



UNIVERSIDAD NACIONAL DE LA MATANZA

INTELIGENCIA DE NEGOCIOS

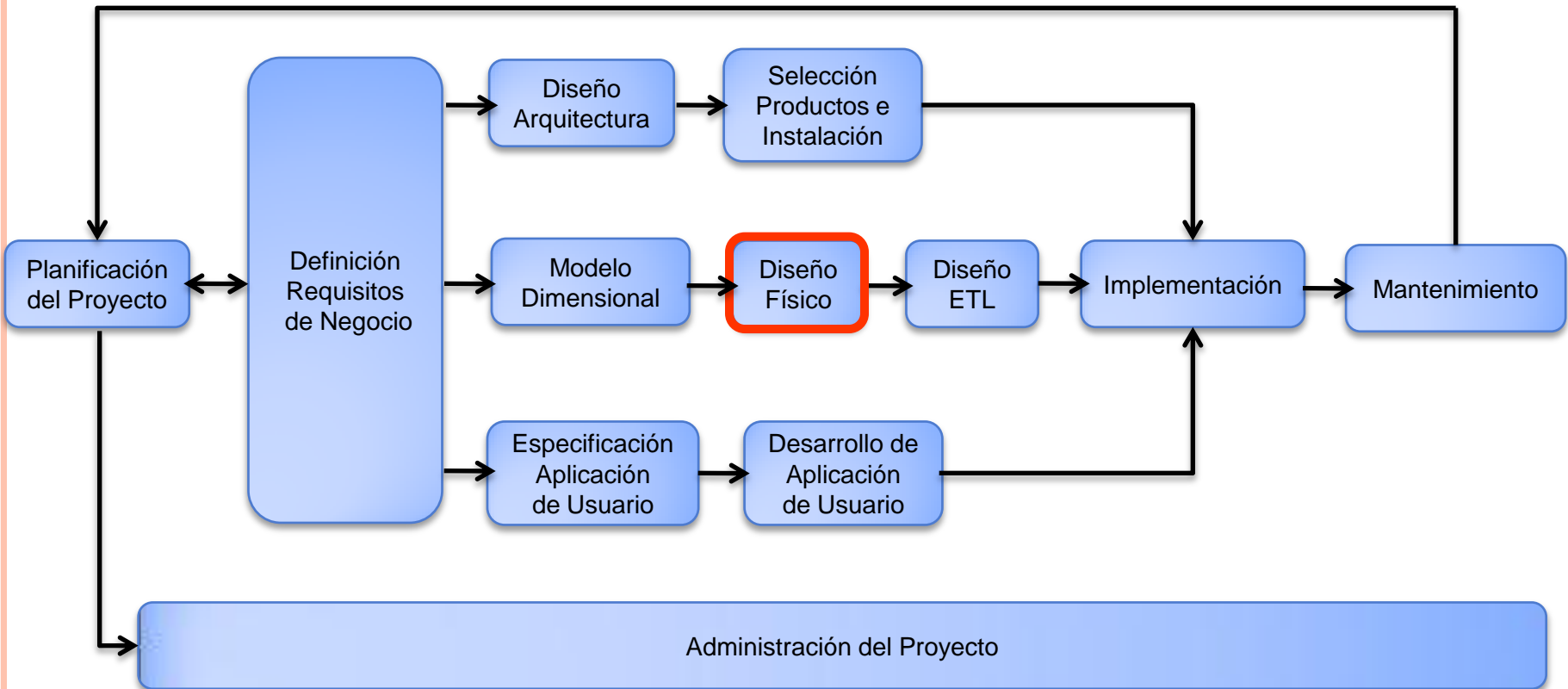
Modelo Físico

Técnicas de Modelado – Parte 2

Profesor: Mg. Diego Basso

Curso 2017

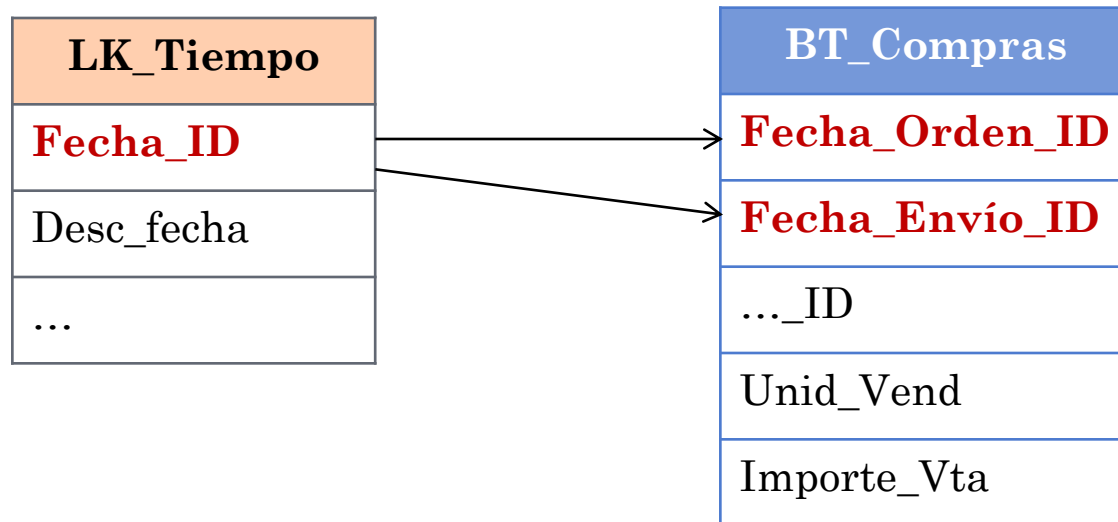
CICLO DE VIDA DE UN PROYECTO DE BI



DIMENSIÓN MÚLTIPLE



- Existen situaciones que referencian a diferentes instancias de una misma dimensión en la tabla de hechos.
- Ejemplo: Una orden es solicitada en una fecha determinada y enviada días después.



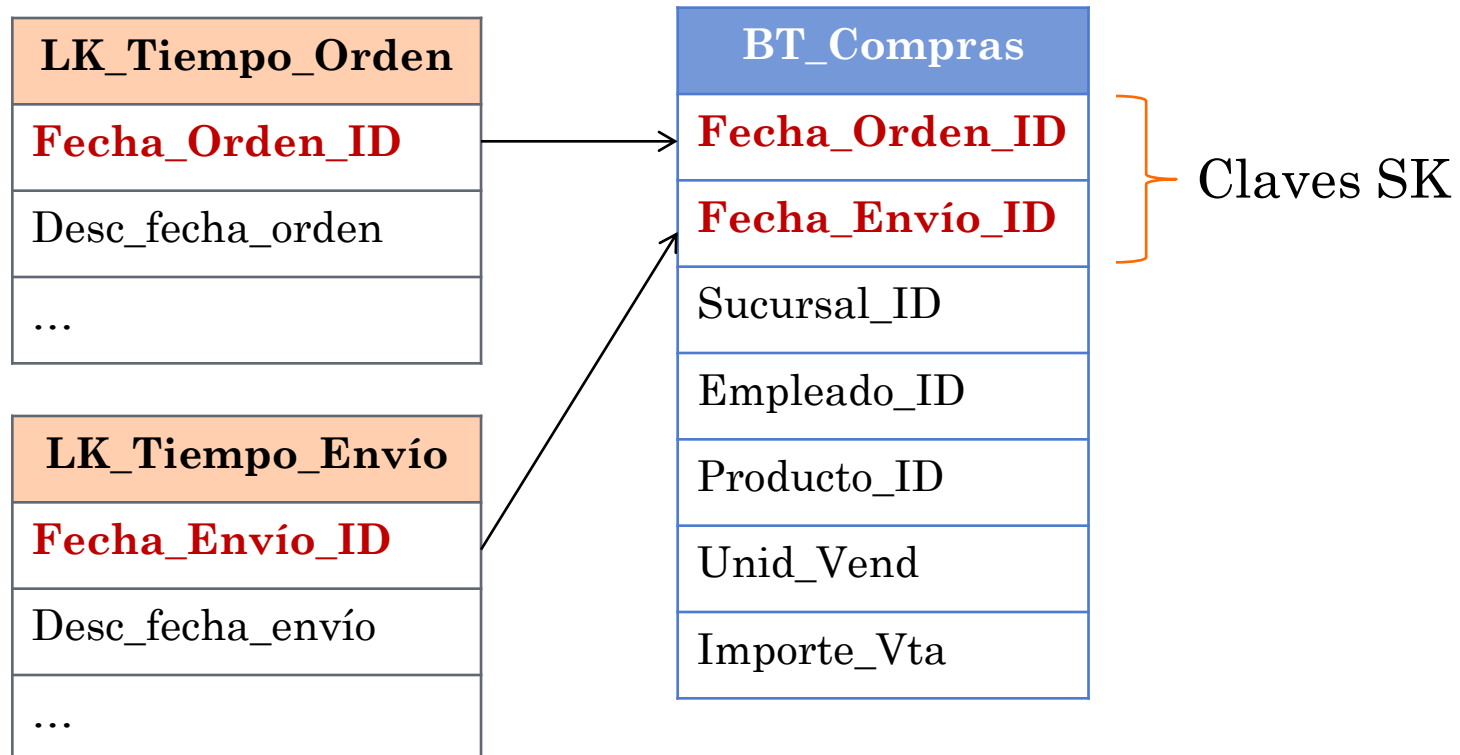
¿Es correcto este modelo?





DIMENSIÓN MÚLTIPLE

- Al momento de querer implementarlo tendríamos el problema de una dimensión de fecha que se relaciona con más de una SK de fecha en la tabla de hechos.
- Tantas tablas de dimensión como SK en la fact table.



DIMENSIÓN DEGENERADA



- Se refiere a dimensiones que tienen un identificador transaccional (no necesariamente numérico) generado en el sistema fuente.
- Ejemplos
 - Número de órdenes de compra
 - Número de factura
 - Número de remito
 - Número de transacción
- Parece una dimensión, pero ¿cuáles son sus atributos?
 - Los atributos que podrían incluirse ya están en la propia dimensión (fecha, vendedor, sucursal, cliente).
 - Aparece cuando la granularidad es la misma que la de la tabla de hechos
 - Nivel de renglón de factura o de orden de compra



DIMENSIÓN DEGENERADA



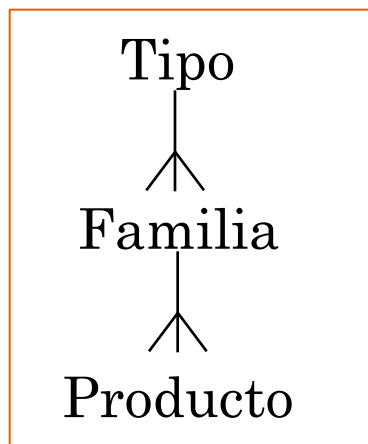
- Esta dimensión no tiene una tabla asociada.
 - Atributo con sólo 2 columnas: Clave SK + N° transacción
- Es un atributo que será utilizado como criterio de análisis y que es almacenado en la *tabla de hechos*.
 - No permite agrupación ni sumarización.
- ¿Para qué sirve?
 - Permite vincular todos los ítems pertenecientes a una misma transacción.
 - Sirve como fuente para procesos que requieran saber los productos que se compraron juntos en la misma transacción.
 - Da la posibilidad de localizar datos en los sistemas fuente.



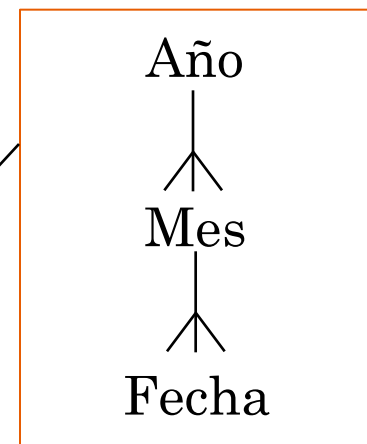
DIMENSIÓN DEGENERADA



DIM_PRODUCTO



DIM_TIEMPO



DIM_CLIENTE



Ventas	
...	ID (SK)
Unid_Vend	
Importe_Venta	
Nro_Ticket	

Dimensión Degenerada



DIMENSIÓN DEGENERADA



BT_VENTAS

Fecha_ID	Cliente_ID	Producto_ID	Unid_Vend	\$ Venta	Nro_Ticket
15/04/2017	1	10	3	\$ 90	1234
15/04/2017	1	11	2	\$ 100	1234
15/04/2017	1	12	5	\$ 150	1234
15/04/2017	3	10	2	\$ 60	2411
15/04/2017	3	13	10	\$ 200	2411

Dimensiones

Medidas

Dimensión
Degenerada



DIMENSIONES DE CAMBIO LENTO



- El horizonte temporal del DW es mayor que el de los sistemas transaccionales.
- El DW debe reflejar el paso del tiempo pero no perder la historia.
 - Un producto cambia de denominación.
 - Una sucursal cambia de distrito.
- Son dimensiones en las cuales sus datos tienden a modificarse a través del tiempo.
- Los atributos de las dimensiones son relativamente estáticos pero cambian a lo largo del tiempo.



DIMENSIONES DE CAMBIO LENTO



¿Qué hay que hacer?

- Para cada atributo hay que definir una estrategia de manejo de cambios.
- El modelo dimensional debe responder a los cambios.
- Interpretación del profesional de negocios.
 - ¿Cómo se quiere guardar la historia de cambios en los atributos?
 - ¿Con qué detalle?



DIMENSIONES DE CAMBIO LENTO



- Existen distintos tipos de técnicas para manejar cambios en los atributos.
- Se denominan SCD (Slowly Changing Dimensions)
 - Manejan en forma diferente la conservación de la historia.
 - No hay una que sea mejor que otra.
 - Se define para **cada atributo de cada dimensión**.



SCD TIPO 0



Se mantiene el valor original

- El atributo de la dimensión no permite cambios.
- Mantiene el valor original del atributo.
 - Los hechos siempre están asociados al valor original del atributo.
- Ejemplos
 - Atributos de la dimensión Tiempo
 - Datos del CUIT de la dimensión Clientes



SCD TIPO 1



Se reescribe el valor perdiendo el valor original

- No conserva la historia de cambios.
- Se modifica el dato en la tabla de dimensión (update).
- El atributo siempre contiene el último valor asignado.
- Es el más fácil de implementar.

Cliente_ID	Cod_Cliente	Nombre	Estado Civil
1	JC001	Jorge Cornejo	Soltero
2	CG015	Carlos González	Casado

Cliente_ID	Cod_Cliente	Nombre	Estado Civil
1	JC001	Jorge Cornejo	Casado
2	CG015	Carlos González	Casado



EJEMPLO SCD TIPO 1



Producto_ID	Cod_Prod	Desc_Prod	Tipo	Rubro
11	A101	Jeans	1	1

Producto_ID	Cod_Prod	Desc_Prod	Tipo	Rubro
11	A101	Jeans	1	2

- Evitar las inconsistencias en el DW.
- Si el atributo es un nivel por el que se suma (parte de una jerarquía), hay que recalcular agregaciones previas.



SCD TIPO 2



Se agrega una nueva fila o registro

- Cuando un atributo cambia de valor se agrega una nueva fila a la tabla de dimensión.
- Los nuevos hechos apuntan a la nueva fila.
- Los hechos anteriores continúan apuntando a la fila anterior.

Cliente_ID	Cod_Cliente	Nombre	Estado Civil
1	JC001	Jorge Cornejo	Soltero
2	CG015	Carlos González	Casado

Cliente_ID	Cod_Cliente	Nombre	Estado Civil
1	JC001	Jorge Cornejo	Soltero
2	CG015	Carlos González	Casado
3	JC001	Jorge Cornejo	Casado



SCD TIPO 2



- Es la técnica más usada.
- No requiere el recálculo de agregaciones.
- Permite guardar toda la historia de cambios.
- Los resultados previos al cambio siguen siendo los mismos.
- Se necesita colocar alguna *marca* o *fechas de vigencia* para saber cuándo se cambió o cuál se cambió primero.



EJEMPLOS SCD TIPO 2



Cliente_ID	Cod_Cliente	Nombre	Estado Civil	FDesde	FHasta
1	JC001	Jorge Cornejo	Soltero	02/06/2005	10/09/2009
3	JC001	Jorge Cornejo	Casado	11/09/2009	05/02/2017
10	JC001	Jorge Cornejo	Divorciado	06/02/2017	

○ Otra forma:

Cliente_ID	Cod_Cliente	Nombre	Estado Civil	Versión
1	JC001	Jorge Cornejo	Soltero	1
3	JC001	Jorge Cornejo	Casado	2
10	JC001	Jorge Cornejo	Divorciado	3

- **Nota:** si hay dudas en la elección del tipo SCD, es aconsejable elegir el 2.



SCD TIPO 3



Se agrega un nuevo atributo

- Guarda una cantidad limitada de valores históricos de atributos seleccionados.
- No incrementa el tamaño de la tabla.
- Sólo guarda el último valor
- Permite ver los datos recientes y los históricos por el *atributo actual* y el *anterior*.

Producto_ID	Cod_Prod	Desc_Prod	Tipo	Rubro	Rubro_Ant
11	A101	Jeans	1	1	-

Producto_ID	Cod_Prod	Desc_Prod	Tipo	Rubro	Rubro_Ant
11	A101	Jeans	1	2	1

EJEMPLO SCD TIPO 3



Cliente_ID	Cod_Cliente	Nombre	Estado Civil	FDesde	Est_Ant
1	JC001	Jorge Cornejo	Soltero	02/06/2005	-

- Jorge se casa, entonces:

Cliente_ID	Cod_Cliente	Nombre	Estado Civil	FDesde	Est_Ant
1	JC001	Jorge Cornejo	Casado	11/09/2009	Soltero

- Jorge se divorcia, entonces:

Cliente_ID	Cod_Cliente	Nombre	Estado Civil	FDesde	Est_Ant
1	JC001	Jorge Cornejo	Divorciado	06/02/2017	Casado

- **Nota:** Este tipo de SCD no es el más utilizado.



ATRIBUTOS CON DIFERENTE SCD



- Dado que la política SCD se define para cada atributo, una misma dimensión puede tener atributos con distinto tipo de SCD.
- Esto hay que tenerlo en cuenta al definir el ETL.



DIMENSIONES RÁPIDAMENTE CAMBIANTES



- Se denominan “Dimensiones Monstruo” (Rapidly Changing Monster Dimensions).
- Tablas de dimensiones con gran cantidad de filas.
- Los atributos descriptivos cambian frecuentemente.
- Causa un aumento desmedido de la cantidad de filas al usar SCD tipo 2.
- Son el tipo SCD 4.



SCD TIPO 4



Se agrega una mini dimensión

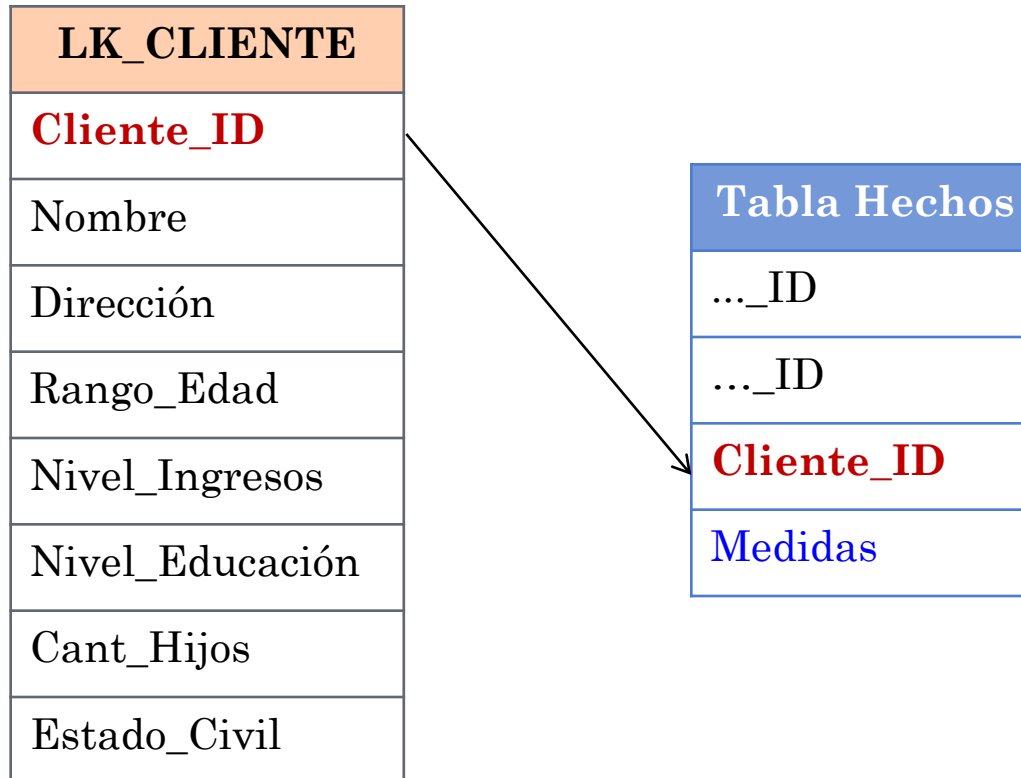
- Abrir la dimensión original en 2 partes:
- Una dimensión con los atributos que cambian con frecuencia (mini dimensión).
- Otra dimensión con el resto de los atributos (dimensión primaria).



EJEMPLO SCD TIPO 4



- Dimensión Cliente con gran cantidad de filas.



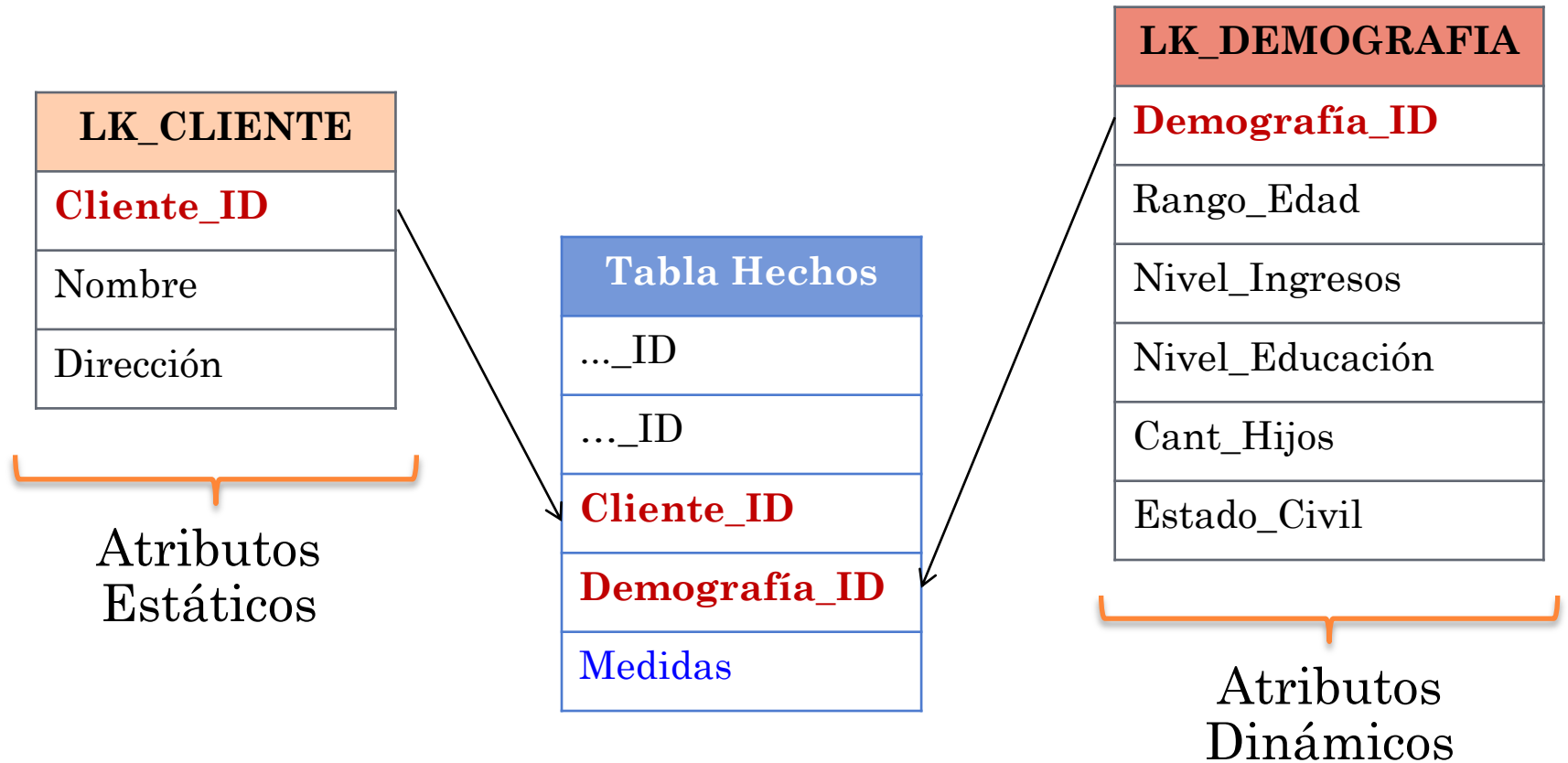
EJEMPLO SCD TIPO 4



- Hay cambios de tipo sociodemográficos.
 - Rango de Edad
 - Nivel de Ingresos
 - Nivel de Educación
 - Cantidad de Hijos
 - Estado Civil
- Se crea una dimensión que agrupa a estos atributos.
 - Cada fila de esta tabla de dimensiones contiene un juego de valores posibles de cada uno de esos atributos.



EJEMPLO SCD TIPO 4



SCD TIPO 4



- La nueva dimensión contiene un conjunto de perfiles sociodemográficos.
- Tiene pocas filas.
- La tabla de hechos contiene:
 - La clave FK de la nueva dimensión.
 - La clave FK de la dimensión original.
 - Los hechos están asociados al valor que tenían los atributos de cambio rápido en el momento en que se generaron.
- **Recordar:** Transformar valores continuos en rangos
 - Cada atributo tendrá una cantidad pequeña de valores posibles.



DIMENSIÓN JUNK



- En las fuentes aparecen atributos con datos de baja cardinalidad (marcas por SI o NO).
 - Ejemplo: Encuestas con preguntas de Si / No
- Estos atributos no forman parte de ninguna dimensión.
- Los usuarios quieren conservarlos “por las dudas”.
- ¿Armamos una dimensión por cada uno de estos atributos?
 - Esto incide directamente en la tabla de hechos.
 - 10 dimensiones con indicadores Si / No $\Rightarrow 2^{10}$ registros.



DIMENSIÓN JUNK



- Una dimensión JUNK es una agrupación de atributos no relacionados que se llevan a una dimensión.
- Se utilizan para almacenar banderas o marcas.
- Solución
 - Definir una dimensión con las combinaciones de todos esos atributos y asignarle una clave subrogada a cada combinación.
 - Técnica similar a la de las mini dimensiones SCD tipo 4.
- Alternativas para la dimensión:
 - Guardar todas las combinaciones posibles.
 - Guardar sólo las que se presentan en los datos.



MEDIDAS O HECHOS - CASOS ESPECIALES



- La medida o hecho por sí solo no genera ninguna cantidad o valor para poder medir.
 - Ejemplo: Se necesita analizar la **cantidad de veces** que un usuario accede a un sitio web.
- En la tabla de hechos se podría agregar una medida “**veces**” con valor **1** para poder operar sobre ella.

BT_ACCESOS

Fecha_ID	Hora_ID	Usuario_ID	Veces
15/04/2017	9:00	1	1
15/04/2017	9:30	2	1
15/04/2017	10:30	1	1
15/04/2017	14:00	3	1

Sumar



TABLAS DE HECHOS SIN MEDIDAS



- Factless Fact Table
- La tabla de hechos no tiene medidas sino que registra un evento o acontecimiento que se da conjuntamente con todas las dimensiones en un determinado momento.
- Ejemplo
 - Asistencia de alumnos a distintos cursos
 - Censo
 - Atención médica
- No existen medidas numéricas para sumar.
- Se agregan contadores para facilitar las consultas.



EJEMPLO TABLAS DE HECHOS SIN MEDIDAS

