

# Modelado Dimensional: técnicas y mejores prácticas

Una guía completa sobre las técnicas de modelado dimensional para sistemas de inteligencia de negocios efectivos.

# Conceptos fundamentales

1

## Recopilar requisitos del negocio

Antes de iniciar el modelado dimensional, el equipo debe comprender las necesidades del negocio y las realidades de los datos subyacentes mediante sesiones con representantes del negocio y expertos en sistemas fuente.

2

## Talleres colaborativos

Los modelos dimensionales deben diseñarse en colaboración con expertos en la materia y representantes de gobierno de datos del negocio, no de forma aislada por personas que no comprenden completamente el negocio.

Estos conceptos fundamentales establecen la base para un modelado dimensional exitoso, asegurando que el sistema final satisfaga las necesidades analíticas reales del negocio.

# Proceso de diseño dimensional de cuatro pasos

## Paso 1: Seleccionar el proceso de negocio

Los procesos de negocio son las actividades clave que impulsan una organización, como tomar pedidos, facturar o atender llamadas. Algunas características clave:

- Son verbos de acción que representan las operaciones del negocio
- Están respaldados por sistemas informáticos que capturan datos valiosos
- Generan métricas clave que los analistas quieren analizar exhaustivamente
- Se encadenan, con las salidas de un proceso convirtiéndose en las entradas del siguiente

## Paso 2: Declarar la granularidad

La granularidad define el nivel de detalle de cada fila en la tabla de hechos, respondiendo a la pregunta "¿Cómo se describe una sola fila?". Algunos ejemplos:

- Una fila por cada escaneo de producto en una transacción de ventas
- Una fila por cada artículo de factura de un médico
- Una fila por cada tarjeta de embarque escaneada en un aeropuerto

La granularidad se expresa en términos de negocio, no solo enumerando dimensiones. Es fundamental definir correctamente la granularidad.

## Paso 3: Identificar las Dimensiones

Las dimensiones son los atributos descriptivos que permiten a los usuarios de negocio comprender y analizar los datos resultantes de los eventos del proceso de negocio. Son esenciales para "decorar" las tablas de hechos con un contexto significativo.



### Quién

Identifica a las personas involucradas, como clientes, empleados o vendedores.



### Qué

Describe los productos, servicios o elementos de interés.



### Dónde

Indica la ubicación, como tiendas, regiones o sucursales.



### Cuándo

Especifica el momento del evento, como fecha, hora o período fiscal.



### Por Qué

Aclara las razones o propósitos detrás de un evento, como las promociones aplicadas.



### Cómo

Detalla el método o las condiciones, por ejemplo, el canal de venta.

Una vez identificadas las dimensiones clave (quién, qué, dónde, cuándo, por qué y cómo), se enumeran todos los atributos discretos y textuales que conformarán cada tabla de dimensión, como "nombre del cliente", "categoría de producto" o "ciudad de la tienda".

## Paso 4: Identificar los hechos

### Definir los hechos

Responde a la pregunta: "¿Qué está midiendo el proceso?"

1

### Datos numéricos

Los hechos típicos son cifras aditivas como cantidades y costos

2

3

### Métricas clave

Los usuarios de negocio quieren analizar estas métricas de rendimiento

#### ▼ Evitar el enfoque impulsado por datos

No te enfoques solo en los datos fuente. Debes considerar los requisitos de los usuarios de negocio también. Un enfoque puramente basado en datos a menudo fracasa.

# Caso de estudio: ventas minoristas

Imaginen que trabajan en la sede central de una gran cadena de supermercados. El negocio cuenta con 100 tiendas distribuidas en cinco estados. Cada tienda tiene una gama completa de departamentos, incluyendo comestibles, alimentos congelados, lácteos, carnes, productos frescos, panadería, florería y artículos de salud/belleza.

Cada tienda tiene aproximadamente 60,000 productos individuales, conocidos como unidades de mantenimiento de existencias (SKU), en sus estantes. Se recopilan datos en varios puntos interesantes de un supermercado.

Algunos de los datos más útiles se recolectan en las cajas registradoras, a medida que los clientes compran productos. El sistema de punto de venta (POS) escanea los códigos de barras de los productos en la caja registradora, midiendo la salida de mercancía de consumo por la puerta principal del supermercado, como se ilustra en el recibo de caja de la imagen. Otros datos se capturan en la puerta trasera de la tienda, donde los proveedores realizan las entregas.

En el supermercado, la gerencia se ocupa de la logística de pedidos, almacenamiento y venta de productos, mientras maximiza las ganancias. La ganancia finalmente proviene de cobrar lo máximo posible por cada producto, reducir los costos de adquisición de productos y los gastos generales, y al mismo tiempo atraer a tantos clientes como sea posible en un entorno altamente competitivo. Algunas de las decisiones de gestión más significativas tienen que ver con los precios y las promociones. Tanto la gerencia de la tienda como el departamento de marketing de la sede dedican mucho tiempo a ajustar los precios y las promociones.

Las promociones en un supermercado incluyen reducciones temporales de precios, anuncios en periódicos e insertos de periódicos, exhibiciones en la tienda de comestibles y cupones. La forma más directa y efectiva de generar un aumento en el volumen de productos vendidos es bajar drásticamente el precio. Una reducción de 50 centavos en el precio de las toallas de papel, especialmente cuando se combina con un anuncio y una exhibición, puede hacer que la venta de toallas de papel se multiplique por 10. Desafortunadamente, una reducción de precio tan grande generalmente no es sostenible porque es probable que las toallas se vendan con pérdidas. Como resultado de estos problemas, la visibilidad de todas las formas de promoción es una parte importante del análisis de las operaciones de un supermercado. Ahora que hemos descrito nuestro estudio de caso de negocio, comenzaremos a diseñar el modelo dimensional.

Allstar Grocery 123 Loon Street Green Prairie, MN 55555 (952) 555-1212	
Store: 0022 Cashier: 00245409/Alan	
0030503347 Baked Well Multigrain Muffins	2.50
2120201195 Diet Cola 12-pack Saved \$.50 off \$5.49	4.99
0070806048 Sparkly Toothpaste Coupon \$.30 off \$2.29	1.99
2840201912 SoySoy Milk Quart	3.19
TOTAL	12.67
AMOUNT TENDERED CASH	12.67
ITEM COUNT:	4
-----	
Transaction: 649	4/15/2013 10:56 AM
-----	
Thank you for shopping at Allstar	
0064900220415201300245409	

# El proceso de cuatro pasos

1

## Seleccionar el proceso de negocio

El primer paso es decidir qué proceso de negocio modelar. Esto implica entender tanto las necesidades del negocio como los datos fuente disponibles.

- ❑ El primer proyecto debe enfocarse en el proceso más crítico para los usuarios y más viable (disponibilidad de datos, calidad y preparación organizacional).

En nuestro caso la gerencia quiere entender mejor las compras de los clientes a través del sistema POS. Por lo tanto, el proceso a modelar son las **transacciones de ventas minoristas POS**.

2

## Declarar la granularidad

Una vez identificado el proceso, se define el nivel de detalle de los datos en el modelo dimensional. La clave es ir al detalle más bajo o "atómico" para maximizar la flexibilidad analítica.

- ❑ Desarrollar modelos dimensionales con la información más detallada y atómica posible, ya que esto permite mayor flexibilidad.

En el caso de estudio, la granularidad más fina es un **producto individual en una línea de transacción POS**. Aunque los usuarios no siempre necesiten ver cada artículo, el detalle atómico permite responder a preguntas complejas (ej., diferencias de ventas entre lunes y domingo, impacto de promociones).

3

## Identificar las dimensiones

Una vez elegida la granularidad de la tabla de hechos, la selección de las dimensiones es más clara. Las dimensiones son los atributos descriptivos que contextualizan los hechos.

- ❑ Una declaración precisa de la granularidad determina las dimensiones principales. Solo agregue dimensiones adicionales si naturalmente toman un único valor para cada combinación de las dimensiones principales; de lo contrario, reevalúe.

Para nuestro caso de estudio, las dimensiones descriptivas son: **fecha, producto, tienda, promoción, cajero y método de pago**. También se incluye el número de ticket de transacción POS como una dimensión especial (dimensión degenerada).

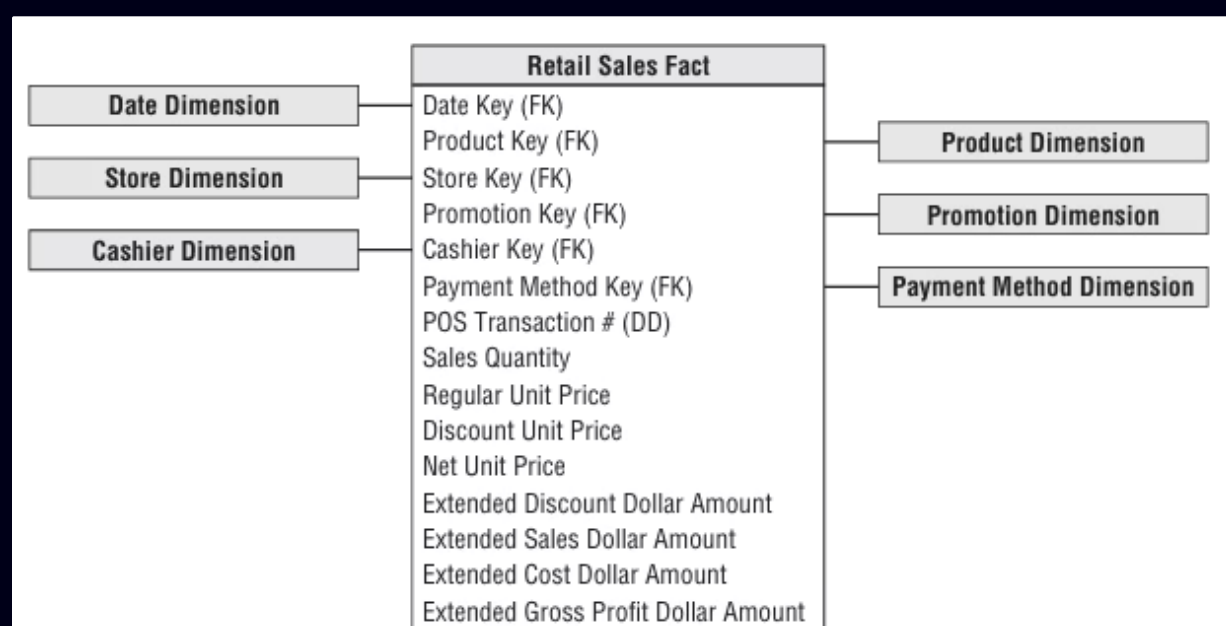
4

## Identificar los hechos

El último paso es determinar cuidadosamente qué hechos aparecerán en la tabla de hechos, siempre manteniendo la coherencia con la granularidad declarada.

Los hechos deben ser "verdaderos" a la granularidad: en este caso, la línea de artículo de producto individual en la transacción POS.

Los hechos típicos del sistema POS incluyen: **cantidad de venta, precio unitario regular, descuento, precios netos pagados, y montos en dólares de descuento y ventas extendidas**. Si está disponible, también se puede incluir el **costo extendido del producto**.



### Hechos aditivos

Cantidades de ventas, descuentos, ingresos y costos son fácilmente sumables a lo largo de todas las dimensiones

### Métricas no aditivas

Métricas como el precio unitario y el margen de ganancia no se pueden sumar directamente

1

2

3

## Margen bruto

Aunque es un hecho calculado, el margen bruto también es perfectamente aditivo

**1 Almacenar métricas calculadas como hechos**

**2 Calcular ratios y porcentajes en la herramienta BI**

**3 Promediar métricas no aditivas con cuidado**

### ▼ Consideraciones sobre Hechos No Aditivos

Algunos cuestionan si los hechos no aditivos como el precio unitario deben almacenarse físicamente en la tabla de hechos, ya que tienen un valor analítico limitado. Sin embargo, en algunos casos, estos hechos pueden ser útiles para filtrar o mostrar valores individuales en los informes.

# Dimensión Fecha

La dimensión fecha es especial porque está prácticamente garantizada en cada modelo dimensional. Características importantes:

- Puede construirse por adelantado con 10-20 años de filas
- Incluso 20 años de días son solo aproximadamente 7,300 filas
- Debe incluir atributos para períodos calendario y fiscales
- Los indicadores deben tener valores significativos (ej. "Día Festivo" en lugar de "Y")

Date Key	Date	Full Date Description	Day of Week	Calendar Month	Calendar Quarter	Calendar Year	Fiscal Year-Month	Holiday Indicator	Weekday Indicator
20130101	01/01/2013	January 1, 2013	Tuesday	January	Q1	2013	F2013-01	Holiday	Weekday
20130102	01/02/2013	January 2, 2013	Wednesday	January	Q1	2013	F2013-01	Non-Holiday	Weekday
20130103	01/03/2013	January 3, 2013	Thursday	January	Q1	2013	F2013-01	Non-Holiday	Weekday
20130104	01/04/2013	January 4, 2013	Friday	January	Q1	2013	F2013-01	Non-Holiday	Weekday
20130105	01/05/2013	January 5, 2013	Saturday	January	Q1	2013	F2013-01	Non-Holiday	Weekday
20130106	01/06/2013	January 6, 2013	Sunday	January	Q1	2013	F2013-01	Non-Holiday	Weekday
20130107	01/07/2013	January 7, 2013	Monday	January	Q1	2013	F2013-01	Non-Holiday	Weekday
20130108	01/08/2013	January 8, 2013	Tuesday	January	Q1	2013	F2013-01	Non-Holiday	Weekday

Date Dimension
Date Key (PK)
Date
Full Date Description
Day of Week
Day Number in Calendar Month
Day Number in Calendar Year
Day Number in Fiscal Month
Day Number in Fiscal Year
Last Day in Month Indicator
Calendar Week Ending Date
Calendar Week Number in Year
Calendar Month Name
Calendar Month Number in Year
Calendar Year-Month (YYYY-MM)
Calendar Quarter
Calendar Year-Quarter
Calendar Year
Fiscal Week
Fiscal Week Number in Year
Fiscal Month
Fiscal Month Number in Year
Fiscal Year-Month
Fiscal Quarter
Fiscal Year-Quarter
Fiscal Half Year
Fiscal Year
Holiday Indicator
Weekday Indicator
SQL Date Stamp
...

# Dimensión Producto

Product Key	Product Description	Brand Description	Subcategory Description	Category Description	Department Description	Fat Content
1	Baked Well Light Sourdough Fresh Bread	Baked Well	Fresh	Bread	Bakery	Reduced Fat
2	Fluffy Sliced Whole Wheat	Fluffy	Pre-Packaged	Bread	Bakery	Regular Fat
3	Fluffy Light Sliced Whole Wheat	Fluffy	Pre-Packaged	Bread	Bakery	Reduced Fat
4	Light Mini Cinnamon Rolls	Light	Pre-Packaged	Sweeten Bread	Bakery	Non-Fat
5	Diet Lovers Vanilla 2 Gallon	Coldpack	Ice Cream	Frozen Desserts	Frozen Foods	Non-Fat
6	Light and Creamy Butter Pecan 1 Pint	Freshlike	Ice Cream	Frozen Desserts	Frozen Foods	Reduced Fat
7	Chocolate Lovers 1/2 Gallon	Frigid	Ice Cream	Frozen Desserts	Frozen Foods	Regular Fat
8	Strawberry Ice Creamy 1 Pint	Icy	Ice Cream	Frozen Desserts	Frozen Foods	Regular Fat
9	Icy Ice Cream Sandwiches	Icy	Novelties	Frozen Desserts	Frozen Foods	Regular Fat

## 1 Atributos descriptivos ricos

La tabla de dimensión de producto puede tener 50 o más atributos descriptivos, cada uno siendo una fuente para filtrar y etiquetar los datos

## 2 Navegación jerárquica y flexible

Los usuarios pueden "perforar hacia abajo" en los informes, agregando atributos a los encabezados de fila, más allá de la jerarquía predefinida

## 3 Almacenamiento flexible de valores numéricos

Los valores numéricos como el precio de lista estándar deben almacenarse donde mejor sirvan al análisis: en la tabla de hechos si se usan para cálculos, o en la dimensión si se usan para filtrar y agrupar

## 4 Preservar significado embebido

Los códigos de producto operativos con significado incrustado deben preservarse en su totalidad en la tabla de dimensión, y también desglosarse en atributos separados

### ▼ Simplicidad de desarrollo


Los modelos dimensionales deben ser consistentes y el desarrollo de aplicaciones debe ser predeciblemente sencillo. Los datos utilizados en cálculos deben estar en las tablas de hechos, y los datos utilizados en restricciones, agrupaciones y etiquetas deben estar en las tablas de dimensiones.

# Claves subrogadas

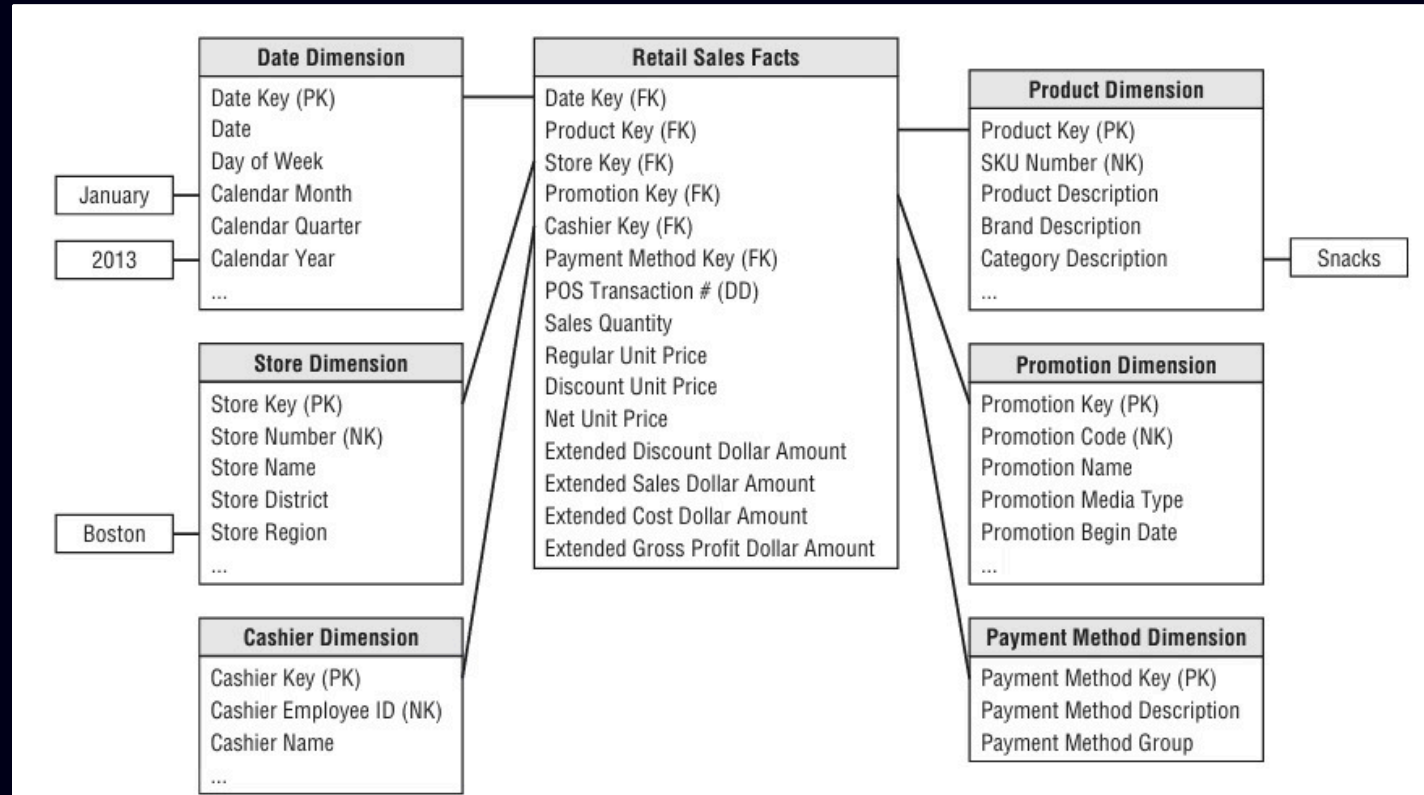
La clave primaria única de una tabla de dimensión debe ser una clave subrogada en lugar de depender del identificador operativo natural.

Ventajas:

- Protege el Data Warehouse de cambios operativos
- Integra múltiples sistemas fuente
- Maneja condiciones nulas o desconocidas
- Soporta el seguimiento de cambios en atributos dimensionales
- **Pueden implementarse con enteros autoincrementales o hashes**

 Cada join entre tablas de dimensión y hechos debe basarse en claves subrogadas sin significado. Debe evitarse usar una clave natural como clave primaria de la tabla de dimensión.

# Modelo dimensional de ventas



# Tipos de tablas de hechos

Existen tres tipos fundamentales de tablas de hechos, independientemente de la industria. Cada tipo tiene un propósito útil y características distintas.

	<b>Transaccional</b>	<b>Foto (Snapshot) periódica</b>	<b>Foto acumulativa</b>
<b>Periodicidad</b>	Punto de tiempo de transacción discreto	Instantáneas recurrentes a intervalos regulares, predecibles	Lapso de tiempo indeterminado para flujo de trabajo/pipeline en evolución
<b>Granularidad</b>	1 fila por transacción o línea de transacción	1 fila por período de instantánea más otras dimensiones	1 fila por ocurrencia de pipeline
<b>Dimensión tiempo</b>	Fecha de transacción	Fecha de instantánea	Múltiples fechas para hitos clave del pipeline
<b>Hechos</b>	Rendimiento de transacción	Rendimiento acumulativo para el intervalo de tiempo	Rendimiento para la ocurrencia del pipeline
<b>Dispersión de la tabla de hechos</b>	Dispersa o densa, según la actividad	Predeciblemente densa	Dispersa o densa, según la ocurrencia del pipeline
<b>Actualizaciones de la tabla de hechos</b>	No hay actualizaciones, a menos que sea corrección de errores	No hay actualizaciones, a menos que sea corrección de errores	Actualizada cada vez que ocurre actividad en el pipeline

# Arquitectura Bus

Para el éxito a largo plazo del DW, se necesita un enfoque arquitectónico e incremental para construir el almacén de datos de la empresa.

## Concepto de Bus

Un bus es una estructura común a la que todo se conecta. Al definir una interfaz de bus estándar, los modelos dimensionales separados pueden ser implementados por diferentes grupos en diferentes momentos y aún así funcionar juntos.

## Matriz de Bus

La matriz de bus documenta y comunica la arquitectura, con procesos de negocio como filas y dimensiones comunes como columnas. Las celdas sombreadas indican qué dimensiones se relacionan con cada proceso.

## Beneficios

Proporciona un marco arquitectónico que guía el diseño general, pero divide el problema en fragmentos de procesos de negocio que pueden implementarse en plazos realistas.

BUSINESS PROCESSES	COMMON DIMENSIONS						
	Date	Product	Warehouse	Store	Promotion	Customer	Employee
Issue Purchase Orders	X	X	X				
Receive Warehouse Deliveries	X	X	X				X
Warehouse Inventory	X	X	X				
Receive Store Deliveries	X	X	X	X			X
Store Inventory	X	X		X			
Retail Sales	X	X		X	X	X	X
Retail Sales Forecast	X	X		X			
Retail Promotion Tracking	X	X		X	X		
Customer Returns	X	X		X	X	X	X
Returns to Vendor	X	X		X			X
Frequent Shopper Sign-Ups	X			X		X	X

# Dimensiones Conformadas

## Dimensiones conformadas idénticas

Las dimensiones conformadas idénticas tienen claves, nombres de atributos, definiciones y valores consistentes a través de todas las tablas de hechos. Esto permite análisis integrados y reportes coherentes.

## Dimensiones conformadas reducidas

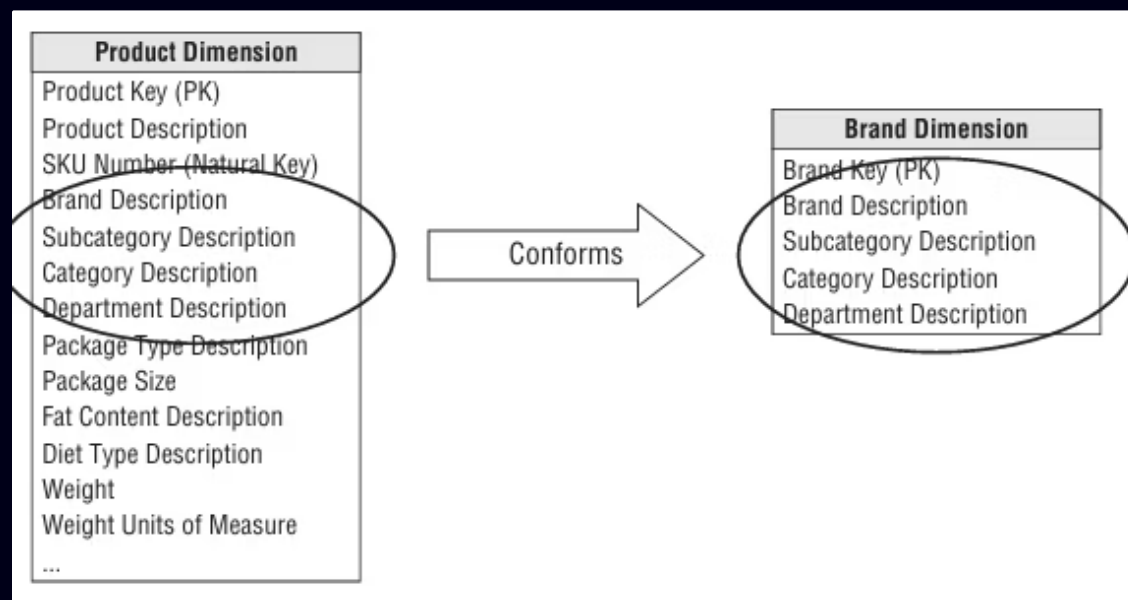
Cuando una tabla de hechos tiene una granularidad más alta que la dimensión base, se puede usar una dimensión conformada reducida que contenga solo un subconjunto de atributos.

## Subconjuntos conformados

Las dimensiones también pueden conformarse cuando representan solo un subconjunto de filas de una dimensión más amplia. Esto permite a los usuarios centrarse en sus líneas de negocio específicas.

### ▼ La clave de la integración

Las dimensiones conformadas son fundamentales para integrar los diferentes procesos de negocio en un almacén de datos coherente. Al garantizar que las definiciones, atributos y valores de las dimensiones sean idénticos, se facilita el análisis transversal y se evitan resultados inesperados.



# Importancia de la Gobernanza de datos

## Desafío clave

Lograr consenso organizacional sobre nombres, definiciones y contenidos de los atributos dimensionales es crucial para la integración de datos en el almacén empresarial.

1

## Objetivos clave

El gobierno de datos debe establecer definiciones, etiquetas y valores maestros comunes para asegurar que "todos hablen el mismo idioma".

2

3

## Liderazgo necesario

El liderazgo del gobierno de datos debe provenir de expertos de negocio, no solo del equipo de TI, para desafiar el status quo y construir consenso.

### ▼ Cómo Impulsa la Agilidad

**Las dimensiones conformadas facilitan el desarrollo ágil de DW/BI al permitir reutilizar dimensiones existentes, acelerando el tiempo de desarrollo. Ignorarlas bajo presión de agilidad resultará en silos analíticos departamentales con categorías inconsistentes, impactando negativamente la toma de decisiones ágil.**

# Dimensiones lentamente cambiantes (SCD)

Los atributos dimensionales no son fijos para siempre; cambian, aunque lentamente, con el tiempo. Los diseñadores dimensionales deben trabajar proactivamente con los representantes de gobierno de datos para determinar la estrategia adecuada para manejar cambios.



## Tipo 0: Mantener original

El valor del atributo dimensional nunca cambia, por lo que los hechos siempre se agrupan por este valor original.

Apropiado para atributos etiquetados como "original" y para la mayoría de los atributos en una dimensión de fecha.



## Tipo 1: Sobrescribir

Se sobrescribe el valor antiguo del atributo en la fila de dimensión, reemplazándolo con el valor actual.

Simple pero pierde todo el historial de cambios de atributos.



## Tipo 2: Agregar nueva fila

Se inserta una nueva fila de dimensión para reflejar el nuevo valor del atributo. Cada fila contiene una versión del perfil de atributos que fue verdadera durante un período de tiempo.

- ❗ Cada atributo de la tabla de dimensión debe tener una estrategia específica para manejar cambios. Se pueden emplear múltiples técnicas dentro de una sola tabla de dimensión.

# Conclusiones



## Proceso de cuatro pasos

El proceso de diseño dimensional de cuatro pasos (seleccionar proceso, declarar granularidad, identificar dimensiones e identificar hechos) proporciona un marco sólido para crear modelos dimensionales efectivos.



## Tipos de tablas de hechos

Los tres tipos fundamentales (transaccional, instantánea periódica y acumulativa) sirven propósitos complementarios y a menudo se necesitan múltiples tipos para obtener una imagen completa del negocio.



## Arquitectura bus

La arquitectura bus con dimensiones conformadas permite un enfoque incremental para construir un sistema de datos integrado, dividiendo el problema en fragmentos manejables.



## Dimensiones de cambio lento

Las técnicas SCD, especialmente el tipo 2, son esenciales para representar correctamente la historia y manejar los cambios en atributos dimensionales de manera que satisfagan las necesidades analíticas del negocio.

