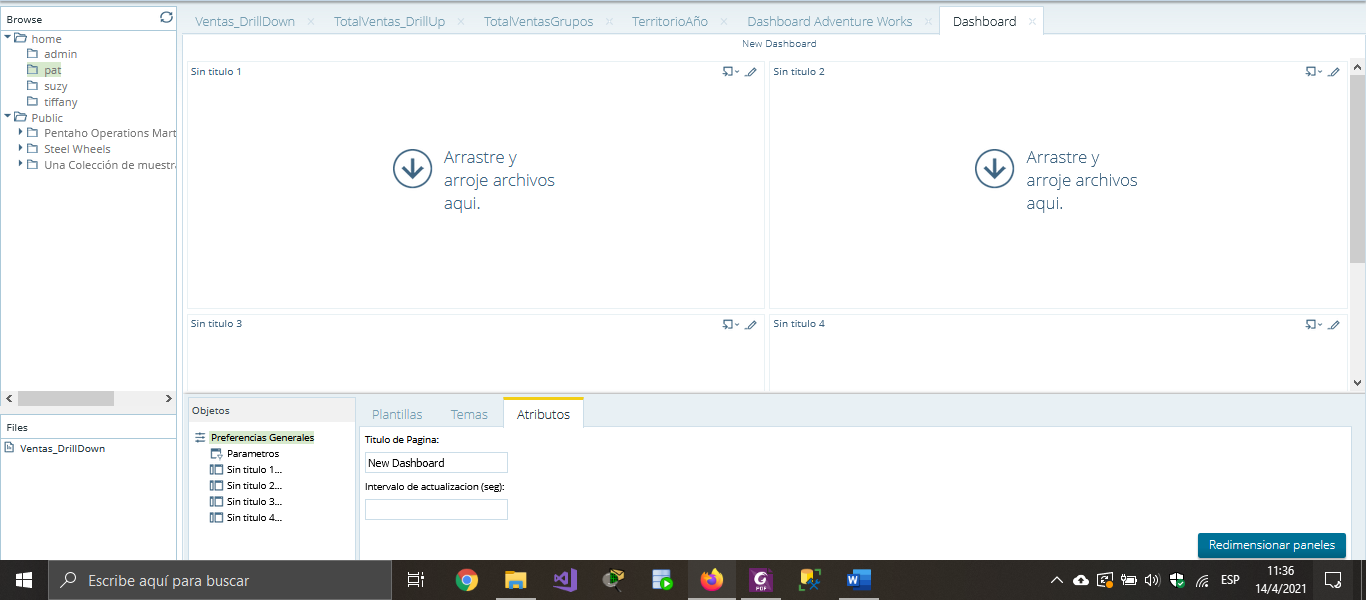
Podremos nombrar nuestro dashboard en Atributos.

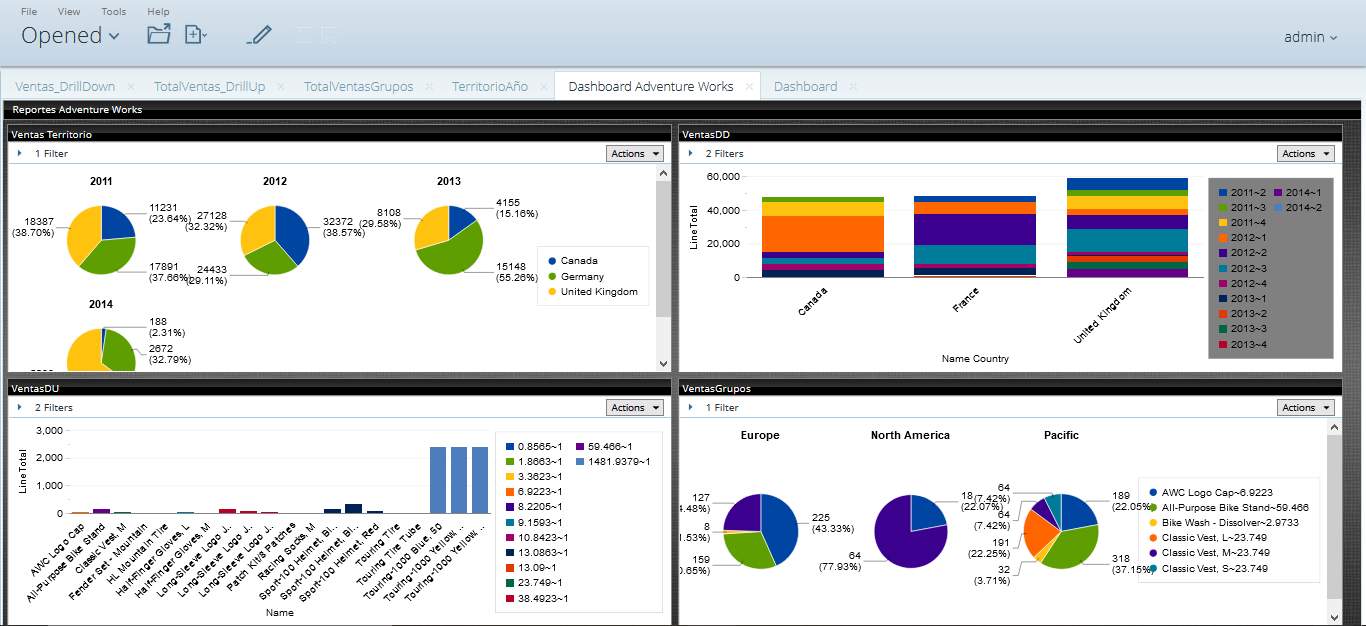


En la barra lateral iremos viendo que salen nuestros reportes creados con anterioridad.



Iremos agregando y arrastrando cada uno de nuestros reportes en cada cuadro que tendremos en el panel principal.



El resultado final será así, con nuestros cuatro reportes desplegados en un dashboard.

# Conclusiones

En conclusión, podemos llegar a decir que para una empresa en la actualidad debe aplicar un datawarehouse para buscar un mejor desarrollo, explotar toda la información de sus distintos departamentos, sacar el mejor provecho y preparar estrategias de manera correcta e integra, la mejor manera de aplicar el datawarehouse adecuado es conocer los requerimientos del negocio y hacer un estudio correcto y extenso de las fuentes que poblaran de datos el mismo.

# Recomendaciones

Podemos recomendar a las empresas que no tengan miedo a la inversión que puedan llegar hacer para que alguien haga este tipo de investigaciones para sus empresas ya que pueden ser de gran utilidad para analizar estrategias empresariales.

Recomendamos analizar bien la base de datos OLTP para asignar bien los datos a utilizar en la base de datos OLAP ya que aquí esta la base principal de la investigación.

Cuando logramos identificar bien las dimensiones y métricas lo demás se ira facilitando mas siendo por decirlo así carpintería.

# Referencias

Microsoft Corporation. (s.f.). Precios de SQL Server. Recuperado 1 abril, 2020, de<https://www.microsoft.com/es-es/sql-server/sql-server-2017-pricing#ft4>

What is a Data Warehouse?. (2020). Retrieved 30 July 2020, from <https://www.oracle.com/database/what-is-a-data-warehouse/>

PowerData, G. (2020). Data Warehouse: todo lo que necesitas saber sobre almacenamiento de datos. Retrieved 30 July 2020, from <https://www.powerdata.es/data-warehouse>