

Desafío Práctico DMD

Autor: BA1819 ,GM181938

2020

Universidad Don Bosco



Datawarehouse y Minería de Datos

Desafío práctico

Integrantes:

Bonilla Avilés, David Alejandro BA1819

Grande Menjivar, Willian Adonis GM181938

Catedrático: Ing. Alexander Siguenza

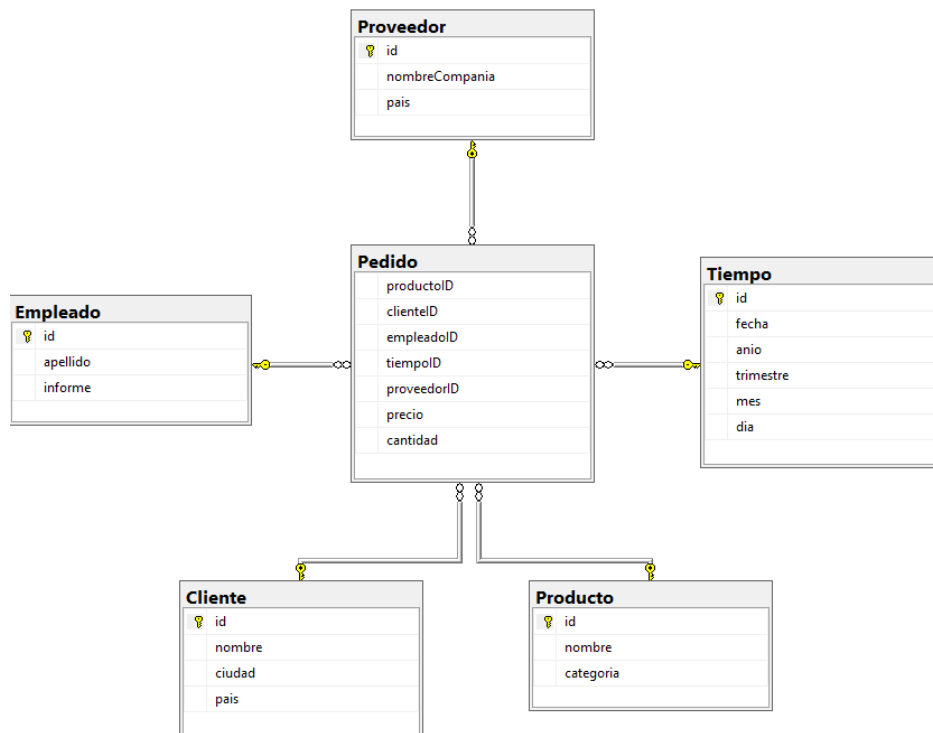
Soyapango, 04 de octubre del 2020

Contenido

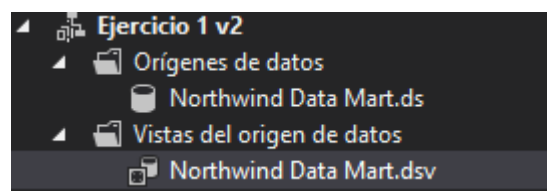
Ejercicio 1	4
Ejercicio 2	7

Ejercicio 1

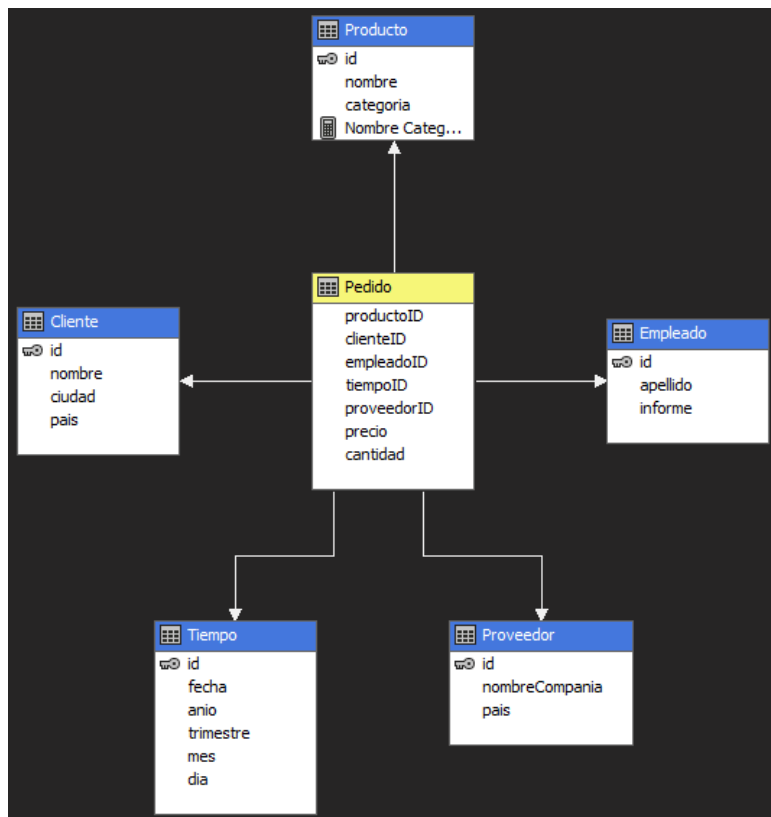
La solución del ejercicio 1, fue bastante directa ya que en el planteamiento presentaba los requerimientos claros de cada una de las partes. En primer lugar se tenía que importar la base de datos Northwind Data Mark.



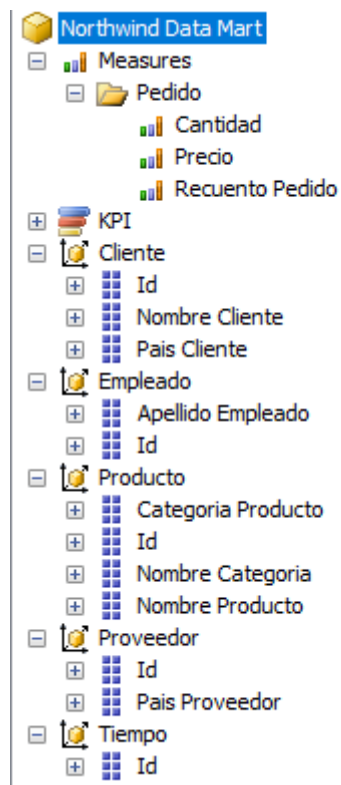
Después de importar la base de datos fue necesario crear la conexión a la base de datos con el integration services, y a su vez el crear una vista del origen de datos, donde se mostraban las tabla que se realizó con el proceso anterior.



Después se creo el cubo dejando como dimensiones Producto", " Empleado", " Cliente", " Tiempo" y "proveedor".



A su vez se modificaron los atributos de las dimensiones para colocarles los nombres correspondientes y realizar el nuevo campo que esta conformado por la concatenación del nombre y categoría del producto:



Es importante recalcar que para crear el atributo de la concatenación dentro de los atributos de las dimensiones del cubo, fue necesario agregar un calculo con nombre en la vista dentro de la tabla correspondiente.

Editar cálculo con nombre

Nombre de columna:

Nombre Categoría

Descripción:

Expresión:

nombre + ',' + categoria

De la siguiente manera quedaría la consulta de datos dentro del cubo, esto demuestra que se puede visualizar el nombre y país del cliente, el apellido del empleado, el país del proveedor, el nombre del producto, la categoría del producto, el campo concatenado del nombre y categoría del producto, que son los requerimientos solicitados.

Nombre Cliente	Pais Cliente	Apellido Empleado	Pais Proveedor	Nombre Producto	Categoría Producto	Nombre Categoría	Cantidad	Precio
Alejandra Cam...	Spain	Davolio	Australia	Perth Pasties	Meat/Poultry	Perth Pasties,Meat/Poultry	10	262
Alejandra Cam...	Spain	Davolio	Canada	Tourtière	Meat/Poultry	Tourtière,Meat/Poultry	5	29.5
Alejandra Cam...	Spain	Davolio	Germany	Nord-Ost Matjes...	Seafood	Nord-Ost Matjeshering,Seafood	10	207
Alejandra Cam...	Spain	Peacock	Brazil	Guaraná Fantás...	Beverages	Guaraná Fantástica,Beverages	6	21.6
Alejandra Cam...	Spain	Peacock	France	Camembert Pierrot	Dairy Products	Camembert Pierrot,Dairy Products	10	340
Alejandra Cam...	Spain	Peacock	Germany	Nord-Ost Matjes...	Seafood	Nord-Ost Matjeshering,Seafood	7	150.09
Alejandra Cam...	Spain	Peacock	Italy	Raviole Annelo	Grains/Cereals	Raviole Annelo,Grains/Cereals	2	31.2

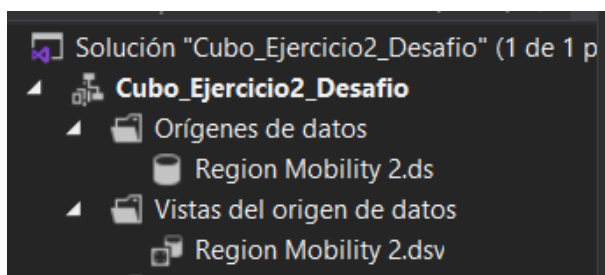
Procedimiento alcanzado: 100%

Ejercicio 2

La solución al problema en el ejercicio 2 fue sencillo en cuanto al ETL, porque lo único que deseamos hacer fue el cargar estos datos en la base de datos directamente desde el archivo .csv todo esto para facilitar la realización del cubo, por lo cual el proceso de ETL solo constó de traspasar los datos al SQL SERVER incluso sin la conversión de datos porque al realizarla, el archivo no mostraba información, tal como se muestra en la siguiente imagen, es como se realizó el proceso:



Una vez verificado que la información del archivo plano, se transfirió correctamente la tabla, fue necesario crear la conexión a la base de datos con el integration services, y a su vez el crear una vista del origen de datos, donde se mostraba la tabla que se realizó con el proceso anterior.

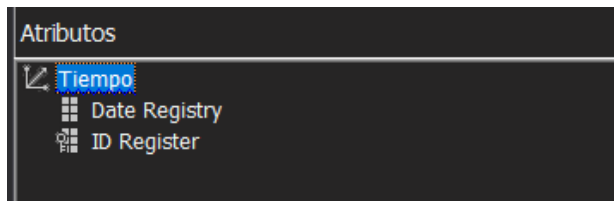


Luego de esto se creó el cubo, dejando así como dimensión "Country" o "País" y luego se agregaron otras 2 dimensiones, las cuales fueron la dimensión "Tiempo" o mejor dicho "Fecha" y por ultimo la dimensión de "Movilidad" dando así a cada uno sus campos respectivos.

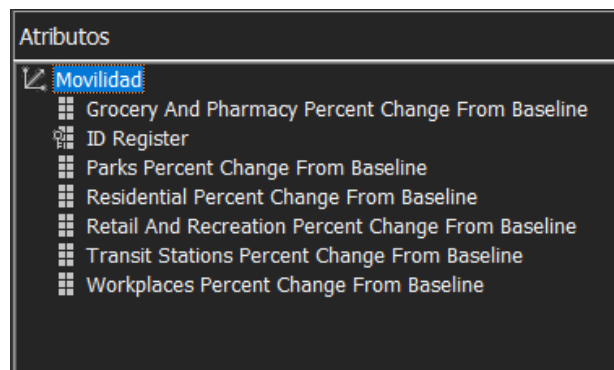


(Se creó una jerarquía con los campos de Country)

Dimensión Fecha:



Dimensión Movilidad:



Bien y finalmente la solución a la pregunta de “Que representan los valores negativos de los campos de la dimensión “Movilidad” es la siguiente, y es que representa que la movilidad en dichos sitios disminuyó.

Tal es el caso de ahuchapan; en ahuchapan la movilidad en parques, residenciales, lugares de trabajo y otros, desde el 22 de febrero al 26 de septiembre disminuía continuamente y en gran medida día con día (En este caso se muestran los valores de Parques)

Country Region Code	Country Region	Sub Region 1	Grocery And Pharmacy Perc...	Parks Percent Change From ...	Date Registry	Recuento Mobilit
/	El Salvador	Ahuachapán...		0	2020-03-08	1
/	El Salvador	Ahuachapán...		-1	2020-02-17	1
/	El Salvador	Ahuachapán...		-1	2020-02-21	1
/	El Salvador	Ahuachapán...		10	2020-02-19	1
/	El Salvador	Ahuachapán...		10	2020-03-04	1
/	El Salvador	Ahuachapán...		11	2020-02-18	1
/	El Salvador	Ahuachapán...		-11	2020-03-13	1
/	El Salvador	Ahuachapán...		12	2020-02-27	1
/	El Salvador	Ahuachapán...		13	2020-02-26	1
/	El Salvador	Ahuachapán...		-14	2020-03-16	1
/	El Salvador	Ahuachapán...		-16	2020-03-17	1
/	El Salvador	Ahuachapán...		2	2020-03-03	1
/	El Salvador	Ahuachapán...		-2	2020-02-28	1
/	El Salvador	Ahuachapán...		-2	2020-03-09	1
/	El Salvador	Ahuachapán...		20	2020-02-20	1
/	El Salvador	Ahuachapán...		21	2020-02-23	1
/	El Salvador	Ahuachapán...		-22	2020-03-18	1
/	El Salvador	Ahuachapán...		23	2020-03-01	1

Otro caso, sería si verificamos a nivel de San Salvador, donde se muestra el índice de personas que dejaron de transitar por dichos lugares; todo esto se realiza haciendo un filtro al cubo, en el que se muestren simplemente los valores de “Sub región 1” en el que sea igual a “San Salvador” y a partir de ahí se miran los valores obtenidos y la movilidad y la forma en la que cambió:

Dimensión	Jerarquía	Operador	Expresión de filtro	Parámetro
Mobility	Sub Region 1	Igual	{ San Salvador Department }	<input checked="" type="checkbox"/>
<Seleccionar dimensión>				<input type="checkbox"/>

Country Region Code	Country Region	Sub Region 1	Grocery And Pharmacy Perc...	Parks Percent Change From ...	Date Registry	Recuento Mobilit
/	El Salvador	San Salvado...	0	-2	2020-02-22	1
/	El Salvador	San Salvado...	0	4	2020-02-19	1
/	El Salvador	San Salvado...	0	-5	2020-03-10	1
/	El Salvador	San Salvado...	1	2	2020-03-07	1
/	El Salvador	San Salvado...	1	-2	2020-02-27	1
/	El Salvador	San Salvado...	1	-4	2020-02-20	1
/	El Salvador	San Salvado...	1	-4	2020-02-24	1
/	El Salvador	San Salvado...	-1	-4	2020-02-23	1
/	El Salvador	San Salvado...	-1	-4	2020-03-08	1
/	El Salvador	San Salvado...	10	-23	2020-03-16	1
/	El Salvador	San Salvado...	11	-15	2020-03-14	1
/	El Salvador	San Salvado...	13	-14	2020-03-13	1
/	El Salvador	San Salvado...	14	4	2020-03-11	1
/	El Salvador	San Salvado...	-14	-41	2020-08-15	1
/	El Salvador	San Salvado...	-16	-47	2020-07-31	1
/	El Salvador	San Salvado...	-17	-51	2020-03-20	1
/	El Salvador	San Salvado...	-19	-42	2020-08-01	1
/	El Salvador	San Salvado...	2	1	2020-03-06	1

Por lo que se concluye que en toda la época de pandemia donde hubo los picos más altos de contagios, la movilidad en todos los lugares como de trabajo, parques, tiendas u otras, cambió en gran manera pero siempre hubieron días en los que disminuía y otros en los que aumentaba, al igual que cabe destacar que hubieron días que no se obtuvieron datos en ningunos de los lugares porque probablemente no los calcularon ni los plantearon en el archivo plano.

Procedimiento alcanzado: 100%