

## Inteligencia de Negocios:

### Slowly Changing Dimensions:

Las dimensiones lentamente cambiantes o SCD (Slowly Changing Dimensions) son dimensiones en las cuales sus datos tienden a modificarse a través del tiempo, ya sea de forma ocasional o constante.

Cuando ocurren estos cambios, se puede optar por seguir una de estas dos opciones:

- Registrar el historial de cambios.
- Reemplazar los valores que sean necesarios.

Una de las características de un Data Warehouse es la capacidad de crear análisis de datos históricos, es decir, analizar datos que en el sistema de origen ya no existen porque han sido modificados o eliminados. Para poder realizar esta acción se utiliza el concepto de SCD, esto nos permite hacer un rastreo de los cambios que se han modificado en el sistema de origen.

Existen diversos tipos de cambio

#### → Tipo 0

Este tipo **no admite ningún cambio**, mantiene los valores originales aún cuando existan cambios en el sistema origen.

Ejemplo de cambio en el origen

Source System	ProductID	Product	ReorderPoint
	802	LL Fork	375
↓			
Source System	ProductID	Product	ReorderPoint
	802	LL Fork	<del>375</del> 350

Dimensión con SCD 0

Product Dimension	pk_ProductKey	Nk_ProductID	Product	ReorderPoint
	1	802	LL Fork	375

Los datos insertados en un inicio se mantienen a lo largo del tiempo. Generalmente es usado para atributos que se espera que no cambien como la fecha de nacimiento o el número de seguridad social.

#### → Tipo 1

En este tipo los datos almacenados en el Data Warehouse son **sobreescritos** con los nuevos valores de los atributos cuando ocurre un cambio en el sistema de origen.

Ejemplo de cambio en el origen

Source System	ProductID	Product	ReorderPoint
	802	LL Fork	375

↓

Source System	ProductID	Product	ReorderPoint
	802	LL Fork	<del>375</del> 350

Dimensión con SCD 1

SCD Type 1: Overwrite	pk_ProductKey	Nk_ProductID	Product	ReorderPoint
	1	802	LL Fork	<del>375</del> 350

Los datos insertados en un inicio en la dimensión van cambiando en el tiempo conforme van cambiando en el sistema origen. Generalmente este tipo es usado para atributos que cuyo historial carece de importancia como el número telefónico.

### → Tipo 2

En este tipo **cada cambio en el sistema origen ocasiona un nuevo registro en la dimensión** dentro del Data Warehouse manteniendo el registro con el valor anterior. Cuando se utiliza el SCD 2 es necesario agregar nuevas columnas a la dimensión que nos indicarán el periodo de validez para cada versión del registro:

- **StartDate:** Fecha en que el registro fue insertado. Se considera el punto de inicio de validez de la versión del registro.
- **EndDate:** Fecha en que el registro pierde validez. Generalmente se utiliza el valor NULL para indicar que es el registro con validez actual, aunque también se puede optar por una fecha estándar (31/12/9999).
- **CurrentFlag:** Opcionalmente se puede utilizar una columna que nos permita visualizar de forma más rápida que registro es el actual.

Ejemplo de cambio en el origen

Source System	ProductID	Product	ReorderPoint
	802	LL Fork	375

↓

Source System	ProductID	Product	ReorderPoint
	802	LL Fork	<del>375</del> 350

Dimensión con SCD 2

SCD Type 2: Row Versioning	pk_ProductKey	Nk_ProductID	Product	ReorderPoint	StartDate	EndDate
	1	802	LL Fork	375	1/1/2019	<del>NULL</del> 2/1/19
	2	802	LL Fork	350	2/2/2019	NULL

Cada que se inserta un nuevo registro en la dimensión la fecha EndDate debe almacenar la fecha anterior al StartDate del nuevo registro.

### → Tipo 3

En este tipo de SCD generalmente **sólo se rastrea el cambio más reciente** por lo que no se cuenta con un historial de cambios. El mismo atributo tendrá diferentes columnas, una

Martin Engelhardt - Rodrigo Nolberque - Micaela Olivera  
columna almacenará el valor actual mientras que las demás almacenarán los valores previos. Las necesidades del negocio definirán cuántos valores previos almacenar.

Ejemplo de cambio en el origen

Source System	ProductID	Product	ReorderPoint
	802	LL Fork	375

↓

Source System	ProductID	Product	ReorderPoint
	802	LL Fork	<del>375</del> 350

Dimensión con SCD 3

SCD Type 3: Column Versioning	pk_ProductKey	Nk_ProductID	Product	ReorderPoint	PrevReorderPoint
	1	802	LL Fork	375	NULL

↓

SCD Type 3: Column Versioning	pk_ProductKey	Nk_ProductID	Product	ReorderPoint	PrevReorderPoint
	1	802	LL Fork	<del>375</del> 350	NULL 375

Cuando un cambio ocurre, el valor actual se modifica con el nuevo valor y los valores previos son modificados con los valores previos más recientes.

### → Tipo 6

Como regla general las dimensiones **combinarán diferentes tipos de Slowly Changing Dimension** en sus atributos. Un atributo puede ser SCD 2 y el resto SCD 1 o SCD 0.

Ejemplo donde el color es SCD 2 y ReorderPoint SCD 1

Source System	ProductID	Product	Color	ReorderPoint
	802	LL Fork	Black	500

↓

Source System	ProductID	Product	Color	ReorderPoint
	802	LL Fork	<del>Black</del> Gray	<del>500</del> 550

Dimensión con múltiple SCD

Dimension Table						
pk_ProductKey	Nk_ProductID	Product	Color	ReorderPoint	StartDate	EndDate
1	802	LL Fork	Black	<del>500</del> 550	1/1/2019	<del>NULL</del> 2/1/2019
2	802	LL Fork	Gray	<del>500</del> 550	2/2/2019	NULL

El cambio de color ocasionó la inserción de un nuevo registro mientras que el cambio de punto de reorden ocasiono la actualización de los valores en todos los registros. Un enfoque diferente para las columnas de tipo SCD 1 es sólo actualizar el registro más actual y dejar los anteriores con el valor previo, aunque el estándar es actualizar todos los registros.

En la otra página vimos que hay 6 tipos, básicamente es lo mismo, pero notamos el tipo 4 que agrega el historial y el 6 que es la combinación del tipo 1,2 y 3 mencionado en la primera fuente.

- Type 0 - The passive method
- Type 1 - Overwriting the old value
- Type 2 - Creating a new additional record
- Type 3 - Adding a new column
- Type 4 - Using historical table
- Type 6 - Combine approaches of types 1,2,3 (1+2+3=6)

Fuentes:

<https://www.aprendiendodatos.com/slowly-changing-dimension-scd-0123-datawarehouse/>  
<https://www.datawarehouse4u.info/SCD-Slowly-Changing-Dimensions.html>