**Data Process I**

Ejercicio Integrador

Implementación de una Solución de Datos

Para implementar una solución de datos al ejercicio integrador, se optó por una serie de supuestos que fomentan a la misma.

En primer lugar, se partía de las premisas de ciertas reglas de negocio a cumplir:

* Un producto puede estar en más de una venta.
* En una venta un producto puede tener más de una cantidad vendida.
* Una Sucursal puede contener más de una venta.
* Un Producto corresponde a una Categoría.
* Un vendedor puede estar en más de una venta.
* El nivel de detalle es diario.

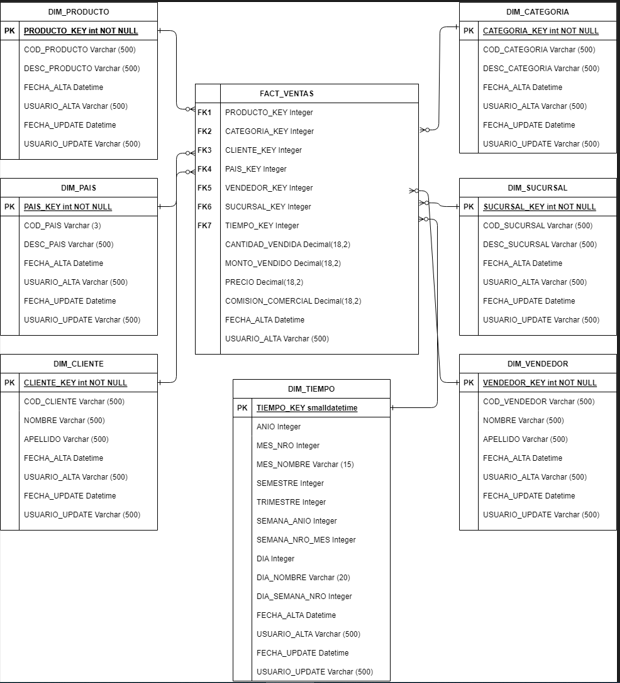
Esto indica a grandes rasgos, que las ventas se indican por día, y que tanto producto, como sucursal y vendedor pueden estar en más de una venta, ya que es una relación de uno a muchos. Es decir, no es una sola vez que se dan estos casos. A su vez en una misma venta, un producto puede venderse en más de una cantidad.

La única regla que me generó algo de dudas fue “Un producto corresponde a una Categoría”, ya que podría estar indicando un modelo de copo de nieve, al asociar cada producto a una categoría a través de tablas.

A esto simplemente lo tomé, como un supuesto lógico y no necesariamente que deba llevar a un cambio en el modelo de negocio. Sino que un producto corresponde a una categoría, como por ejemplo un vendedor puede corresponder a una sucursal, o una sucursal a un país, sin necesidad de estar en una relación enlazada entre ellas.

En el archivo AXEL\_MORIENA\_DW\_COMERCIAL\_VALIDACIONES, se encuentra una query que indica que cada producto pertenece a una categoría, sin necesidad de formar un modelo copo de nieve.

Por lo tanto, el modelo de negocio que se optó, es de un Esquema Estrella siguiente:



En la tabla DIM\_VENDEDOR, en el PDF ofrecido se indica el COD\_VENDEDOR como INT pero se tomó como un error, ya que en todos los demás casos es un Varchar(500), y no serviría quitarle la V del principio del código, ya que es información que significa de que tabla viene.

Con respecto a las comisiones comerciales hay valores muy bajos, cuasi nulos y podemos interpretar como supuesto de que tal vez sean ventas en que no tienen comisiones o se ofrece otro trato entre vendedor y gerente.

Se le puede generar a la empresa una muestra de las comisiones muy pequeñas y que luego ellos nos indiquen si es correcto o ha sido algún problema de carga, falta de datos, etc.

Por las muestras explicadas en **Validaciones punto 2 (Dos).** del archivo de validaciones, se tomará estos valores como correctos, como si la empresa nos hubiera indicado que es porque no se pueden ingresar ventas sin comisiones.

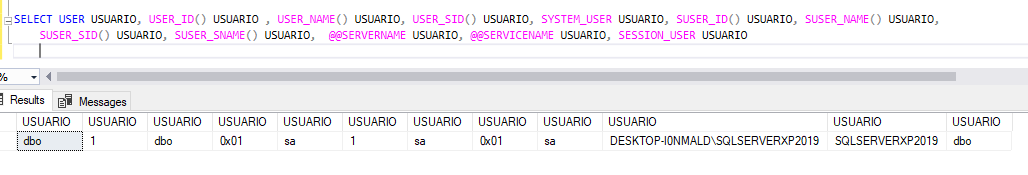
Siguientemente se observa error en los signos de 18 registros de ventas que se decidió cambiar en la carga hacia la Fact\_Ventas definitiva para que queden todos en positivo.

Se crean además, vistas para la carga de datos ordenados por fecha, ya que quedan inexactos de la manera en que están en el Excel.

Con respecto a la FECHA\_UPDATE y USUARIO\_UPDATE, si bien su función principal es saber en qué fecha y qué usuario registró cambios en nuestro DW, por ejemplo, si hubo un cambio de categoría o alguna actualización por error de tipeo, incorrección o equivocación.

Para lo que nos compete, es un caso que no va a pasar y para chequear que realmente funcione toda la “infraestructura”, se va a considerar dichos campos que se llenen (o actualicen) cada vez que se ejecuten los procedimientos almacenados que rellenen nuestras tablas finales. Para así saber, independientemente si hubo o no cambios, cuando fue la última vez que se actualizo la tabla (a través de dichos SP) independientemente si hubo o no cambios.

Otro supuesto integrado, es que en los procedimientos almacenados que se ejecuten para la carga de las tablas finales de dimensiones, el nombre de usuario de quién lo carga se lo deberá indicar a mano cada persona, una vez que ejecute el SP en cuestión. Esto, con el fin de indicar quién es el encargado de la carga de esos datos y sabiendo que, si no se le ingresa ningún nombre, no se ejecutará el procedimiento que realice la carga. Se decidió de esta manera, ya que se intentó con los siguientes campos que trae SQL SERVER y ninguno serviría para el caso en cuestión. Ya que no lo diferenciaría del resto.



En este caso, el usuario que generó las tablas será 'Axel' y quien ejecutó el procedimiento por última vez 'Moriena'. Excepto en la tabla DIM\_TIEMPO, que aún no se actualizó porque será cada nuevo año de datos, y la tabla FACT\_VENTAS.

En ningún procedimiento almacenado puse la cláusula SET NOCOUNT ON

Por último, el supuesto final será el de toda la ingeniería. Y se dará por las dudas, para que se entienda todo el proceso y no queden dudas al respecto.

Es decir, se ingresan tablas Staging para cada tabla desde el Excel, en estos campos se van a ingresar nuevos registros para continuar con la actualización de información. Una vez completas las tablas de STG, se incorporarán a las tablas Intermedias. Por ello, cada procedimiento va limpiando los datos para las cargas finales y siempre se realiza un TRUNCATE para ir eliminando datos viejos, ya que los archivos del origen de datos serán distintos y en cada carga se comienza desde cero para evitar duplicados.

Una vez realizadas todas estas cargas a través de los últimos procedimientos se actualizan las tablas finales (DIMENSIONES y DE HECHO), aquí sí, si hay duplicados ( o datos cargados anteriormente) no se cargarán y se ingresarán las fechas y el usuario que realizó esta ultima carga.