Normalización de datos

Proceso de estructurar una base de datos para *minimizar la redundancia* y asegurar la *integridad* de los datos.

Normalización

Se logra mediante la división de una base de datos en tablas más pequeñas y la definición de relaciones entre ellas.

Sigue una serie de reglas conocidas como *formas normales* (normal forms).

Minimizar la redundancia: Reducir al mínimo la duplicidad de datos.

Objetivo principal

Asegurar Integridad: Mantener la coherencia de los datos mediante la eliminación de anomalías de inserción, actualización y eliminación.

Mejorar la Organización: Hacer que la base de datos sea más flexible y escalabre, permitiendo la fácil edición y mantenimiento para su reutilización.

¿Porque es importante?

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)

Tenemos que saber del *porqué* y el *cómo* se implementan relaciones y estos tipos de modelos dimensionales.

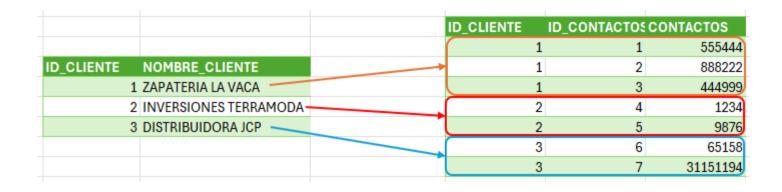
Formas de Normalización

Tabla no normalizada (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	65158,31151			2	1000	2000

• Se tienen tres formas...

1er caso



- 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	ID_CIUDAD	CIUDAD_CLIENTE
1	ZAPATERIA LA VACA	BOGOTA	1	BOGOTA
2	INVERSIONES TERRAMODA	BOGOTA	2	BUCARAMANGA
3	DISTRIBUIDORA JCP	BUCARAMANGA		

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

3er caso

D_CLIENTE 🔽 I	NOMBRE_CLIENTE	CIUDAD_CLIENTE -	TELEFONO 1	TELEFONO 2	TELEFONO 3	CANT VR UNIT	- X	TAL
1 7	ZAPATERIA LA VACA	BOGOTA	555444	888222	444999	5	1000	8000
2	INVERSIONES TERRAMODA	BOGOTA	1234	9876		3	2000	6000
3	DISTRIBUIDORA JCP	BUCARAMANGA	65158,31151			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/No normalizada (clientes)

ID_CLIENTE ▼ NOMBRE_CLIENTE ▼	CIUDAD_CLIENTE	▼ TELEFONO 1 ▼	TELEFONO 2 🔽	TELEFONO 3 🔻	CANT VR UNIT	v	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	65158,31151			2	1000	2000

• Tabla Normalizada

ID_CLIENTE V NOMBRE_CLIENTE	▼ ID_CIUDAD ▼ ID_CONT	TACTO 🔽 CANT	▼ VR L	INIT 🔽
1 ZAPATERIA LA VACA	1	1	5 \$	1.000
2 INVERSIONES TERRAMOD)A 1	4	3 \$	2.000
3 DISTRIBUIDORA JCP	2	6	2 \$	1.000

Resumiendo

Normalización

ID_CLIENTE V NOMBRE_CLIENTE	▼ ID_CIUDAD ▼ ID_CO	NTACTO 🔽 CANT	▼ VR U	NIT 🔽
1 ZAPATERIA LA VACA	1	1	5 \$	1.000
2 INVERSIONES TERRAMOD	A 1	4	3 \$	2.000
3 DISTRIBUIDORA JCP	2	6	2 \$	1.000

- Orientado a máquinas.
- Relaciones tienden a ser complejas entre tablas.
- Procesamiento rápido de transacciones.
- Elimina redundancia de datos.

Denormalización

ID_CLIENTE ▼ NOMBRE_CLIENTE ▼	CIUDAD_CLIENTE	Z TELEFONO 1 ✓	TELEFONO 2 🔽	TELEFONO 3 🔽	CANT VR UNIT	▼ 1	OTAL 🔽
1 ZAPATERIA LA VACA	BOGOTA	555444	888222	444999	5	1000	5000
2 INVERSIONES TERRAMODA	BOGOTA	1234	9876		3	2000	6000
3 DISTRIBUIDORA JCP	BUCARAMANGA	65158,31151			2	1000	2000

- Orientado a humanos.
- Relaciones simples (Ej. Todo el modelo puede estar en una única tabla).
- Velocidad de respuesta al usuario (Cada tabla puede incluir montos calculados)
- Utilización de redundancias para facilidad al usuario.