Requerimientos

* Monto total vendido por vendedor junto a sus utilidades y comisión durante el año 2006
* Cantidad de autos vendidos por marca y por agencia en el turno de la mañana
* Total de utilidades obtenidas por un vendedor y por tipo de vehículo dentro de una agencia en un mes
* Venta más alta por agencia mostrando el vendedor con detalle y monto total vendido
* Mostrar los años con más vehículos vendido de cada lugar, con su cantidad vendida y su vendedor
* El vehículo más costoso vendido por la agencia La estrella en el mes de febrero
* Auto más vendido por agencia y por vendedor durante los últimos 5 años

**Dimensiones previas:**

vendedor

comisión

Tiempo

marca

agencia

vehículo

lugar

**Dimensiones:**

**vendedor**:id vendedor, Nombre ,dirección,teléfono ,rut

Origen:

* OLTP del sistema excel:
  + Hoja: Data
  + Columna: Vendedor
* OLTP del modelo relacional:
  + Tabla vendedor

**Tiempo** :id tiempo ,dia\_venta, mes\_venta ,año\_venta,dia\_turno, mes\_turno ,año\_turno,TipoTurno

Origen:

* OLTP del sistema excel:
  + Hoja: Data
  + Columna: Fecha,Turno
* OLTP del modelo relacional:
  + Tabla venta

**agencia:** id agencia, nombre,ubicación

Origen:

* OLTP del sistema excel:
  + Hoja: Data
  + Columna: Agencia,Lugar

**vehículo**: id vehiculo, modelo, marca,valor

Origen:

* OLTP del sistema excel:
  + Hoja: Data
  + Columna: Vehiculo,marca
* OLTP del modelo relacional:
  + Tabla vehiculo\_nuevo

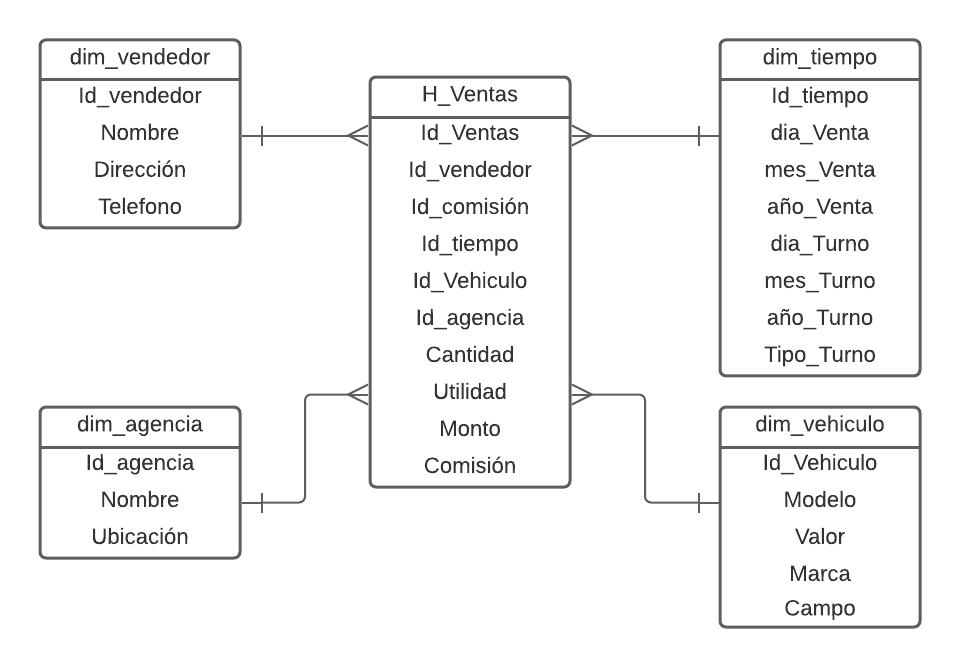
**Hechos:**

Ventas: id ventas, cantidad, utilidad, monto ,tipo\_Turno,fecha\_turno,comisión

Origen:

* OLTP del sistema excel:
  + Hoja: Data
  + Columna: Monto,Utilidades,Cantidad,Turno,Fecha,Comisión
* OLTP del modelo relacional:
  + Tabla venta

**Modelo dimensional:**

****

**CASO DE PROCESO ETL**

El gerente de la concesionaria “TodoMotor” una empresa con más de 7 años en el rubro que cuenta con sucursales en la capital de Perú y en el norte de Chile, específicamente en Antofagasta, y comenzó a expandirse hacia el sur de Chile en la ciudad de Temuco. Por lo tanto los datos que se manejan en la actualidad irán en aumento, y se necesita que estos datos pueda darles información relevante y que sirva para la toma de decisiones en la empresa.

Las 2 sucursales ubicadas en la capital de Perú y en Antofagasta, trabajan respaldando la información en planillas Excel donde registran su información diaria ya sea de ventas o compras de vehículos y algún que otro detalle. Por otra parte, la nueva sucursal cuenta con una base de datos.

También cuentan con un servidor web donde ingresa información de vehículos nuevos que la concesionaria tiene a la venta, esto a modo de hacer publicidad a su empresa.

El gerente contrató a un ingeniero en informática y este le sugirió que para la buena toma de decisiones en su empresa es necesario realizar un proceso ETL, donde se debe comenzar a extraer los datos de las sucursales actuales, estos datos serán transformados, para luego ser almacenados en un DataWarehouse, con esto lograra tener una mejor visión de la condición actual y futura de su empresa. .

**ANÁLISIS Y DISEÑO DEL DW**

El ingeniero informático a cargo de esta tarea realizó un modelo de acuerdo a los requerimientos presentados por el gerente, estos requerimientos son los siguientes:

* Monto total vendido por vendedor junto a sus utilidades y comisión durante el año 2006
* Cantidad de autos vendidos por marca y por agencia en el turno de la mañana
* Total de utilidades obtenidas por un vendedor y por tipo de vehículo dentro de una agencia en un mes
* Venta más alta por agencia mostrando el vendedor con detalle y monto total vendido
* Mostrar los años con más vehículos vendido de cada lugar, con su cantidad vendida y su vendedor
* El vehículo más costoso vendido por la agencia La estrella en el mes de febrero
* Auto más vendido por agencia y por vendedor durante los últimos 5 años

Para llegar a un modelo dimensional el ingeniero debió antes realizar un análisis a los distintos atributos y tablas. Este análisis comienza por describir las tablas de hechos y dimensiones a utilizar.

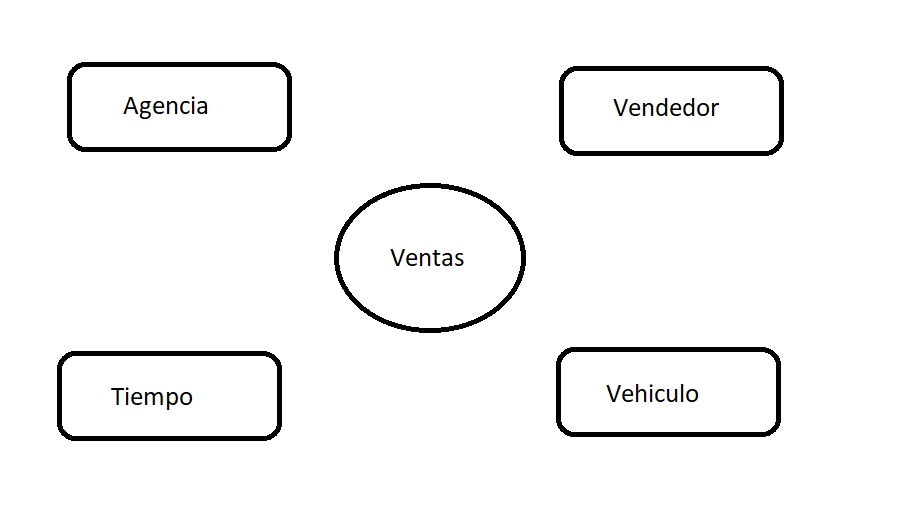
La siguiente tabla de hecho corresponde a la ventas realizadas en la concesionaria

|  |  |
| --- | --- |
| **Tabla de hecho** | **Descripción** |
| H\_Ventas | Incluye todo el proceso de ventas de la concesionaria |

Las siguientes dimensiones son las perspectivas de negocios sobre las que se analizará el proceso de ventas de la concesionaria.

|  |  |
| --- | --- |
| **Dimensiones** | **Descripción** |
| dim\_vendedor | Información asociada al vendedor |
| dim\_agencia | Nombre de la agencia |
| dim\_tiempo | Fecha de venta y turno de vendedores |
| dim\_vehiculo | Información asociada al vehículo |

El modelo conceptual resultante sería el siguiente:



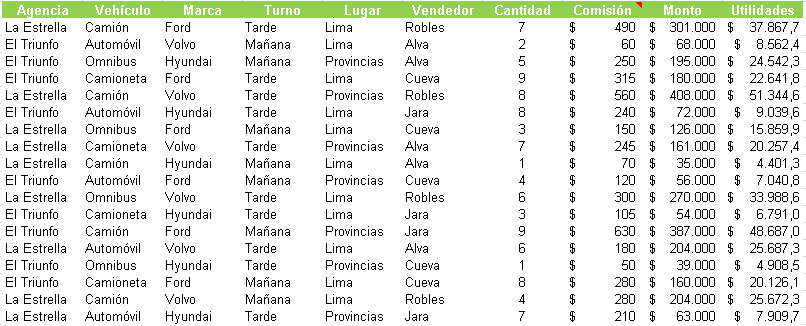
**Extracción**

Fase de análisis

Para comenzar con el proceso de extracción debemos analizar los datos que se encuentran en la base de datos y en la planilla excel. En la planilla utilizaremos la hoja Data, en la cual extraemos las siguientes columnas de datos:

* Agencia
* Vehículo
* Marca
* Turno
* Lugar
* Cantidad
* Comisión
* Monto
* Utilidades



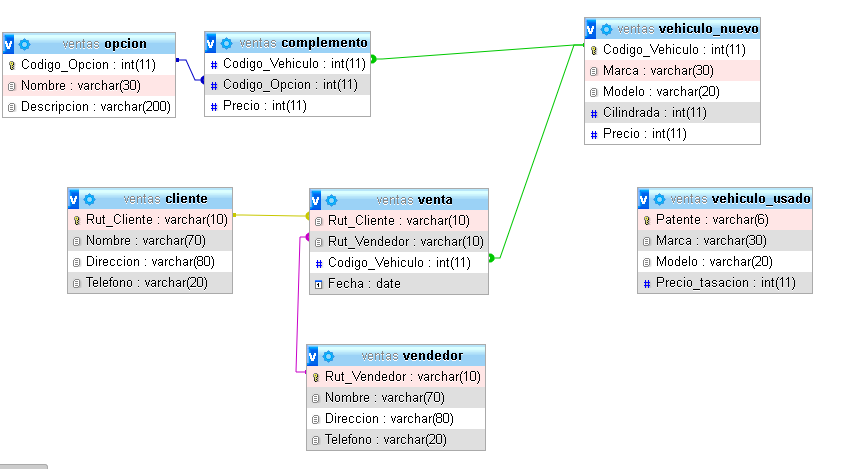


El proceso de ventas realizado por la sucursal ubicada en Temuco, esta representado por el diagrama relacional de la siguiente imagen:

Diagrama Relacional: representa la información a través de entidades, relaciones, cardinalidades, claves, atributos y jerarquías de generalización.

De este diagrama relacional de acuerdo con los requerimientos dados por la gerencia se utilizará las siguientes tablas y atributos

|  |  |
| --- | --- |
| Tabla | Atributos |
| Vehiculo\_nuevo | Modelo,marca,valor |
| Vendedor | Rut,nombre,direccion,telefono |
| Venta | fecha |



De este modelo base de datos se descarta las tablas vehiculo\_usado debido a que no está relacionada con los requerimientos dados por el gerente y además que la concesionaria solo vende autos nuevos, igualmente se descarta la tabla de opción y complemento, donde estas corresponden a accesorios de vehículos que la empresa tiene planeado incluir en un futuro próximo y por último la tabla cliente la cual no es necesaria para los requerimientos dados.

**Establecer correspondencias**

Luego de seleccionar las tablas y atributos a utilizar, el ingeniero establece las relaciones que tendrán estos campos con el modelo conceptual

Las relaciones identificadas con el modelo relacional fueron las siguientes:

* La tabla” Vendedor” se relaciona con la perspectiva” Vendedor”.
* La tabla” Vehiculo\_nuevo” con la perspectiva ”Vehiculo”.
* El campo ”fecha” de la tabla ”Venta” con la perspectiva ”Tiempo” (debido a que es la fecha principal en el proceso de venta)

Las relaciones identificadas con la planilla excel fueron las siguientes:

* La columna ”Agencia” se relaciona con la perspectiva” Agencia”.
* La tabla ”Vehiculo\_nuevo” con la perspectiva ”Vehiculo”.
* La columna ”turno” con la perspectiva ”Tiempo” (debido a que se identifica al turno del vendedor quien realizo el proceso de venta)
* La columna ”cantidad” con el hecho ”Venta” (debido a que representa la cantidad de vehiculos vendidos)
* La columna ”comision” con el hecho ”Venta” (debido a que representa la comision por vehiculos vendidos)
* La columna ”monto” con el hecho ”Venta” (debido a que representa el monto por vehiculos vendidos)
* La columna ”Utilidades” con el hecho ”Venta” (debido a que representa la utilidad por vehiculos vendidos)

FALTA ARREGLAR AQUII

De acuerdo a las correspondencias establecidas, se analizaron los campos residentes en cada fuente de datos.

*Con respecto a la perspectiva ”Clientes”, los datos disponibles son los siguientes:*

tabla a la que se hacia referencia, a través de dos métodos diferentes. Primero se examinó la base de datos para intuir los significados de cada campo, y luego se consultó con el encargado del sistema sobre algunos aspectos de los cuales no se comprendía su sentido.

De todas formas, y como puede apreciarse en el diagrama de relacional antes expuesto, los nombres de los campos son bastante explícitos y se deducen con facilidad, pero aún así fue necesario investigarlos para evitar cualquier tipo de inconvenientes.

**identificación de atributos, tipo de dato y nivel de información**

|  |  |  |
| --- | --- | --- |
| Nombre atributo | Tipo de dato | Nivel de información |
| Agencia | Cadena | Identifica al nombre de agencia |
| Vehículo | Cadena | 1 nivel (equivalente a tipo de vehículo) |
| Marca | Cadena | 1 nivel (equivalente a marca de vehículo) |
| Turno | Cadena | 1 nivel (equivalente a mañana/tarde) |
| Lugar | Cadena | Identificador al lugar |
| Vendedor | Cadena | Identificador al vendedor |
| Cantidad | Número entero | Campo para omitir, dado que el sistema se encargará de estos cálculos |
| Comisión | Número entero | Campo para omitir, dado que el sistema se encargará de estos cálculos |
| Monto | Número entero | Campo para omitir, dado que el sistema se encargará de estos cálculos |
| Utilidades | Número entero | Campo para omitir, dado que el sistema se encargará de estos cálculos |
| Rut | Cadena | 1 nivel (equivalente al identificador del vendedor) |
| Dirección | Cadena | Descripción de la vivienda del vendedor |
| Teléfono | Número entero | Descripción del contacto al vendedor |
| FechaVenta | Fecha | 3 niveles (año, mes, dia) |

Luego de analizar los datos que se extraerán de ambas fuentes, se analiza el formato de estos, para ponerlos en un formato que coincidan en el DW y así no tener problemas al momento de mostrar la información recopilada.

Para la tabla de hecho de ventas, tendremos:

|  |  |  |
| --- | --- | --- |
|  | **H\_Ventas** | **Tipo** |
| (PK) | ID\_Ventas | Clave primaria |
| (FK) | Id\_vendedor,Id\_tiempo,Id\_Vehiculo,Id\_agencia | Claves foráneas |
|  | Cantidad,Utilidad,Monto,Comisión | Atributos |

Para la tabla de dimensiones, tendremos:

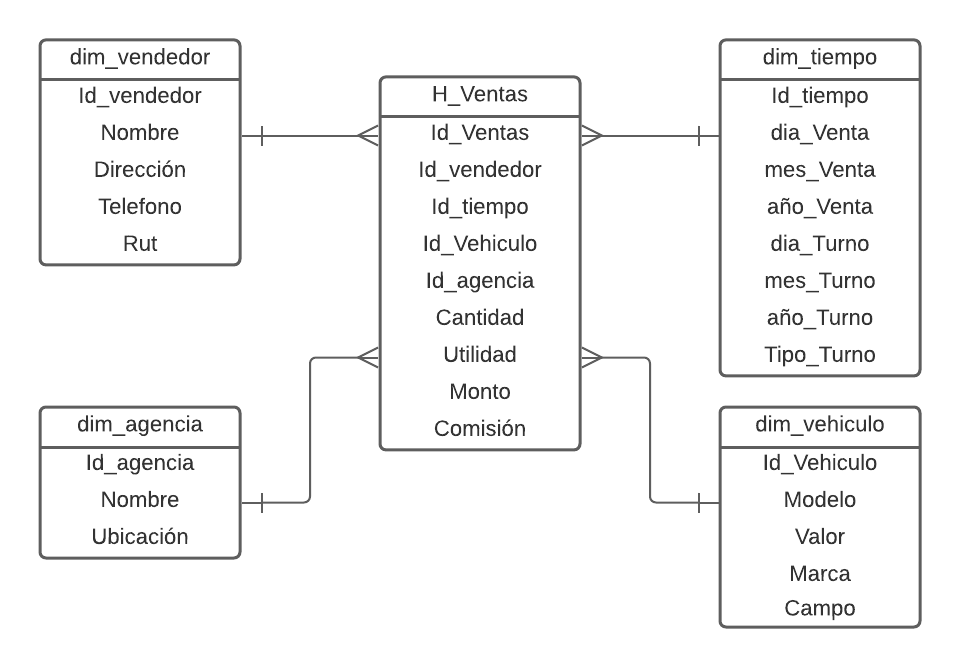
|  |  |  |
| --- | --- | --- |
|  | **Dim\_Agencia** | **Tipo** |
| (PK) | ID\_Ventas | Clave primaria |
| (FK) | Id\_vendedor,Id\_tiempo,Id\_Vehiculo,Id\_agencia | Claves foráneas |
|  | Cantidad,Utilidad,Monto,Comisión | Atributos |

|  |  |  |
| --- | --- | --- |
|  | **Dim\_Vendedor** | **Tipo** |
| (PK) | ID\_Ventas | Clave primaria |
| (FK) | Id\_vendedor,Id\_tiempo,Id\_Vehiculo,Id\_agencia | Claves foráneas |
|  | Cantidad,Utilidad,Monto,Comisión | Atributos |

|  |  |  |
| --- | --- | --- |
|  | **Dim\_Tiempo** | **Tipo** |
| (PK) | ID\_Ventas | Clave primaria |
| (FK) | Id\_vendedor,Id\_tiempo,Id\_Vehiculo,Id\_agencia | Claves foráneas |
|  | Cantidad,Utilidad,Monto,Comisión | Atributos |

|  |  |  |
| --- | --- | --- |
|  | **Dim\_Vehiculo** | **Tipo** |
| (PK) | ID\_Ventas | Clave primaria |
| (FK) | Id\_vendedor,Id\_tiempo,Id\_Vehiculo,Id\_agencia | Claves foráneas |
|  | Cantidad,Utilidad,Monto,Comisión | Atributos |

El esquema resultante una vez acabado el analisis es el siguiente:

****

Por otro lado, el gerente realizó un análisis de prioridades de cada requisito, lo hizo en una escala de prioridad de 1 a 4, siendo 1 completamente prioritario y 4 no prioritario para la actividad.

A continuación, se describen los requerimientos funcionales con su correspondiente prioridad asignada.

|  |  |  |
| --- | --- | --- |
| Número | Requerimiento | Prioridad |
| 1 | Monto total vendido por vendedor junto a sus utilidades y comisión durante el año 2006 | 2 |
| 2 | Cantidad de autos vendidos por marca y por agencia en el turno de la mañana | 2 |
| 3 | Total de utilidades obtenidas por un vendedor y por tipo de vehículo dentro de una agencia en un mes | 2 |
| 4 | Venta más alta por agencia mostrando el vendedor con detalle y monto total vendido | 1 |
| 5 | Mostrar los años con más vehículos vendido de cada lugar, con su cantidad vendida y su vendedor | 1 |
| 6 | El vehículo más costoso vendido por la agencia La estrella en el mes de febrero | 3 |
| 7 | Auto más vendido por agencia y por vendedor durante los últimos 5 años | 3 |

Para realizar esta transformación se utilizara la herramienta Microsoft Visual Studio, esta herramienta además de ser una aplicación gratuita, tiene la extensión necesaria para realizar este proceso de manera correcta.

Para comenzar a realizar el proceso de extracción, se ha acordado un horario en el cual no se esté utilizando los sistemas. Por lo tanto este proceso se realizará de lunes a sábado de 22:00 a 00:00 horas , para así evitar un colapso o una ralentización del sistema.

**Fase práctica**

Ya comenzando con el proceso de transformación, el ingeniero decide realizar una copia de seguridad de la planilla excel ante cualquier situación de riesgo que podría suceder.