Universidad Don Bosco



Docente: Ing. Alexander Alberto Sigüenza

Materia: Datawarehouse y Minería de Datos

Presentado por:

Ruiz Saz, Carlos José RS181977

Rosales Mendoza, Patrick Ernesto RM181976

Porcentaje realizado:100%

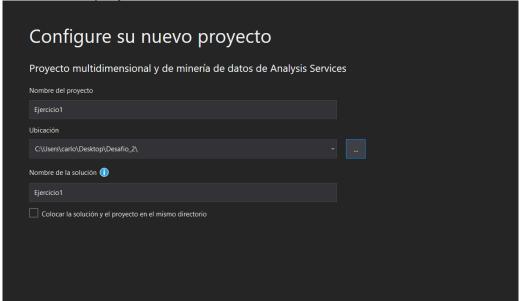
Ciclo: 02-2020

Contenido

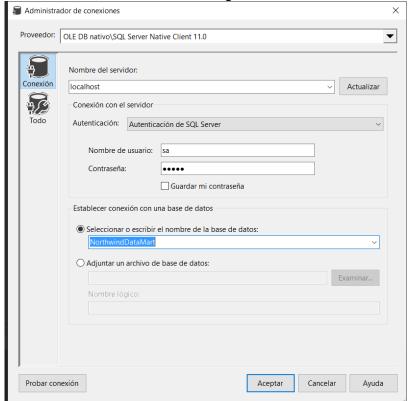
Ejercicio 1	3
Ejercicio 2	
Fase 2: Creación de cubo y análisis de datos	
rase 2. Creacion de cubo y analisis de datos	

Ejercicio 1

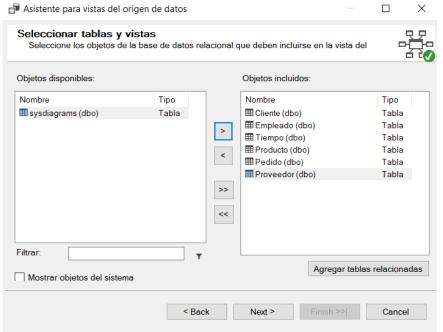
Creamos un proyecto:



Creamos la conexión con el origen de datos:

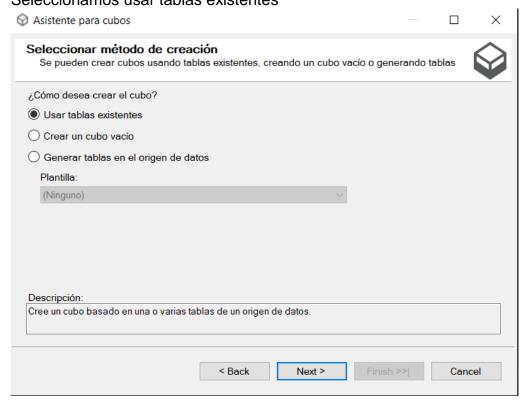


Creamos la vista del origen de datos, seleccionando las siguientes tablas:

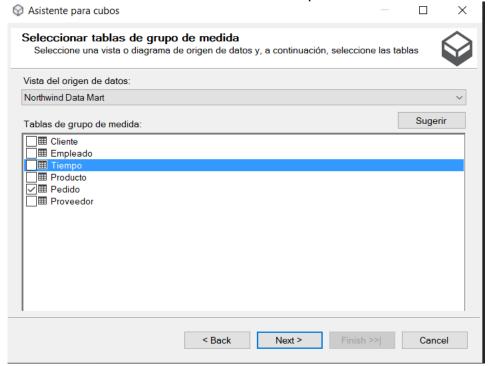


Vista del origen de datos ■ Proveedor ම id nombreCompania pais ■ Pedido productoID Empleado Ⅲ Producto clienteID empleadoID ම id ළම jd tiempoID apellido nombre proveedorID informe categoria Nombre | Categ... precio . cantidad III Tiempo III Cliente ම id ම id fecha nombre ciudad anio trimestre pais mes dia

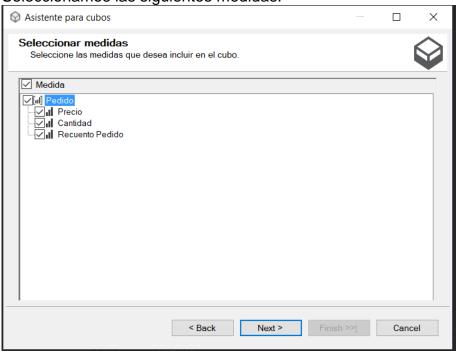
Creación del cubo: Seleccionamos usar tablas existentes



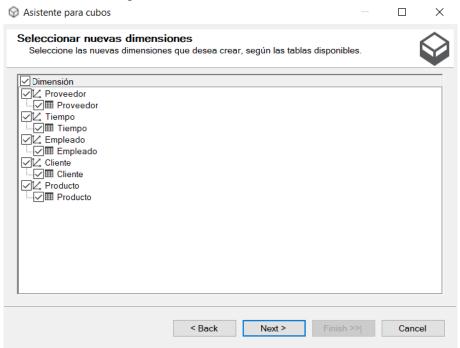
Nuestra tabla de hechos debe ser la tabla pedido.



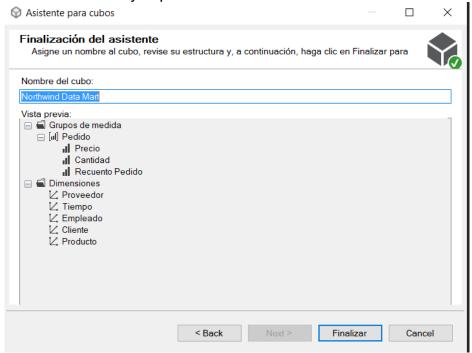
Seleccionamos las siguientes medidas:

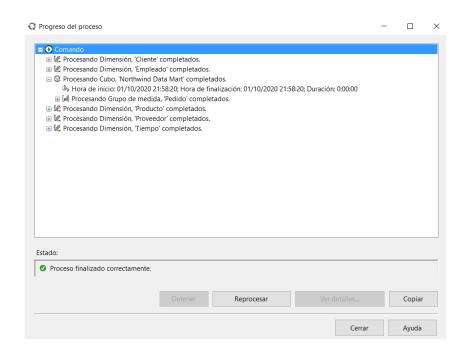


Tendremos las siguientes dimensiones:



Creamos el cubo y lo procesamos:





Modificaremos la dimensión Proveedor y deberá quedar así:



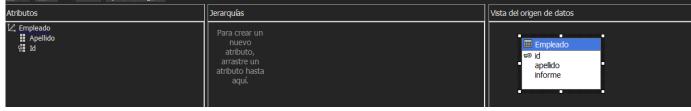
Modificaremos la dimensión Cliente y deberá quedar así:



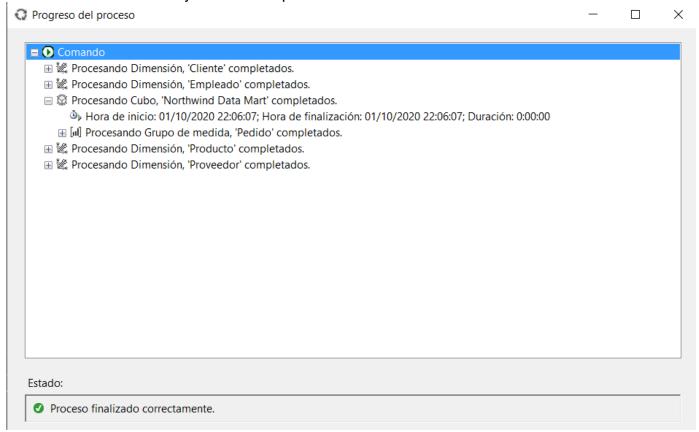
Modificaremos la dimensión producto y deberá quedar así:



Modificaremos la dimensión empleado y deberá quedar así:



Guardamos los cambios y volvemos a procesar el cubo



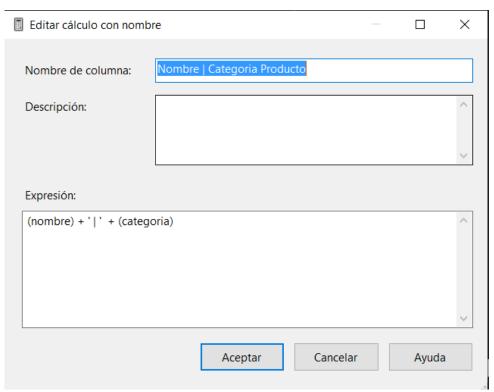
Exploramos el cubo y mostramos los siguientes datos:

- Nombre y País del cliente
- Apellido del empleado
- País del proveedor
- Nombre y Categoría del producto

•		•				
Nombre	Pais	Apellido	Pais	Nombre	Categoria	Cantidad
Alejandra Camino	Spain	Davolio	Australia	Perth Pasties	Meat/Poultry	10
Alejandra Camino	Spain	Davolio	Canada	Tourtière	Meat/Poultry	5
Alejandra Camino	Spain	Davolio	Germany	Nord-Ost Matjeshering	Seafood	10
Alejandra Camino	Spain	Peacock	Brazil	Guaraná Fantástica	Beverages	6
Alejandra Camino	Spain	Peacock	France	Camembert Pierrot	Dairy Products	10
Alejandra Camino	Spain	Peacock	Germany	Nord-Ost Matjeshering	Seafood	7
Alejandra Camino	Spain	Peacock	Italy	Ravioli Angelo	Grains/Cereals	2
Alejandra Camino	Spain	Peacock	UK	Teatime Chocolate Bisc	Confections	1
Alejandra Camino	Spain	Peacock	USA	Steeleye Stout	Beverages	4
Alexander Feuer	Germany	Buchanan	Australia	Vegie-spread	Condiments	6
Alexander Feuer	Germany	Buchanan	Finland	Lakkalikööri	Beverages	10
Alexander Feuer	Germany	Buchanan	France	Raclette Courdavault	Dairy Products	12
Alexander Feuer	Germany	Buchanan	Italy	Mozzarella di Giovanni	Dairy Products	30
Alexander Feuer	Germany	Leverling	Australia	Pavlova	Confections	20
Alexander Feuer	Germany	Leverling	Canada	Tarte au sucre	Confections	20
Alexander Feuer	Germany	Leverling	Netherla	Zaanse koeken	Confections	12
Alexander Feuer	Germany	Peacock	Italy	Gorgonzola Telino	Dairy Products	10

En este caso el valor medible fue el número que producto vendidos.

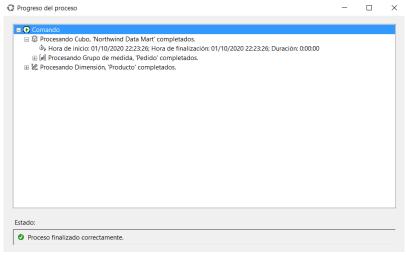
Ahora vamos a crear un campo, en el cual concatenaremos la categoría del producto y el nombre del producto.



Modificamos la dimensión producto, ya que ahora le agregaremos como atributo el campo calculado.



Guardamos, procesamos el cubo y lo exploramos.



Al explorar y usar el campo calculado, podemos observar el siguiente formato:

(Nombre del producto | Categoria del producto)

Nombre Categoria Producto	Cantidad
Alice Mutton Meat/Poultry	800
Aniseed Syrup Condiments	328
Boston Crab Meat Seafood	1053
Camembert Pierrot Dairy Products	1391
arnarvon Tigers Seafood	431
Chai Beverages	738
hang Beverages	987
Chartreuse verte Beverages	713
hef Anton's Cajun Seasoning Condiments	397
nef Anton's Gumbo Mix Condiments	298
hocolade Confections	138
ôte de Blaye Beverages	524
scargots de Bourgogne Seafood	504
o Mix Grains/Cereals	396
otemysost Dairy Products	1033
- Harak I Parisa Dua di Caka	643

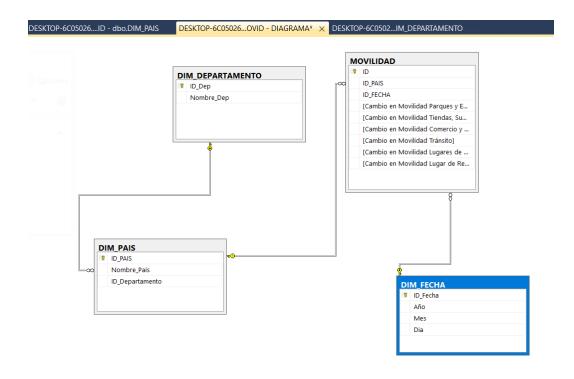
En este caso el valor medible es la cantidad de productos, agrupados por nombre y categoría del producto.

Ejercicio 2

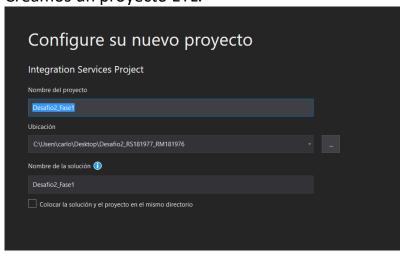
Este ejercicio, se va a dividir en 2 fases:

- ETL para cargar los datos en una base de datos
- Creación del cubo para el análisis de la información

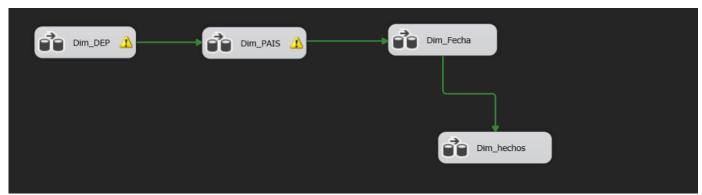
Diagrama de la base de datos a utilizar:



Fase 1 : ETL Creamos un proyecto ETL.

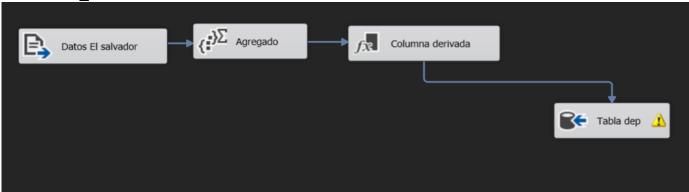


El flujo de control del ETL sería de la siguiente forma:



Cada una de las tareas del flujo de datos tiene el objetivo de llenar una tabla

Tarea Dim_DEP:

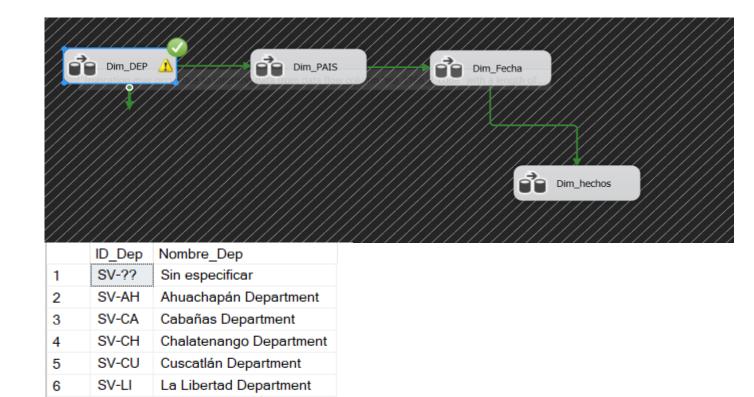


En este caso obtenemos los datos, los agrupados por código del departamento y nombre del departamento y hay ciertos registros que no tienen departamento, por lo tanto, la columna derivada se hace lo siguiente:

Nombre de columna	Columna derivada	Expresión	Tipo de datos	Lo
iso_3166_2_code	Reemplazar 'iso_316	iso_3166_2_code == "" ? "SV-??" : iso_3166_2	cadena [DT_STR]	5
sub_region_1	Reemplazar 'sub_regi	sub_region_1 == "" ? "Sin especificar" : sub_re	cadena [DT_STR]	5

En caso de estar vacío le asignamos un código SV-?? y en el nombre del departamento colocamos sin especificar luego esto pasa a la base de datos.

Resultados:



Tarea Dim Pais.

SV-MO

SV-PA

SV-SA

SV-SM

SV-SO

SV-SS

SV-SV

SV-UN

SV-US

7

8

9

10

11

12

13

14 15 Morazán Department

Santa Ana Department

San Miguel Department

Sonsonate Department

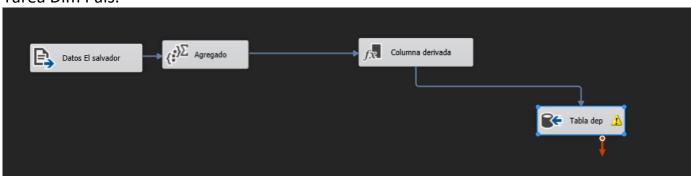
San Salvador Department

San Vicente Department

La Unión Department

Usulután Department

La Paz Department



En este caso se hace la conexión, se hace el agrupado por código del país, nombre del país e ID del departamento que sería la clave foránea, y como en la tarea anterior, en columna derivada hacemos una condición para que todos los Id's que vayan vacíos ahora tengan como id SV-?? adicional a eso en el código del país se concatenan las últimas 2 letras del código del departamento y finalmente mandamos los datos a la base de datos.

Nombre de columna	Columna derivada	Expresión	Tipo de datos
iso_3166_2_code	Reemplazar 'iso_316	iso_3166_2_code == "" ? "SV-??" : iso_3166_2_code	cadena [DT_S
country_region_code	Reemplazar 'country	country_region_code + RIGHT(iso_3166_2_code,2)	cadena [DT_S

Resultados:



	ID_PAIS	Nombre_P	ID_Departame
1	SV	El Salvador	SV-??
2	SVAH	El Salvador	SV-AH
3	SVCA	El Salvador	SV-CA
4	SVCH	El Salvador	SV-CH
5	SVCU	El Salvador	SV-CU
6	SVLI	El Salvador	SV-LI
7	SVMO	El Salvador	SV-MO
8	SVPA	El Salvador	SV-PA
9	SVSA	El Salvador	SV-SA
10	SVSM	El Salvador	SV-SM
11	SVSO	El Salvador	SV-SO
12	SVSS	El Salvador	SV-SS
13	SVSV	El Salvador	SV-SV
14	SVUN	El Salvador	SV-UN
15	SVUS	El Salvador	SV-US



En este caso se hace la conexión, se agrupa por el campo de la fecha y en la columna derivada se hace lo siguiente:

Nombre de columna	Columna derivada	Expresión	Tipo de datos	Lo
year	<agregar colu<="" como="" td=""><td>RIGHT(date,4)</td><td>cadena Unicode [DT</td><td>4</td></agregar>	RIGHT(date,4)	cadena Unicode [DT	4
dia	<agregar colu<="" como="" td=""><td>LEFT(date,2)</td><td>cadena Unicode [DT</td><td>2</td></agregar>	LEFT(date,2)	cadena Unicode [DT	2
mes	<agregar colu<="" como="" td=""><td>SUBSTRING(date,4,2)</td><td>cadena Unicode [DT</td><td>2</td></agregar>	SUBSTRING(date,4,2)	cadena Unicode [DT	2
id	<agregar colu<="" como="" td=""><td>LEFT(date,2) + SUBSTRING(date,4,2) + RIGHT(</td><td>cadena Unicode [DT</td><td>8</td></agregar>	LEFT(date,2) + SUBSTRING(date,4,2) + RIGHT(cadena Unicode [DT	8

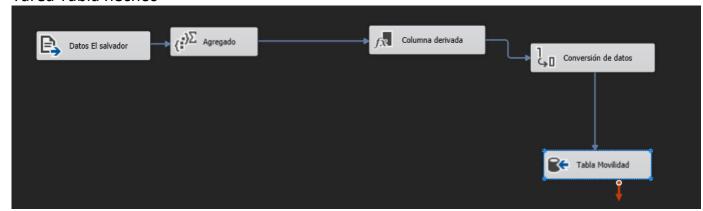
Cortamos la fecha para obtener el año, día, mes y para el id de la fecha concatenamos todo lo que hemos cortado, convertimos los datos a String y por último mandamos eso a la dimensión fecha de la base de datos.

Resultados:



En total serían 226 registros de fechas.

Tarea Tabla hechos

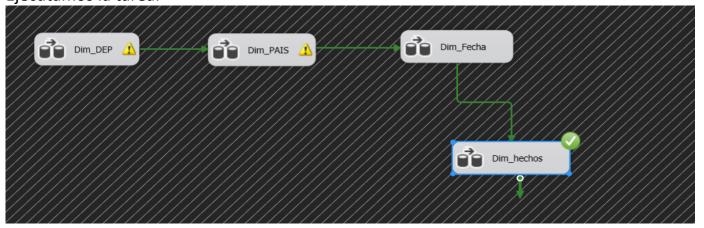


En este caso realizamos la conexión, y agrupamos por todos los porcentajes, código de país, código de departamento, y fecha.

En columna derivada haremos lo siguiente:

Nombre de columna	Columna derivada	Expresión	Tipo de datos	Lo
id_fecha	<agregar colu<="" como="" td=""><td>$LEFT(date,\!2) + SUBSTRING(date,\!4,\!2) + RIGHT($</td><td>cadena Unicode [DT</td><td>8</td></agregar>	$LEFT(date,\!2) + SUBSTRING(date,\!4,\!2) + RIGHT($	cadena Unicode [DT	8
country_region_code	Reemplazar 'country	country_region_code + RIGHT(iso_3166_2_cod	cadena [DT_STR]	5

Esto con el fin de obtener los id's de las dimensiones fecha y país. Luego convertimos id_fecha a string y mandamos los datos a la base de datos. Ejecutamos la tarea.



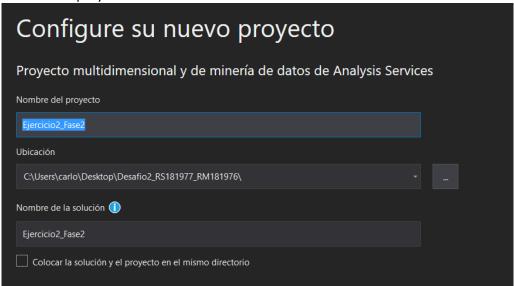
Resultados en la base de datos:

	ID	ID_PAIS	ID_FECHA	Cambio en	Cambio en Movilidad Tiendas, Supermercados y Farm	Cambio en Movilidad Comercio y Recrea	Cambio en Movilidad Trán	Cambio en Movilidad Lugares de Tra	Cambio en Movilidad Lugar de Reside
1	1	SVSS	07042020	-77	-57	-79	-71	-74	38
2	2	SV	19062020	-53	-30	-56	-62	-48	23
3	3	SVSA	10032020	-5	3	-1	-11	4	-1
4	4	SVSM	26092020	0	0	0	0	-21	8
5	5	SVMO	23082020	0	0	0	0	-28	0
6	6	SVSA	02052020	-63	-51	-69	-70	-49	26
7	7	SVLI	02052020	-69	-53	-75	-55	-55	30
8	8	SVLI	25052020	-71	-56	-78	-56	-65	38
9	9	SVLI	26032020	-75	-54	-81	-68	-71	40

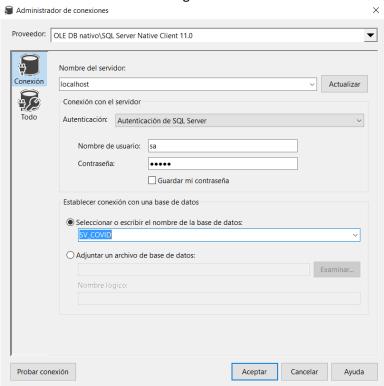
En el caso del id de la tabla hechos es un autoincrementable.

Con esto finalizaríamos la fase 1 del ejercicio 2

Creamos un proyecto.



Creamos la conexión con el origen de datos:



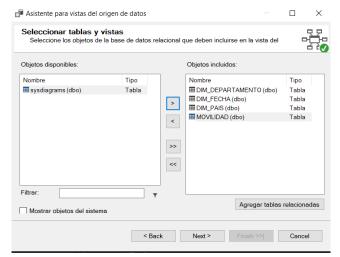
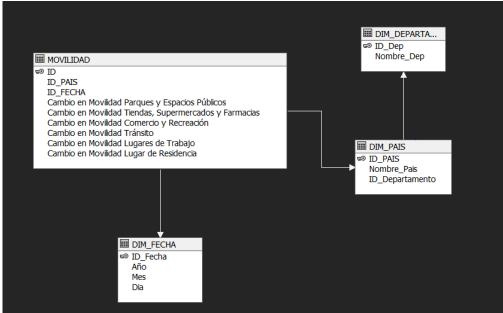
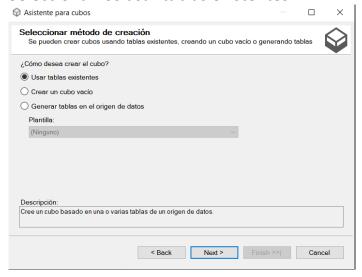


Diagrama de la vista el origen de datos



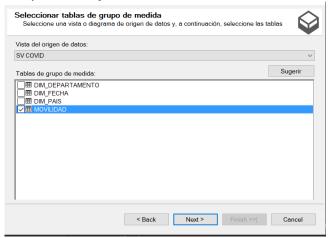
Creación del cubo:

Seleccionamos usar tablas existentes

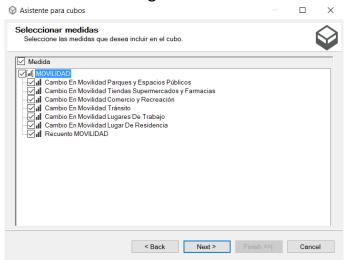


Nuestra tabla de hechos es la tabla de movilidad, ya que contiene los valores medibles de

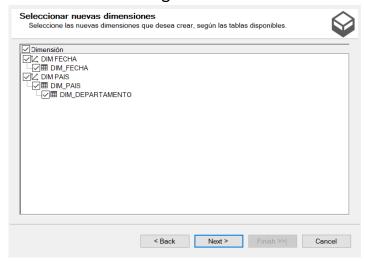
los porcentajes de movilidad:



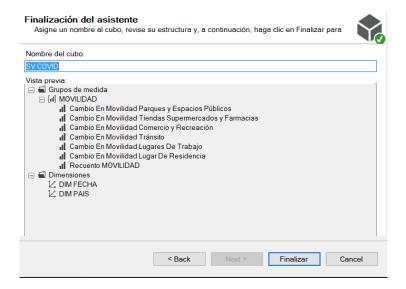
Seleccinamos las siguientes medidas:

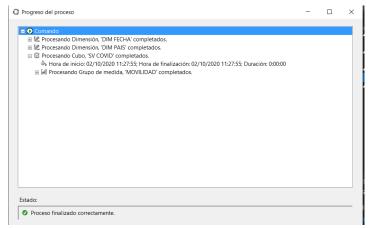


Seleccionamos las siguientes dimensiones:

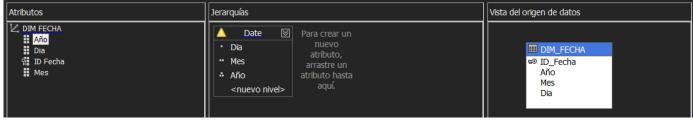


Damos clic en finalizar, creamos el cubo y lo procesamos:





Modificaremos la dimensión fecha y deberá tener los siguientes atributos adicionalmente contará con una jerarquía de fecha:



Modificaremos la dimensión país y deberá tener los siguientes atributos:



Procesamos el cubo.

■ O Comando					
⊞ ₩ Procesando Dimensión, 'DIM FECHA' completados.					
⊞ ﷺ Procesando Dimensión, 'DIM PAIS' completados.					
☐ 🔯 Procesando Cubo, 'SV COVID' completados.					
(a) Hora de inicio: 02/10/2020 11:32:39; Hora de finalización: 02/10/2020 11:32:39; Duración: 0:00:00					
⊞ [ul] Procesando Grupo de medida, 'MOVILIDAD' completados.					
Estado:					

Análisis a nivel de país

	•				
Nombre Pais	Cambio En Movilidad Comercio y Recreación	Cambio En Movilidad Lugar De Residencia	Cambio En Movilidad Lugares De Trabajo	Cambio En Movilidad Parques y Espacios	Cambio En Movilidad Tiendas Supermerc
El Salvador	-109647	26063	-126222	-123118	-54453
Cambio En Movilidad Tránsito					
-85729	9				

Se puede observar que a nivel de país se redujo bastante la movilidad por el hecho de la pandemia y lo podemos comprobar de acuerdo a los indicadores negativos que se nos presenta en la consulta.

En cambio, podemos ver un valor positivo que es el cambio en movilidad de lugar de residencia y aquí el valor es positivo debido a que las personas pasaron el mayor tiempo en su zona residencial.

Análisis a nivel de país por mes

Nombre Pais	Mes	Cambio En Movilidad Comercio y Recreación	Cambio En Movilidad Lugar De Residencia	Cambio En Movilidad Lugares De Trabajo	Cambio En Movilidad Parques y Espacios Públicos	Cambio En Movilidad Tiendas Supermercados		
El Salvador	02	-141	-80	1240	-454	56		
El Salvador	03	-11354	2300	-9600	-12886	-4574		
El Salvador	04	-24889	4985	-25269	-27082	-13269		
El Salvador	05	-26506	5039	-24446	-28550	-14432		
El Salvador	06	-20400	4285	-19620	-23198	-9935		
El Salvador	07	-17757	3879	-18254	-21232	-8516		
El Salvador	08	-8600	3221	-16745	-9716	-3783		
El Salvador	09	0	2434	-13528	0	0		
Cambi	io Er	n Movilidad Tránsito						

Cambio En Movilidad Tránsito
-155
-8457
-18086
-20291
-16959
-14677
-7104
0

Al hacer un análisis por mes podemos ver que justo en el mes de marzo, es decir cuando el virus del covid empezó a propagarse por Centroamérica hubo una baja en el valor de la

movilidad, esto por las cuarentenas obligatorias que ordenaba el gobierno, podemos decir que la baja empezó a darse a partir del mes de marzo, en este caso el motivo fue la evitar la propagación del virus.

Vemos esta reducción del valor entre los meses de marzo y agosto.

En el mes de septiembre podemos ver que según los indicadores 0, quiere decir que hemos vuelto a la "normalidad" en cuanto a la movilidad en ciertos lugares y actividades, a lo cual hay que ver que la movilidad en lugar de residencia se ha visto reducida, ya que han aumentado la movilidad en otros lugares.

Cabe destacar que el valor en movilidad de trabajo en el mes de septiembre, sigue estando muy inferior a un valor normal, esto nos puede indicar 3 situaciones:

- Desempleo: Muchas personas han sido afectadas por la crisis que ha generado el virus
- Home office: Ciertas empresas, se han adaptado a esta modalidad de trabajo, ya que en el tiempo en que el virus se propagaba ferozmente, se pudo ver una mejora en cuanto a la modalidad de trabajo remoto
- Capacidad de personal: Ciertas empresas decidieron no tener en sus oficinas al 100% de sus empleados, solamente iban a la oficina, los empleados que desempeñaban cargos en los cuales era fundamental su presencia en la institución.

Análisis a nivel de departamento Cambio En Movilidad Lugares De Trabajo Cambio En Movilidad Parques y Espacios Públicos Cambio En Movilidad Tiendas Supermercados Ahuachapán... -8139 38 -7552 -3283 Chalatenang... -8254 -7772 -7616 Cuscatlán D... -2233 -8192 -8063 La Paz Depa... -7965 -5776 -6701 La Unión De... -7901 -7874 -9858 -575 San Miguel D... -9518 -8478 -8247 -6330 San Salvado... -10101 5141 -10059 -10284 -6302 Santa Ana D... -9505 3973 -8626 -8922 -6066 Sin especificar -9879 -9412 -6275 -9065 Usulután De... -9457 -8211 -6961 -6022

Cambio En Movilidad Tránsito
-5959
0
-1383
-5726
-7959
-14624
0
0
-10202
-9924
0
-10352
-9984
-9616

En algunos departamentos, podemos ver un 0 en el indicador de la movilidad en cierta

actividad, esto quiere decir, que hubo departamentos en donde salir a cierto lugar, la movilidad siempre se mantuvo "normal".

Por ejemplos en la movilidad de supermercados:

Podemos ver como Chalatenango, Cabañas, Cuscatlán, Morazán y san Vicente vemos que estos tienen un indicador de 0, esto quiere decir que la movilidad ni se redujo, ni aumento, por lo tanto, como se dijo anteriormente la movilidad era "normal"