**SALESIANOS**

**UNIVERSIDAD DON BOSCO**

**FACULTAD DE INGENIERÍA**

**DATAWARE HOUSE Y MINERIA DE DATOS**

**DESAFIO 2**

**CICLO 01-2021**

**GRUPO TEÓRICO**

01

**CARRERA:**

INGENIERÍA EN CIENCIAS DE LA COMPUTACION

**PRESENTADO POR:**

OLMEDO LOPEZ, JONATHAN ALEXIS OL190123

MIRANDA MARTELL, SALVADOR ALEJANDRO MM130477

**CATEDRÁTICO:**

KARENS MEDRANO

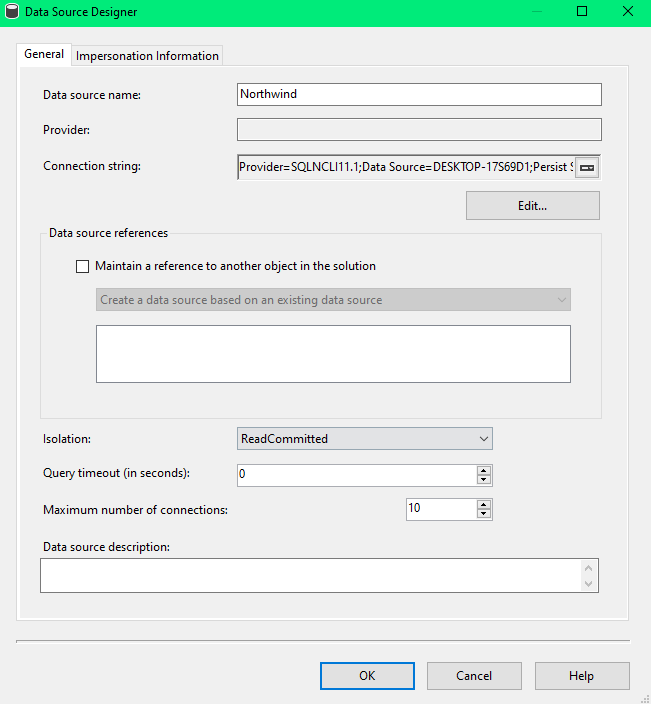
**06 DE JUNIO DEL 2021**

**SAN SALVADOR, EL SALVADOR, C.A.**

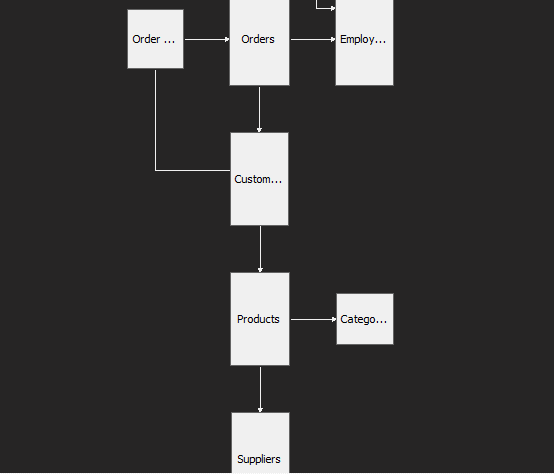
**ANALISIS**

1. **Elaboración de un cubo en donde se muestren los siguientes campos, nombre y país del cliente, apellido del empleado, país del proveedor, y nombre y categoría del producto.**

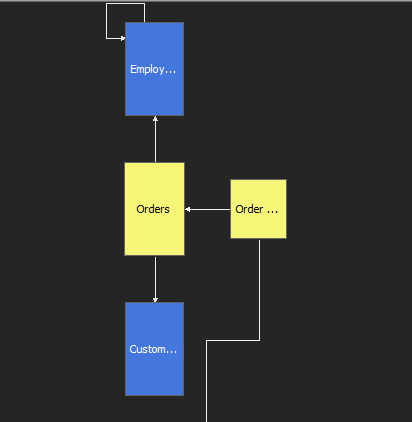
Conexión de visual studio a SQL.

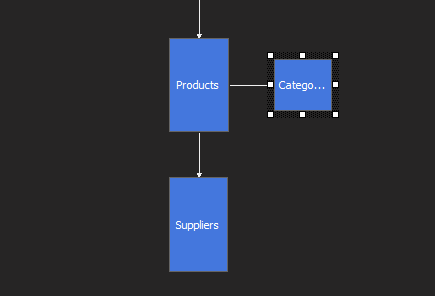


Vista de origines de datos con las tablas a filtrar.

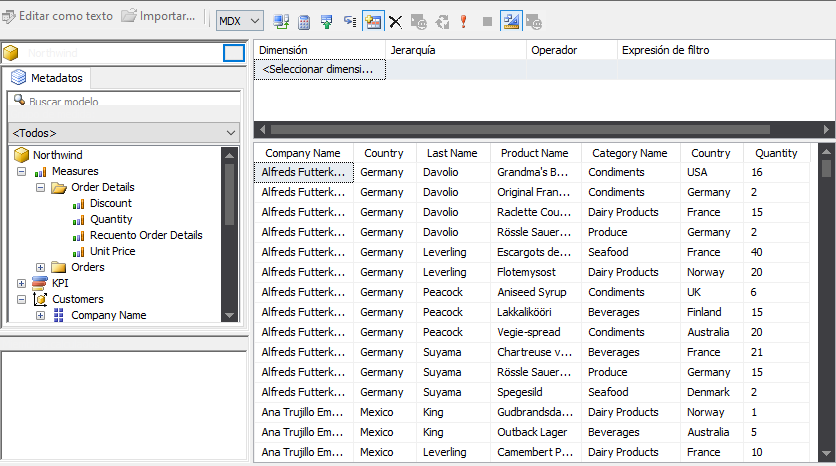


Cubo realizado.

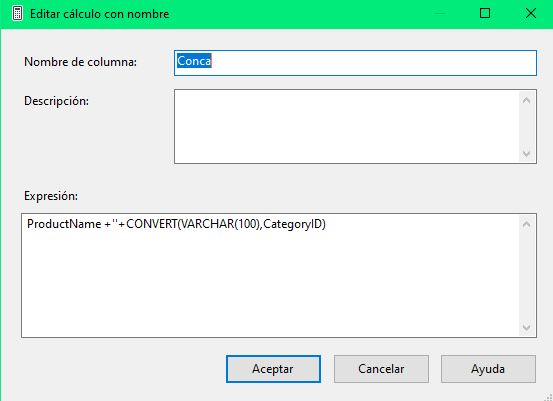




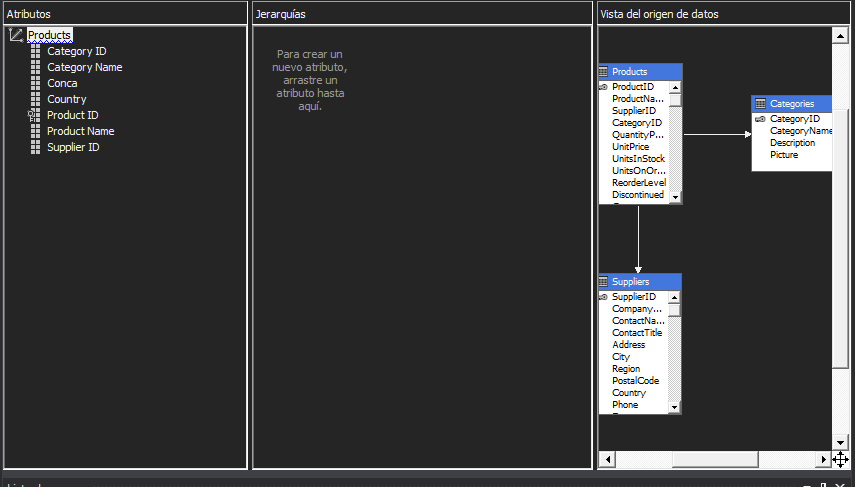
Datos mostrados en el cubo.



Campo concatenando la categoría del producto con el nombre del producto.

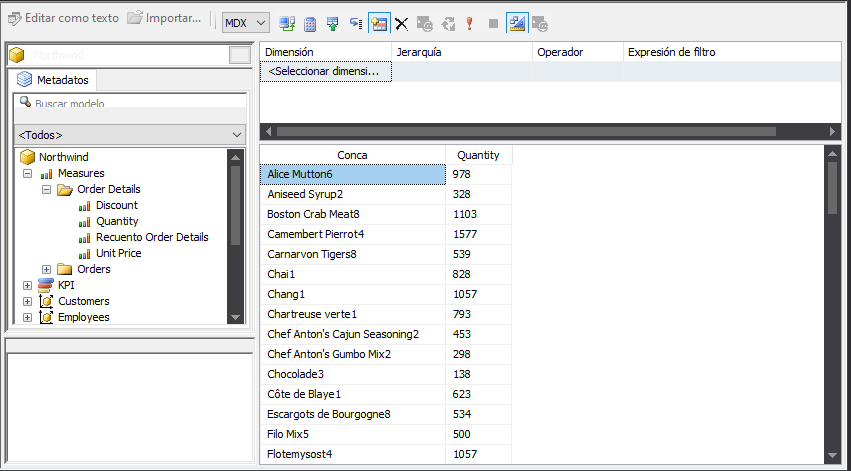


Se crea el campo concatenando la categoría del producto con el nombre del producto.



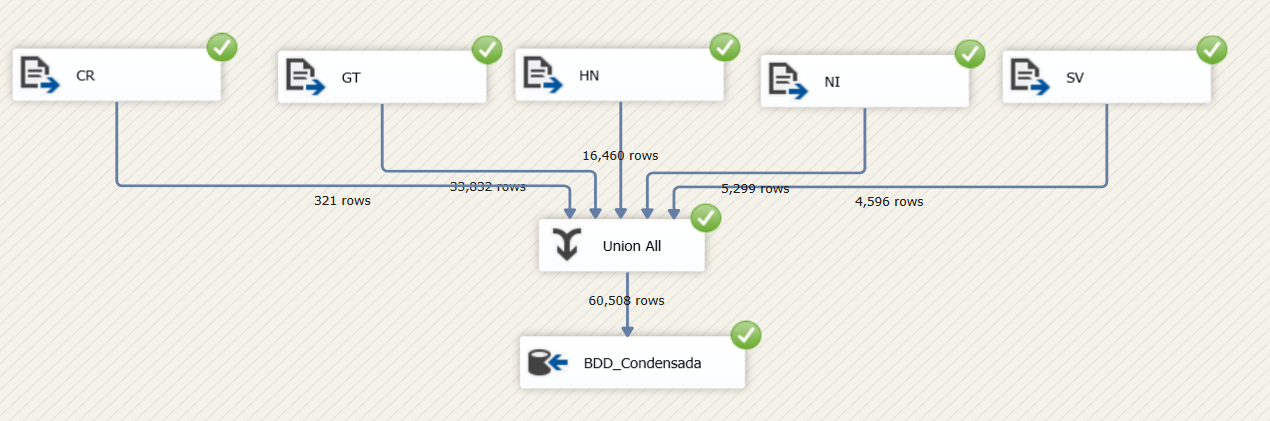
Luego se agrega el campo concatenado a los atributos de Products.

Por último, se ejecuta el cubo con el campo concatenado y se puede visualizar el nombre del producto junto con el ID de la categoría.



1. **Elabore un cubo en donde pueda visualizar el análisis que han realizado del archivo seleccionado. Que criterios tomar para el análisis, será responsabilidad de cada uno ver qué información es relevante para el análisis.**

Para el análisis de la información de la movilidad por país, se creó primero el ETL que cargaría los datos de todos los países centroamericanos en una tabla que tendría el condensado de toda esta información previa al análisis.



Debido a la falta de clave primaria, en el ETL, se crea la tabla con la información condensada añadiendo un campo extra que se denomina id, que es solo un entero con autocrecimiento que permite diferenciar cada uno de los datos contenidos en la tabla, aunque no tendrá propiamente un uso en el análisis, si es requerido para la creación del cubo.

CREATE TABLE "Covid\_movilidad" (

"id" int identity (1,1),

"country\_region\_code" varchar(50),

"country\_region" varchar(50),

"sub\_region\_1" varchar(50),

"sub\_region\_2" varchar(50),

"metro\_area" varchar(50),

"iso\_3166\_2\_code" varchar(50),

"census\_fips\_code" varchar(50),

"place\_id" varchar(50),

"date" varchar(50),

"retail\_and\_recreation\_percent\_change\_from\_baseline" varchar(50),

"grocery\_and\_pharmacy\_percent\_change\_from\_baseline" varchar(50),

"parks\_percent\_change\_from\_baseline" varchar(50),

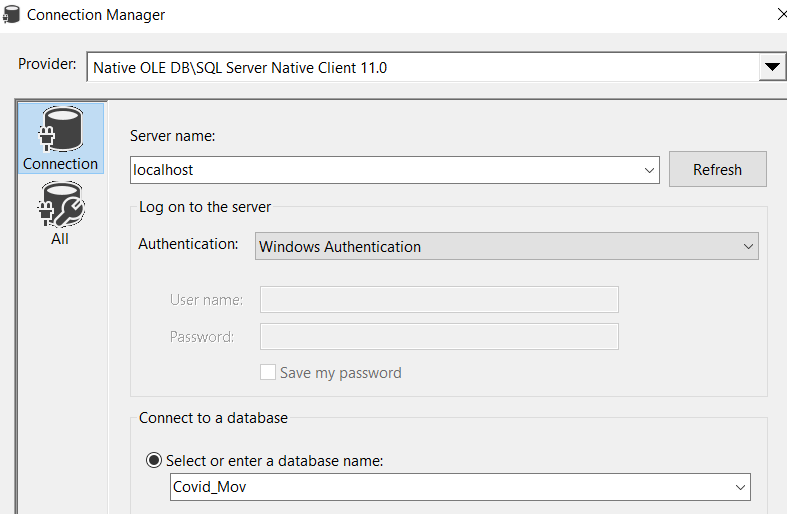
"transit\_stations\_percent\_change\_from\_baseline" varchar(50),

"workplaces\_percent\_change\_from\_baseline" varchar(50),

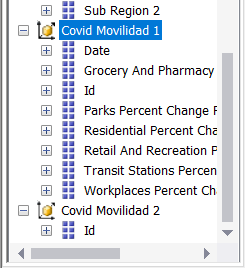
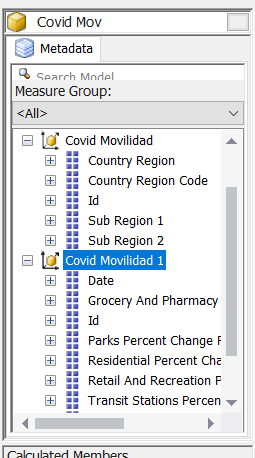
"residential\_percent\_change\_from\_baseline" varchar(50)

)

Una vez creado el ETL y cargada la información, se trabajó en el cubo OLAP por medio de Visual Studio. Se tomaría como Data Source la base de datos que contendría la tabla con todo el condensado de la información.

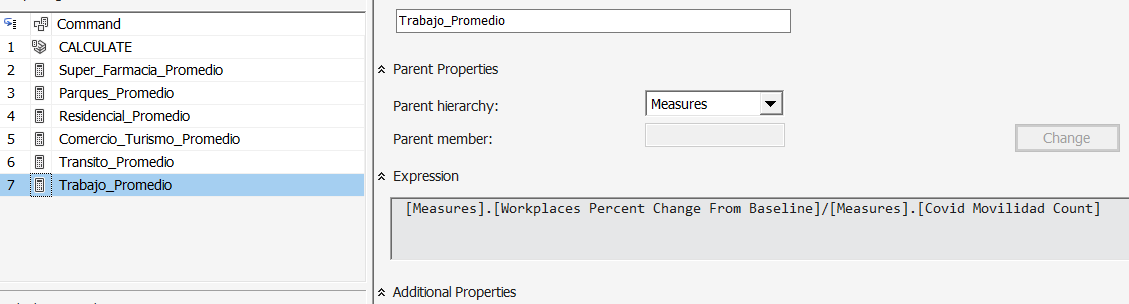


Una vez hecha la conexión, se trabaja en las dimensiones del cubo. Una tendría la información de fechas y datos estadísticos de la movilidad. Otra tendría los países y sus regiones de estar disponibles (no es el caso para Costa Rica). Y la otra, tendría el recuento de datos.

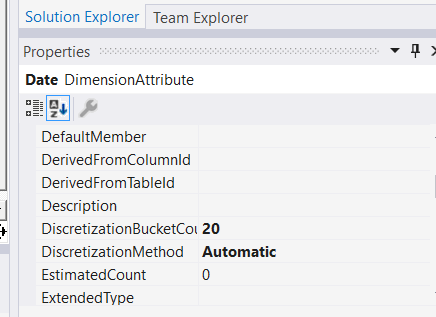


El análisis se enfocará en el promedio que cambió porcentualmente el movimiento. Estos archivos contienen información de cuánto es el cambio del movimiento habitual en un tipo de localidad durante la pandemia de COVID 19 comparado con los datos históricos de dichas fechas. Un dato de 0 implicaría que no ha existido ningún cambio y loa afluencia de la población es exactamente igual a sus registros históricos. Si este fuera positivo, significa un incremento, y si es negativo, conlleva una reducción en la afluencia de gente teniendo siempre el parámetro de 0 como la normalidad.

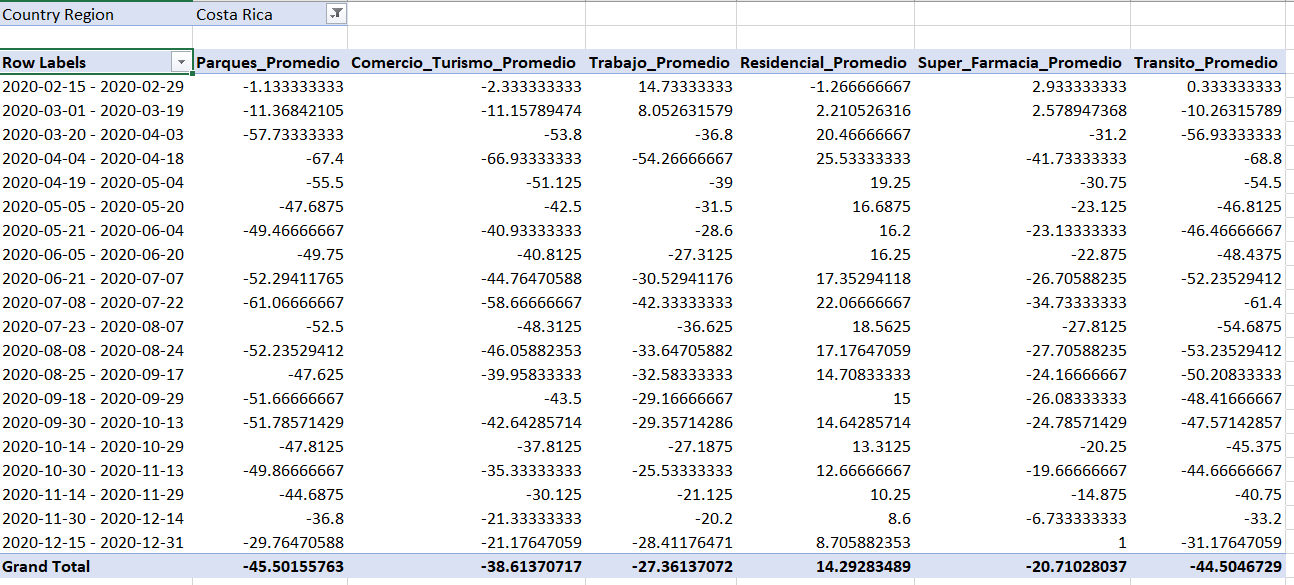
Se requirió añadir miembros calculados al cubo para poder realizar el calculo de los datos ya que la tabla incluye los datos por día, así que para facilitar el análisis, se añadió el calculo promedio del cambio porcentual para cada una de los distintos tipos de localidades.



Para que este cálculo funcione, es requerido que se tome más de una sola fecha por bloque, por lo que se configuró la fecha para que se muestre solo en 20 bloques en lugar de mostrarse de manera individual los aproximadamente 300 días que están contenidos; los datos finalmente se mostrarán como un promedio de cada 15 días, que al ser 15 días el tiempo que solía tomarse para aprobar cada cuarentena y medida en cada país, iba a permitir ver el impacto en cada bloque de tiempo y su varianza respecto a otros bloques.



Ya una vez estructurado el cubo y comprobando que cada medida y cálculo funcione a la perfección, se pone a funcional el cubo. Para propósitos de facilidad en la filtración de datos y acomodarlos de manera más conveniente para su análisis, se ocupa la opción de exportar a Excel el cubo. Ya en Excel, siendo una tabla pívot que puede modificarse su contenido con total libertad, se procede a analizar los datos.



* **Parques, comercio-turismo y tránsito:** Se observa que en Marzo empezó la reducción de afluencia, llegando a su pico en los primeros día de abril, lo que nos puede decir que existió un cierre paulatino y que llegó a ser total en los primeros días de Abril. Se mantuvo una porcentaje de afluencia negativa mayor al 45% hasta la última quincena de noviembre, donde empieza a verse un incremento, siendo un indicador de la reapertura de estas actividades en Costa Rica.
* **Trabajo:** En el caso de los lugares de trabajo, existen solamente dos picos que sobresaltan, uno de -54% en la primer quincena de Abril y otro en el periodo del 8 al 22 de Julio de -42%, el resto del tiempo, se mantuvo entre -20% y -33%, lo que indica que los lugares de trabajo no tuvieron un tratamiento de cierra total a diferencia de parques, centros comerciales, centros turísticos y las vías de tránsito. En estos picos negativos donde la afluencia es menor, podemos deducir que son los tiempos coincidentes con las cuarentenas estrictas impuestas por el gobierno costarricense cuando los casos de COVID-19 iban al alza.

Ahora bien, es de resaltar que no existe un incremento en la afluencia en las últimas fechas del año a diferencia de otras localidades. Esto a causa tanto las vacaciones del mes de Diciembre y que muchos puestos de trabajo en Costa Rica han decidido mantener la modalidad de trabajo en casa, por lo cual ya no se requiere de la presencia de las personal en los puestos de trabajo como si era requerido antes.

* **Supermercados y Farmacia:** Es de conocimiento general que estos lugares jamás fueron cerrados en ningún país del mundo debido a ser esenciales para el abastecimiento de alimentos, artículos de primera necesidad y medicinas. Sin embargo, era normal que familias enteras realizarán juntas las compras en estos lugares. Al observar una reducción al 45-55% negativo, se puede deducir que aunque nunca fueron cerrados estos lugares, si hubo una limitante de las personas que podían ingresar por familia, llegando a reducirse este número a solo una persona por familia. Nuevamente, en la segunda quincena de noviembre, se verifica que estas restricciones se fueron removiendo al incrementar el porcentaje de afluencia de un -45% a -31% para las fechas de final de año.
* **Residencial:** Al ser la mayor medida de prevención el pedir a la población que se quedara en casa al tiempo que todos los comercios y lugares no esenciales cerraban de manera física, el porcentaje de gente que se mantuvo en lugares residenciales se incrementó hasta un máximo de 25% en los tiempos de mayor restricción. Cabe recalcar que estos incrementos no son tan significativos como la reducción en otras localidades debido a que en general, son solo unas horas al día en que la población permanece fuera de casa, siendo por razones laborales o de estudio las razones por las cuales se mantienen fuera. Quitando estas horas laborales y el estudio fuera de casa, es un aproximado de 8 horas extra dentro de casa, y no necesariamente aplica para toda la familia debido al desempleo de algunos miembros, o que simplemente son personas que se dedican al hogar, por lo tanto, solo se incrementa la afluencia durante ese porcentaje del día que se mantenía vacío el hogar.