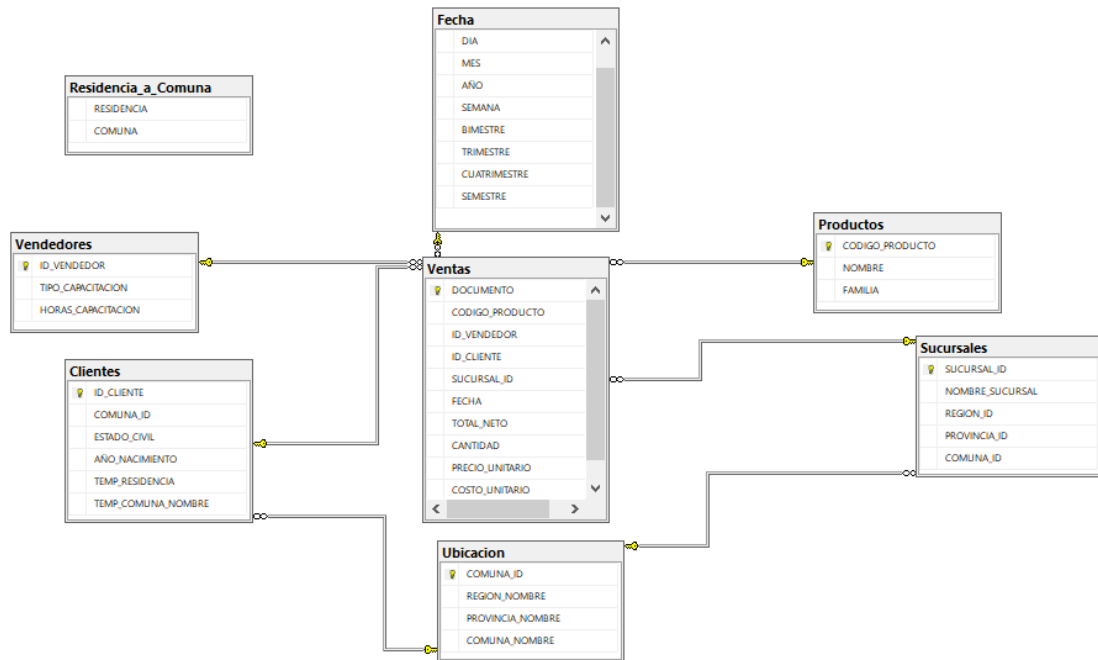


## Trabajo Práctico Final BD2

Enzo Ferrari

### Diagrama Estrella del Datawarehouse



Para el diagrama estrella, se toma como tabla de hechos las ventas de la empresa, ya que este es la actividad de interés para analizar de la empresa.

Según los requerimientos del negocio, las tablas de dimensiones son:

- Vendedores: Origen base de datos transaccional.
- Productos: Origen base de datos transaccional.
- Clientes: Origen base de datos transaccional.
- Sucursales: Origen base de datos transaccional.
- Fecha: Calculado.

Los motivos por los cuales se eligieron estas dimensiones y sus columnas fueron discutidas en la primera entrega del TP.

La tabla Ubicacion es una tabla que se usa tanto como para contextualizar a los Clientes y Sucursales respecto a su ubicación, ya que es preferible que se usen los mismos parámetros para esto. Esto traerá problemas.

Datos faltantes, discrepancias: Las tablas COMUNA, PROVINCIA y REGION de la base de datos transaccional proveen con la totalidad de la información de las comunas, provincias y regiones de Chile, según los Códigos Únicos Territoriales:

<https://www.subdere.gov.cl/documentacion/c%C3%B3digos-%C3%BAnicos-territoriales-actualizados-al-06-de-septiembre-2018>

De esta manera se pueden vincular sin problemas los registros de la tabla Sucursales con los de Ubicacion.

Sin embargo, existen problemas respecto a la ubicación en la tabla Clientes. Los valores que incumben a esta tabla son los de RESIDENCIA, que son las comunas en donde se encuentra cada cliente. Estos valores presentan problemas:

- Contienen 'non-breaking spaces'. Primero se deben tratar los datos.

- Muchos registros no son realmente comunas, sino más bien, son balnearios, pueblos, etc. Deben ser transformados a la comuna a la cual pertenecen.

El primer problema es solucionado fácilmente con consultas de SQL en las herramientas de SSIS, para el segundo se deben tomar de datos externos. Para esto se usa la tabla Residencia\_a\_Comunas, donde se tomaron todas las residencias presentes en Clientes que no coincidían con ninguna comuna en COMUNAS (desde una consulta SQL) y luego, usando inteligencia artificial y posterior revisión, se creó un csv con las columnas RESIDENCIAS y COMUNA, donde para cada residencia desconocida se tiene la comuna a la cual pertenece (residencias.csv).

Tabla fecha: Se crearon los registros mediante consultas de SQL. En la BD transaccional datan ventas del año 2015, por precaución, se crearon registros en la tabla Fecha desde 2010 hasta 2110, por cada día del año. Se incluyen también columnas como día, mes, año, semana, bimestre, etc. para facilitar la visualización.

Año\_Nacimiento: En la tabla Clientes, se decide usar el año de nacimiento y no la edad, así se evita cambiar los registros año a año. Se toma como año actual 2023.

El resto de la carga de datos se realizó con simples entradas y salidas de OLE DB.