



Dimensiones lentamente cambiantes: guía práctica

Una guía completa para entender y gestionar los cambios en las dimensiones de datos a lo largo del tiempo, manteniendo la integridad histórica y permitiendo análisis precisos.

¿Qué es una Dimensión?

Una tabla que contiene información descriptiva (contextual o textual) relacionada con un evento o proceso de negocio medido en una tabla de hechos.

Propósito

Responder preguntas acerca de los hechos como:

- ¿Quién? (Cliente, empleado, proveedor)
- ¿Qué? (Producto, servicio)
- ¿Dónde? (Ubicación, tienda)
- ¿Cuándo? (Fecha, hora)
- ¿Cómo y por qué?

Atributos

Columnas descriptivas que proporcionan contexto, usados para:

- Agrupar datos (ventas por marca, categoría)
- Filtrar consultas (ventas por región o por cliente)

Características Técnicas de las Dimensiones

Aspectos Clave

- Tienen una clave primaria independiente de la fuente de datos que se relaciona con las tablas de hechos, denominada **clave subrogada**
- Están desnormalizadas para simplificar la interpretación y el análisis

id_dim_producto	src_id_producto	descripcion_producto	categoria_producto
1	prod1	Limpiador limpito 500 cc	Limpieza
2	prod2	Limpiador limpito 1 litro	Limpieza
3	prod100	Celular último modelo	Electrónica

Ejemplo: Tabla de dimensión de productos

Dimensiones Lentamente Cambiantes

¿Qué es?

Técnica para gestionar cambios en los datos de una dimensión a lo largo del tiempo.

Propósito

- Mantener un historial de cambios o actualizar información en una dimensión
- Permitir análisis precisos y consistentes al considerar los cambios en datos históricos

Tipos de Dimensiones Lentamente Cambiantes

1

Tipo 0

No se realiza ningún cambio en el valor del atributo.

2

Tipo 1

Sobreescribe el valor del atributo. Se pierde el historial pero se mantiene actualizada la información.

3

Tipo 2

Se agrega una nueva fila en la dimensión con el nuevo valor del atributo, manteniendo el historial completo.

4

Tipo 3

Se agrega una columna adicional para registrar el valor actual y el previo del atributo.

5

Tipo 4

Se crea una mini-dimensión para almacenar los atributos que cambian rápidamente.

Ejemplo: ¿Qué pasa si un producto cambia de categoría? Podemos actualizar el atributo (tipo 1) o agregar una fila para mantener historia (tipo 2).

Implementación de SCD Tipo 2

Características

- Al contar con una clave independiente del origen, podemos tener más de un registro para un mismo id de origen
- Se incorporan columnas adicionales para tener una traza precisa de la historia:
 - Fecha de inicio
 - Fecha de expiración
 - Indicador de validez

id_dim	src_id	categoría	fecha_inicio	fecha_fin	es_valido
3	prod100	Electrónica	1900-01-01	2025-01-08	0
4	prod100	Celulares	2025-01-09	9999-12-31	1

Ejemplo: Producto que cambió de categoría "Electrónica" a "Celulares"

Consideraciones para SCD Tipo 2



Fecha de Expiración Inicial

Defina la fecha de expiración inicial como 31 de diciembre de 9999 para evitar valores nulos y asegurar que el registro esté siempre vigente por defecto.



Control de Historial

Al añadir un nuevo registro, la fila anterior debe ser expirada con una fecha anterior para mantener un historial preciso de los cambios.



Opciones de Fecha de Inicio

Considere dos enfoques para la fecha de inicio: usar una fecha estándar como 1 de enero de 1900, o la fecha de creación original del registro de origen.

Implementación SQL de Dimensiones Lentamente Cambiantes

Primer paso: Comparar registros

```
-- Crear tabla temporal para identificar cambios
DROP TABLE IF EXISTS tmp_product_dim_stg;
CREATE TABLE tmp_product_dim_stg AS
SELECT
  src.producto_id,
  src.descripcion, -- SCD Tipo 1
  src.categoria, -- SCD Tipo 2
  trg.id_dim_productos,
  trg.categoria AS old_categoria,
  trg.descripcion AS old_descripcion,
  CAST(CASE
    WHEN trg.id_dim_productos IS NULL THEN 'I'
    WHEN ... THEN 'U'
    ELSE 'S'
  END AS CHAR(1)) AS upd_flag,
  CAST(CASE
    WHEN ... <> ... THEN '2'
    ELSE '1'
  END AS CHAR(1)) AS scd_type,
  CURRENT_DATE AS valid_from,
  '9999-12-31' AS valid_to
FROM stg.stg_productos src
LEFT JOIN dw.dim_productos trg
ON src.producto_id = trg.producto_id
AND trg.valid_now = 1;
```

Segundo paso: Aplicar cambios

```
BEGIN TRANSACTION;

-- Actualización de SCD Tipo 1 (descripcion)
UPDATE dim_productos
SET descripcion = stg.descripcion
FROM tmp_product_dim_stg stg
WHERE stg.upd_flag = 'U' AND stg.scd_type = '1'
AND dim_productos.producto_id = stg.producto_id;

-- Expirar versiones actuales para SCD Tipo 2
UPDATE dim_productos
SET valid_to = CURRENT_DATE - INTERVAL '1 DAY',
  valid_now = '0'
FROM tmp_product_dim_stg stg
WHERE dim_productos.producto_id = stg.producto_id
AND dim_productos.valid_now = '1'
AND stg.upd_flag = 'U' AND stg.scd_type = '2';

-- Insertar nuevos registros y versiones SCD Tipo 2
INSERT INTO dim_productos (...)
SELECT ...
FROM tmp_product_dim_stg stg
WHERE stg.upd_flag = 'I'
OR (stg.upd_flag = 'U' AND stg.scd_type = '2');

COMMIT;
```