

Normalización de datos



Objetivo principal

Reducir al mínimo datos repetidos o redundantes.

Es un proceso de diseño y reglas llamadas normales.

Es de fácil edición y mantenimiento (de fácil reutilización)

¿Porque es importante?

Los analistas Bi necesitan estar en la capacidad de **normalizar**, **denormalizar** y transformar datos en un modelo **dimensional**. (los datos de origen a veces no están muy bien estructurados y debemos estar en la capacidad de solucionarlo)

También porque tenemos que saber de porqué y el cómo se implementan relaciones y estos tipos de modelos dimensionales.

Formas de Normalización

Tabla Denormalizada (clientes)

ID_CLIENTE	NOMBRE_CLIENTE	CIUDAD_CLIENTE	TELEFONO 1	TELEFONO 2	TELEFONO 3	CANT	VR UNIT	TOTAL
1	ZAPATERIA LA VACA	BOGOTA	555444	888222	444999	5	1000	5000
2	INVERSIONES TERRAMODA	BOGOTA	1234	9876		3	2000	6000
3	DISTRIBUIDORA JCP	BUCARAMANGA	65158, 31151194			2	1000	2000

- Se tienen tres formas...

1er caso

ID_CLIENTE	NOMBRE_CLIENTE
1	ZAPATERIA LA VACA
2	INVERSIONES TERRAMODA
3	DISTRIBUIDORA JCP

ID_CLIENTE	ID_CONTACTOS	CONTACTOS
1	1	555444
1	2	888222
1	3	444999
2	4	1234
2	5	9876
3	6	65158
3	7	31151194

- Las tablas deben ser de registros únicos.
- Cada columna contiene un solo valor.
- No se repiten grupos.

2do caso

ID_CLIENTE	NOMBRE_CLIENTE	CIUDAD_CLIENTE
1	ZAPATERIA LA VACA	BOGOTA
2	INVERSIONES TERRAMODA	BOGOTA
3	DISTRIBUIDORA JCP	BUCARAMANGA

ID_CIUADAD	CIUDAD
1	BOGOTA
2	BUCARAMANGA

- Un campo o columna debe ser dependiente enteramente de una llave primaria.

3er caso

ID_CLIENTE	NOMBRE_CLIENTE	CIUDAD_CLIENTE	TELEFONO 1	TELEFONO 2	TELEFONO 3	CANT	VR UNIT	TOTAL
1	ZAPATERIA LA VACA	BOGOTA	555444	888222	444999	5	1000	5000
2	INVERSIONES TERRAMODA	BOGOTA	1234	9876		3	2000	6000
3	DISTRIBUIDORA JCP	BUCARAMANGA	65158, 31151194			2	1000	2000

- Ningún campo no clave DEPENDERÁ de otro campo no clave.
- (Teórico), en la actualidad como existe mayor capacidad de procesamiento se puede hacer.

Resultado...

- Tabla Denormalizada (clientes)

ID_CLIENTE	NOMBRE_CLIENTE	CIUDAD_CLIENTE	TELEFONO 1	TELEFONO 2	TELEFONO 3	CANT	VR UNIT	TOTAL
1	ZAPATERIA LA VACA	BOGOTA	555444	888222	444999	5	1000	5000
2	INVERSIONES TERRAMODA	BOGOTA	1234	9876		3	2000	6000
3	DISTRIBUIDORA JCP	BUCARAMANGA	65158, 31151194			2	1000	2000

- Tabla Normalizada

ID_CLIENTE	NOMBRE_CLIENTE	ID CIUDAD	ID CONTACTO	CANT	VR UNIT
1	ZAPATERIA LA VACA	1	1	5	\$ 1.000
2	INVERSIONES TERRAMODA	1	4	3	\$ 2.000
3	DISTRIBUIDORA JCP	2	6	2	\$ 1.000

Resumiendo

Normalización

ID_CLIENTE	NOMBRE_CLIENTE	ID CIUDAD	ID CONTACTO	CANT	VR UNIT
1	ZAPATERIA LA VACA	1	1	5	\$ 1.000
2	INVERSIONES TERRAMODA	1	4	3	\$ 2.000
3	DISTRIBUIDORA JCP	2	6	2	\$ 1.000

- Orientado a Maquinas (servidores).
- Relaciones tienden a ser complejas entre tablas.
- Procesamiento rápido de transacciones.
- Elimina redundancia de datos.

Denormalización

ID_CLIENTE	NOMBRE_CLIENTE	CIUDAD_CLIENTE	TELEFONO 1	TELEFONO 2	TELEFONO 3	CANT	VR UNIT	TOTAL
1	ZAPATERIA LA VACA	BOGOTA	555444	888222	444999	5	1000	5000
2	INVERSIONES TERRAMODA	BOGOTA	1234	9876		3	2000	6000
3	DISTRIBUIDORA JCP	BUCARAMANGA	65158, 31151194			2	1000	2000

- Orientados a humanos.
- Relaciones simples. (Ejem: todo el modelo puede estar en una tabla)
- Velocidad de respuesta al usuario. (ejemplo de esto es que cada tabla puede incluir montos calculados)
- Utilizamos redundancia para simplificar el modelo.