

Diseño Dimensional

Los cuatro pasos del diseño dimensional

1. Seleccionar el proceso de negocio
2. Declarar el nivel de granularidad
3. Identificar las dimensiones
4. Identificar los hechos



1. Identificar procesos de negocio

- Se expresan en verbos. Las dimensiones que lo acompañan brindan contexto a cada evento del proceso.
- Están soportados por sistemas operacionales
- Generan o capturan KPIs
- Se desencadenan de un ingreso de datos y tienen como salida métricas.

¿Los procesos son igual a las funciones de un departamento?

2. Declarar la granularidad

- Es definir lo que representa cada registro de la tabla de hechos.
- Muestra las realidades físicas del sistema operacional que captura los eventos del proceso de negocio.

Ejemplos de definiciones de granularidad:

- Una fila por scan de los productos individuales de una venta a un cliente
- Una fila por línea de detalle de la cuenta de un doctor
- Una fila por cada boleto escaneado en la puerta de un aeropuerto
- Una fila por la instantánea de los niveles de inventario de cada bodega
- Una fila por cuenta de banco por mes

3. Identificar dimensiones

- Se derivan de la pregunta, “**¿Cómo la gente del negocio describe los datos que resultan de los eventos de medición de resultados del proceso?**”
- Al tener clara la granularidad, las dimensiones se pueden identificar ya que representan:
 - **Quién**
 - **Qué**
 - **Dónde**
 - **Cuándo**
 - **Por qué**
 - **Cómo**

4. Identificar los hechos

- Se determinan al contestar la pregunta, “¿**Qué mide este proceso?**”
- Todos los hechos candidatos deben cumplir con la definición de granularidad.
- Los hechos son típicamente valores numéricos.

Ejemplo

Un supermercado

Trabajamos para una cadena de supermercados. El negocio cuenta con 100 tiendas desplegadas en 5 estados. Cada tienda tiene los siguientes departamentos: abarrotes, comida congelada, lácteos, carnes, verduras y vegetales, panadería, floristería y productos de salud/belleza. Cada tienda tiene cerca de 60,000 productos individuales, llamados stock keeping units (SKUs) en sus anaqueles.

Los datos se recolectan en diferentes puntos en la tienda. Algunos de los datos más útiles se recolectan en las cajas registradoras cuando los clientes realizan sus compras de productos. El sistema point of sale (POS) hace un scan de los códigos de barra en la caja registradora, midiendo la venta por cliente en la puerta de la tienda.

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 | 4.99 |
| Saved \$.50 off \$5.49 | |
| 0070806048 Sparkly Toothpaste | 1.99 |
| Coupon \$.30 off \$2.29 | |
| 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 supermercado

Otra data es capturada en la puerta trasera de la tienda en donde los proveedores hacen sus entregas.

En la tienda, la gerencia se preocupa por la logística de ordenar, colocar y vender productos mientras se maximiza la ganancia. La ganancia se obtiene de cobrar lo máximo posible por cada producto, bajar sus costos de adquisición y al mismo tiempo atraer a la mayor cantidad de clientes como sea posible en un entorno altamente competitivo.

Algunas de las decisiones más importantes a tomar tienen que ver con establecimiento de precios y promociones. Tanto la gerencia de la tienda como el equipo corporativo de mercadeo invierten mucho tiempo en precios y promociones.

El supermercado

Las promociones en una tienda incluyen reducciones de precio temporales, anuncios e insertos en periódicos, visuales en la tienda y cupones. La forma más efectiva de aumentar el volumen de ventas de un producto es bajar el precio dramáticamente. Una reducción de 50 centavos en el precio de las toallas de papel, especialmente cuando se apoya con un anuncio y un visual en tienda, pueden causar que las ventas se multipliquen en un factor de 10.

Desafortunadamente, estas reducciones de precios no son sostenibles porque probablemente las toallas se venden con pérdida. Como resultado de esto, la visibilidad de todo tipo de promociones es una parte importante del análisis de las operaciones del supermercado.

Paso 1

Seleccionar el proceso de negocio

Debemos entender el proceso de negocio y las fuentes de datos disponibles.

El primer proyecto de DW/BI debería enfocarse en el proceso más crítico para los usuarios así como el más asequible.

Caso de estudio:

La gerencia quiere entender de mejor forma las compras de los clientes capturadas por el sistema de POS.

Paso 2

Declaración de granularidad

¿Cuál es el nivel de detalle en el que los datos estarán disponibles en el modelo dimensional?

Los datos atómicos proveen la mayor flexibilidad analítica porque pueden ser filtrados y agrupados de muchas formas.

Caso de estudio:

Venta por producto individual en una transacción de POS, asumiendo una sumarización por producto cuando se vende más de un elemento.

Paso 3

Identificar dimensiones

son consultados

La declaración de granularidad determina las dimensiones primarias de la tabla de hechos.

Se agregan más dimensiones si éstas naturalmente toman un valor bajo la combinación de las dimensiones primarias. Si rompe la granularidad se descalifica o se debe revisar la granularidad.

Caso de estudio:

Fecha, producto, tienda, promoción, cajero, método de pago. El ticket de la transacción de POS es una dimensión especial.

Paso 4

Identificar los hechos

son numericos

La definición de granularidad facilita escoger los hechos a analizar.

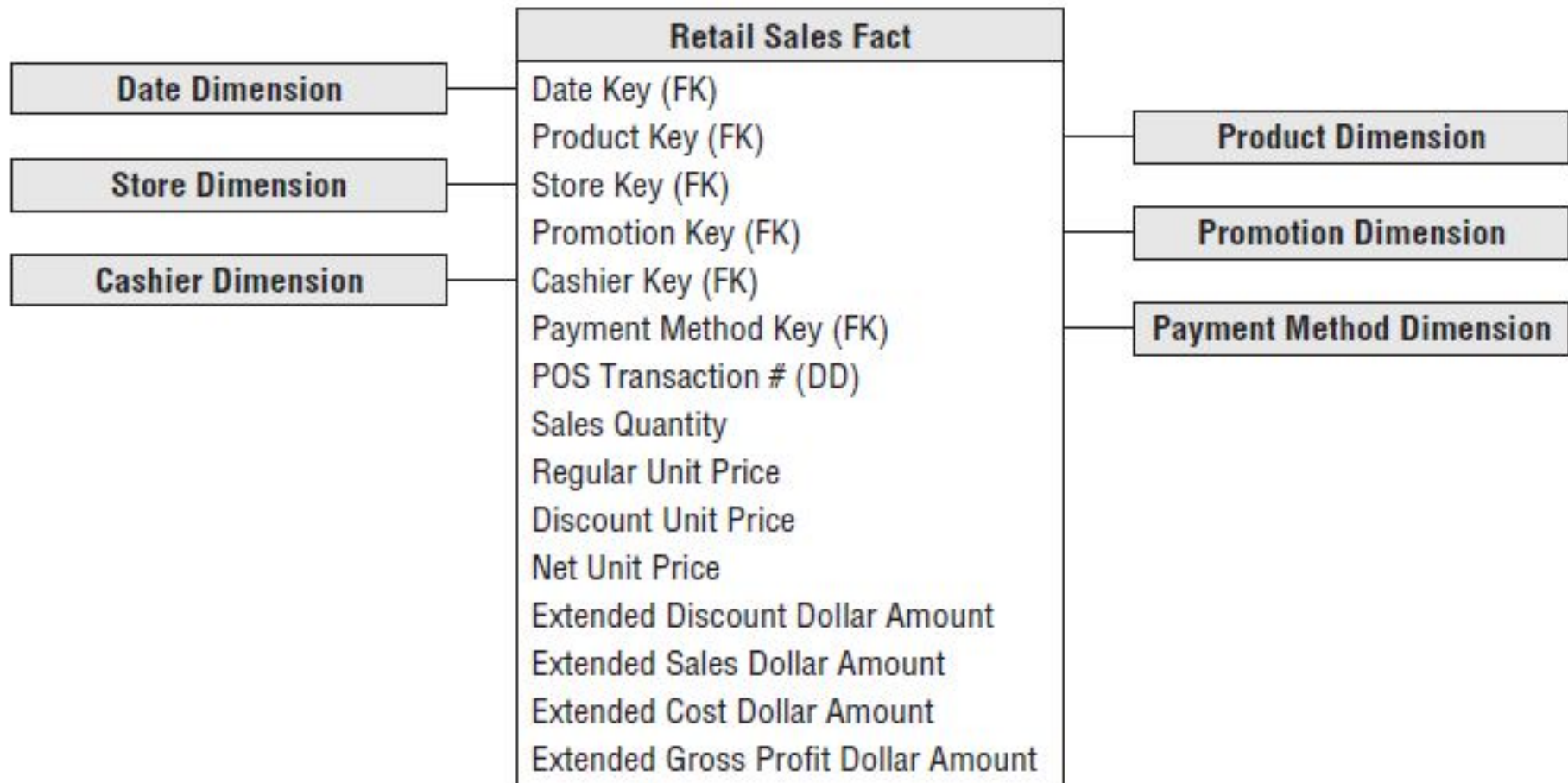
Caso de estudio:

Los hechos que recolecta el POS incluye la cantidad de venta, precio regular, descuento, precio pagado, descuentos extendidos y venta extendida en dólares.

La venta extendida en dólares es igual a la venta en cantidades multiplicada por el precio neto unitario.

El descuento extendido en dólares es la venta en cantidades multiplicada por el monto de descuento por unidad.

Si está disponible, el costo por producto en dólares se puede considerar.



Hechos derivados

esto debe de estar calculado en la tabla de hechos

Hechos que se calculan en función de hechos naturales.

Caso de Estudio:

Ganancia bruto = venta extendida en dólares - costo extendido en dólares

¿Lo calculamos cada vez o lo almacenamos? ¿Tabla o vista?

Para evitar inconsistencias, se calcula en el ETL

Hechos no aditivos

Porcentajes y ratios son no aditivos. El numerador y denominador debe almacenarse como hecho en la tabla de hechos.

| | Fecha | Cantidad | Precio | Venta |
|------------|----------|----------|--------|--------|
| Producto 1 | 3/2/2021 | 1 | \$1.00 | \$1.00 |
| Producto 1 | 3/3/2021 | 4 | \$0.50 | \$2.00 |
| | | 5 | ???? | \$3.00 |

Se debe utilizar un promedio ponderado

Tabla de hechos transaccional

El grano se expresa en el contexto de la transacción, cómo un registro por transacción o un registro por línea de transacción.

La cantidad de registros depende de los eventos transaccionales y puede crecer en volumen rápidamente.

Son altamente dimensionales.

Sus métricas son típicamente aditivas

Estimación del tamaño de la tabla de hechos

Se debe entender la cantidad de registros que se generan en un período de tiempo razonable.

Caso de estudio:

Si la venta bruta son \$4 billones al año y el precio promedio por ticket es \$2.00, se tienen aproximadamente 2 billones de registros.

Tablas Dimensionales

Dimensión de Fechas

Si usamos la fecha de la transacción en la tabla de hechos, ¿Podemos obviar usar una dimensión de fechas?

Hay muchos atributos de fecha que no se pueden calcular en SQL como número de semana, períodos fiscales, temporadas, asuetos, fines de semana. Esta lógica de calendario se debe manejar en la dimensión de fechas.

Los modelos dimensionales necesitan una dimensión de fechas explícita.

| 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 |
| ... |

Consideraciones con el tiempo

La dimensión de fechas es sobre la que más se hacen filtros en el esquema, por lo que debe ser lo más pequeña y manejable posible.

Si se necesita llevar registro de horas, minutos, segundos se debe crear una dimensión separada de tiempo para evitar una explosión de registros en la dimensión de fechas.

Atributos textuales

Se utilizan para filtrar.

Es mejor utilizar valores con significado en el dominio que valores crípticos.

Monthly Sales

Period: June 2013
Product Baked Well Sourdough

Monthly Sales

Period: June 2013
Product Baked Well Sourdough

| Holiday Indicator | Extended Sales Dollar Amount |
|-------------------|------------------------------|
| N | 1,009 |
| Y | 6,298 |

OR

| Holiday Indicator | Extended Sales Dollar Amount |
|-------------------|------------------------------|
| Holiday | 6,298 |
| Non-holiday | 1,009 |

Dimensión de Productos

Describe todos los productos del supermercado.

Se administra por medio de un maestro de productos y se actualiza en intervalos frecuentes.

Se deben aplanar todas las relaciones de 1 a muchos (desnormalizar).

| 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 |

| Product Dimension |
|--------------------------|
| Product Key (PK) |
| SKU Number (NK) |
| Product Description |
| Brand Description |
| Subcategory Description |
| Category Description |
| Department Number |
| Department Description |
| Package Type Description |
| Package Size |
| Fat Content |
| Diet Type |
| Weight |
| Weight Unit of Measure |
| Storage Type |
| Shelf Life Type |
| Shelf Width |
| Shelf Height |
| Shelf Depth |
| ... |

Drill Down

| Department Name | Sales Dollar Amount |
|-----------------|---------------------|
| Bakery | 12,331 |
| Frozen Foods | 31,776 |

Drill down by brand name:

| Department Name | Brand Name | Sales Dollar Amount |
|-----------------|-------------|---------------------|
| Bakery | Baked Well | 3,009 |
| Bakery | Fluffy | 3,024 |
| Bakery | Light | 6,298 |
| Frozen Foods | Coldpack | 5,321 |
| Frozen Foods | Freshlike | 10,476 |
| Frozen Foods | Frigid | 7,328 |
| Frozen Foods | Icy | 2,184 |
| Frozen Foods | QuickFreeze | 6,467 |

Or drill down by fat content:

| Department Name | Fat Content | Sales Dollar Amount |
|-----------------|-------------|---------------------|
| Bakery | Nonfat | 6,298 |
| Bakery | Reduced fat | 5,027 |
| Bakery | Regular fat | 1,006 |
| Frozen Foods | Nonfat | 5,321 |
| Frozen Foods | Reduced fat | 10,476 |
| Frozen Foods | Regular fat | 15,979 |

Dimensión de Tiendas (Geografía)

Describe todas las tiendas de la cadena de supermercados.

Para este caso, representa la dimensión geográfica.

Se puede utilizar las agrupaciones jerárquicas que maneja el negocio así como agrupaciones geográficas.

| Store Dimension |
|------------------------|
| Store Key (PK) |
| Store Number (NK) |
| Store Name |
| Store Street Address |
| Store City |
| Store County |
| Store City-State |
| Store State |
| Store Zip Code |
| Store Manager |
| Store District |
| Store Region |
| Floor Plan Type |
| Photo Processing Type |
| Financial Service Type |
| Selling Square Footage |
| Total Square Footage |
| First Open Date |
| Last Remodel Date |
| ... |

Dimensión de Promociones

Es una dimensión causal porque describe los factores que producen un cambio en las ventas de un producto.

Las condiciones que potencialmente afectan la venta no necesariamente son registrados en el sistema de POS.

Incluye reducciones temporales de precios, puntas de góndola, publicidad y cupones.

Se le considera una **dimensión casual** porque describe factores que se considera afectan las ventas.

¿Qué pasa cuando no hay promociones?

| Promotion Dimension |
|----------------------|
| Promotion Key (PK) |
| Promotion Code |
| Promotion Name |
| Price Reduction Type |
| Promotion Media Type |
| Ad Type |
| Display Type |
| Coupon Type |
| Ad Media Name |
| Display Provider |
| Promotion Cost |
| Promotion Begin Date |
| Promotion End Date |
| ... |

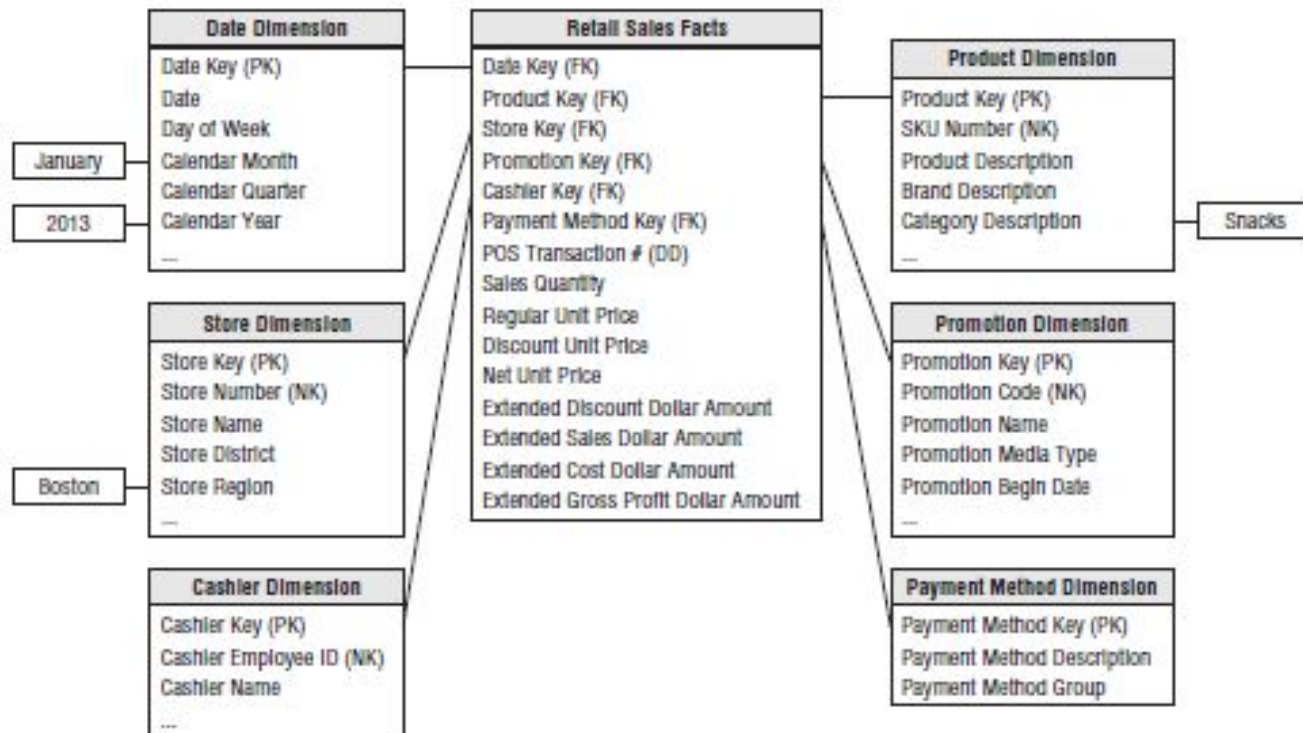
Dimensiones degeneradas - Número de transacción

El número de transacción es útil para agrupar todos los productos comprados en una transacción.

Mantiene la relación con el sistema transaccional.

¿Se debe asignar una llave subrogada a dimensiones degeneradas?

Esquema de ventas



Ejemplo de queries

| Calendar Week Ending Date | Promotion Name | Extended Sales Dollar Amount |
|--------------------------------------|-----------------------|-----------------------------------------|
| January 6, 2013 | No Promotion | 2,647 |
| January 13, 2013 | No Promotion | 4,851 |
| January 20, 2013 | Super Bowl Promotion | 7,248 |
| January 27, 2013 | Super Bowl Promotion | 13,798 |

| Department Name | No Promotion Extended Sales Dollar Amount | Super Bowl Promotion Extended Sales Dollar Amount |
|----------------------------|----------------------------------------------------------|------------------------------------------------------------------|
| January 6, 2013 | 2,647 | 0 |
| January 13, 2013 | 4,851 | 0 |
| January 20, 2013 | 0 | 7,248 |
| January 27, 2013 | 0 | 13,798 |

¿Qué debemos evitar?

Esquema Snowflake con dimensiones normalizadas

Debemos evitar estos modelos basados en desempeño y facilidad de uso.

- Perdemos simplicidad al hacer joins complejos de las tablas normalizadas.
- Estos joins complejos se traducen en menor desempeño de queries.
- El espacio en disco salvado al normalizar es marginal comparado con el tamaño que pueden llegar a tener las tablas de hechos.

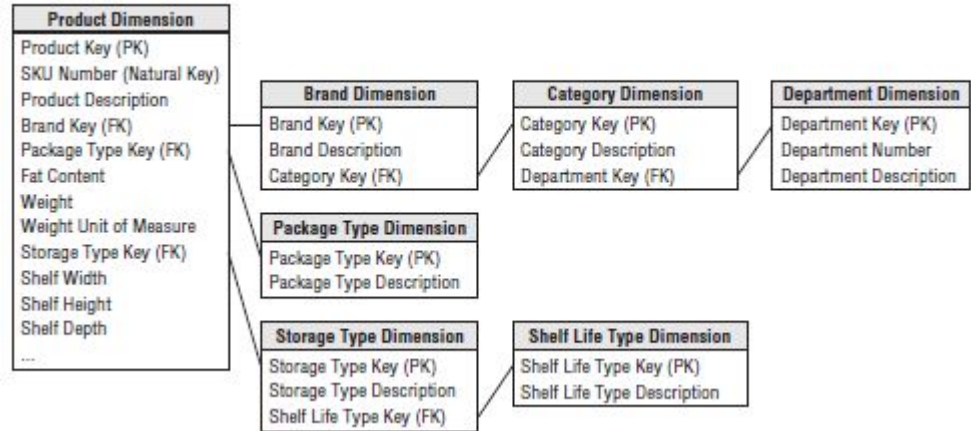
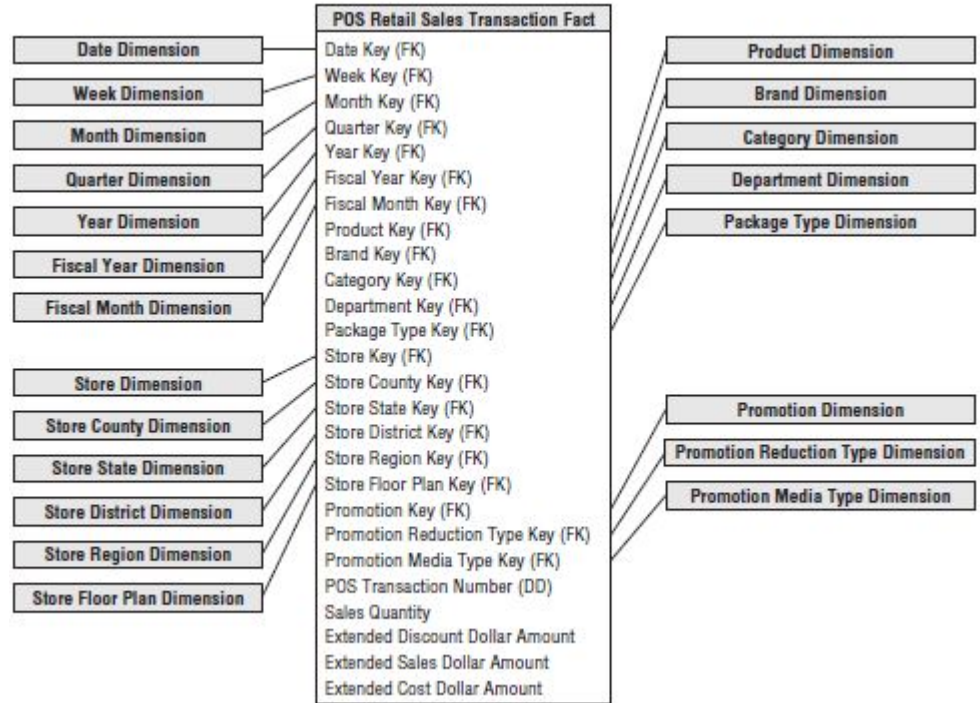


Tabla de hechos ciempiés

- Un gran número de dimensiones de jerarquías correlacionadas deben agruparse en una sola dimensión.



Tipos de fact tables

Transactional Fact Tables

Un registro existe en la tabla de hechos representa un evento que ocurrió en un momento de tiempo específico.

Periodic Snapshot Fact tables

Se utilizan para medir desempeño acumulado en periodos de tiempo regulares.

Accumulating Snapshot Fact Tables

Se utiliza para representar procesos que tienen un inicio y un fin definido con un conjunto de pasos intermedios

Ejemplo de Accumulating Snapshot Fact Tables

| CLAIM KEY | CUSTOMER KEY | POLICY KEY | CLAIM DATE | INVESTIGATION DATE | DAYS TO INVESTIGATION | REVIEW DATE | DAYS TO REVIEW | DESCISION DATE | DAYS TO DECISION | PAYMENT DATE | DAYS TO PAYMENT | PROCESS COMPLETE FLAG |
|-----------|--------------|------------|------------|--------------------|-----------------------|-------------|----------------|----------------|------------------|--------------|-----------------|-----------------------|
| 345623 | 23113 | 5567889 | 1/3/2017 | 5-Jan | 2 | 9-Jan | 6 | 10-Jan | 7 | 13-Jan | 10 | 1 |
| 345624 | 78345 | 4223131 | 1/3/2017 | 6-Jan | 3 | 9-Jan | 6 | 12-Jan | 9 | 17-Jan | 14 | 1 |
| 345625 | 44560 | 7997675 | 1/3/2017 | 9-Jan | 6 | 13-Jan | 10 | 16-Jan | 13 | | | 1 |
| 345626 | 12123 | 2342345 | 1/3/2017 | 4-Jan | 1 | 11-Jan | 8 | 13-Jan | 10 | 17-Jan | 14 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 341123 | 88767 | 2627433 | 2/8/2017 | 9-Feb | 1 | 10-Feb | 2 | 13-Feb | 5 | 14-Feb | 6 | 1 |
| 341124 | 94332 | 5454626 | 2/8/2017 | 9-Feb | 1 | 10-Feb | 2 | 10-Feb | 2 | | | 0 |
| 341125 | 54565 | 4356742 | 2/8/2017 | 9-Feb | 1 | 13-Feb | 5 | | | | | 0 |
| 341126 | 33356 | 2745649 | 2/8/2017 | 13-Feb | 5 | | | | | | | 0 |

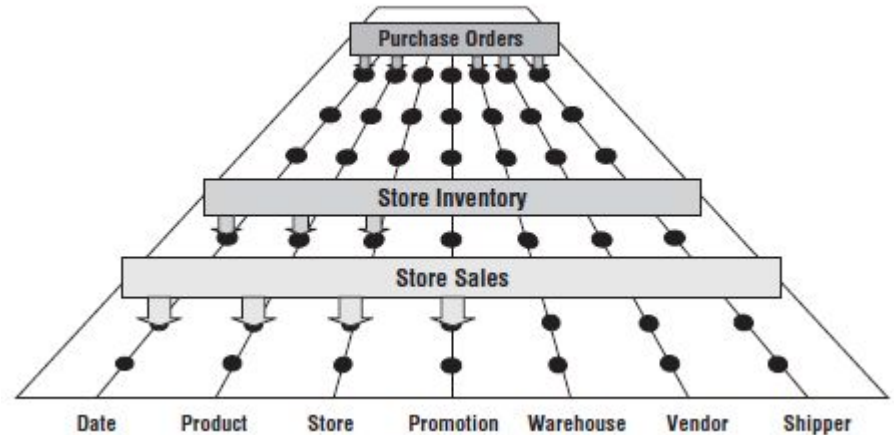
NOTE: The current date is 2/15/2017

| FACT_CLAIM_PROCESSING ▲ | |
|-------------------------|------------------------------|
| P * | CLAIM_KEY NUMBER |
| P * | CUSTOMER_KEY NUMBER |
| P * | POLICY_KEY NUMBER |
| * | CLAIM_DATE DATE |
| | INVESTIGATION_DATE DATE |
| | DAYS_TO_INVESTIGATION NUMBER |
| | REVIEW_DATE DATE |
| | DAYS_TO_REVIEW NUMBER |
| | DESCISION_DATE DATE |
| | DAYS_TO_DESCISION NUMBER |
| | PAYMENT_DATE DATE |
| | DAYS_TO_PAYMENT NUMBER |

Enterprise Data Warehouse Bus Matrix

Diferentes modelos dimensionales pueden ser implementados aprovechando una infraestructura dimensional común.

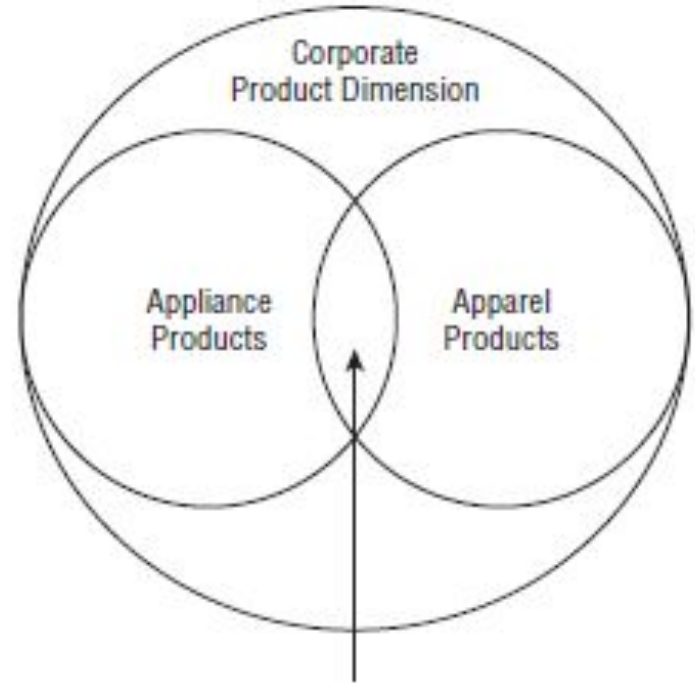
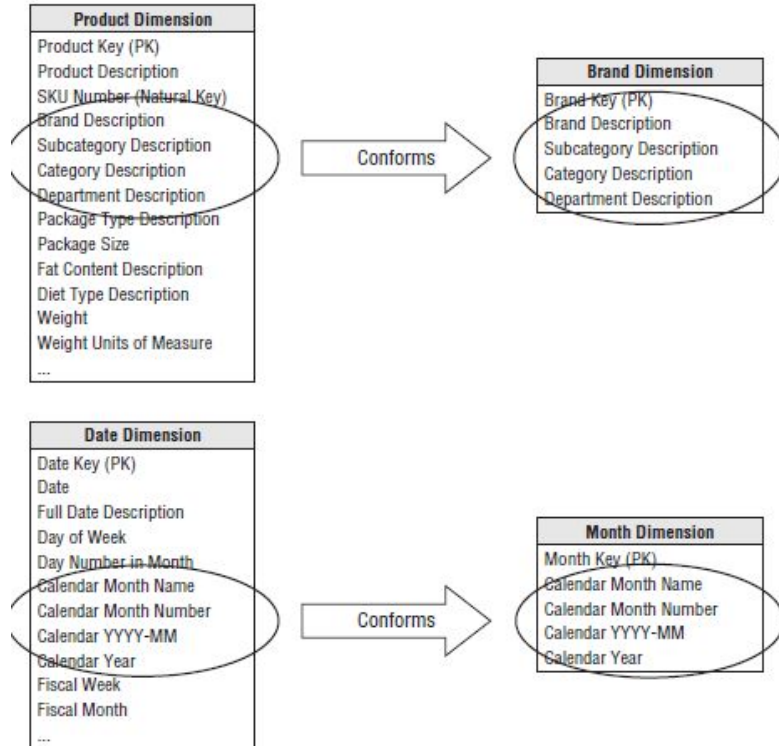
Las columnas del bus representan las dimensiones comunes utilizadas en la organización.



Enterprise Data Warehouse Bus Matrix

| 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



Slow changing dimensions

Las dimensiones no son independientes del tiempo.

A pesar de que los atributos dimensionales son relativamente estáticos, pueden cambiar, aunque lentamente.

Se debe establecer la estrategia apropiada para trabajar con los cambios en atributos que cambian lentamente de cada dimensión.

SCD - Tipo 0: Mantener original

El atributo dimensional nunca cambia.

Usamos llaves subrogadas como llaves primarias, manteniendo como atributo la llave natural.

SCD - Tipo 1: Sobreescribir

Se se sobreescribe el atributo antiguo en la dimensión, reemplazándolo con el valor actual.

El atributo siempre refleja el valor actual.

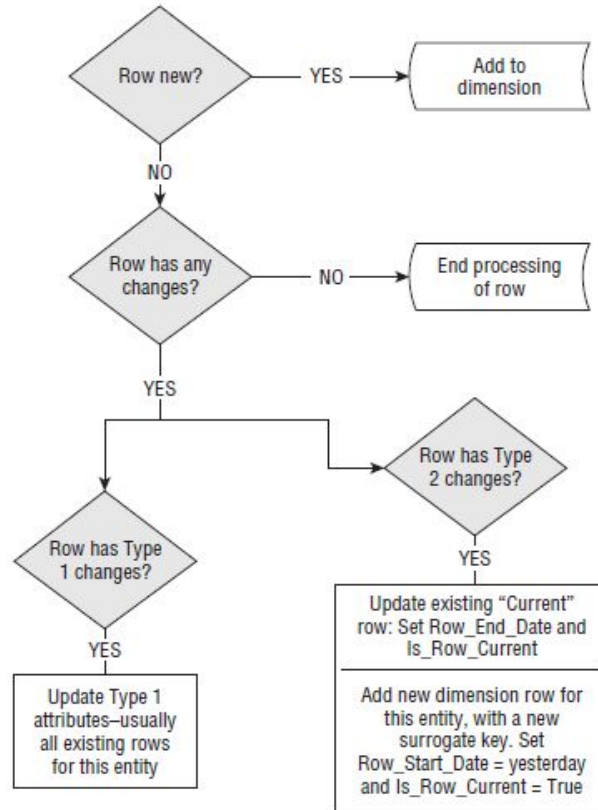
Original row in Product dimension:

| Product Key | SKU (NK) | Product Description | Department Name |
|-------------|----------|---------------------|-----------------|
| 12345 | ABC922-Z | IntelliKidz | Education |

Updated row in Product dimension:

| Product Key | SKU (NK) | Product Description | Department Name |
|-------------|----------|---------------------|-----------------|
| 12345 | ABC922-Z | IntelliKidz | Strategy |

SDC en el proceso de ETL



SCD - Tipo 2: Agregar un nuevo registro

Se agrega un nuevo registro a la dimensión con la misma llave natural pero con una nueva llave subrogada.

Se agregan fechas de control para determinar la validez de cada registro. Esto permite obtener datos precisos al momento de consultar la fact table.

Original row in Product dimension:

| Product Key | SKU (NK) | Product Description | Department Name | Introduction Date | ... | Row Effective Date | Row Expiration Date | Current Row Indicator |
|-------------|----------|---------------------|-----------------|-------------------|-----|--------------------|---------------------|-----------------------|
| 12345 | ABC922-Z | IntelliKidz | Education | 2012-12-15 | ... | 2012-01-01 | 9999-12-31 | Current |

Rows in Product dimension following type 2 change to Department Name and type 1 change to Introduction Date:

| Product Key | SKU (NK) | Product Description | Department Name | Introduction Date | ... | Row Effective Date | Row Expiration Date | Current Row Indicator |
|-------------|----------|---------------------|-----------------|-------------------|-----|--------------------|---------------------|-----------------------|
| 12345 | ABC922-Z | IntelliKidz | Education | 2012-01-01 | ... | 2012-01-01 | 2013-01-31 | Expired |
| 25984 | ABC922-Z | IntelliKidz | Strategy | 2012-01-01 | ... | 2013-02-01 | 9999-12-31 | Current |

SCD - Tipo 3: Agregar un nuevo atributo

Permite mostrar datos nuevos e históricos al utilizar el dato anterior, creando realidades alternativas.

Original row in Product dimension:

| Product Key | SKU (NK) | Product Description | Department Name |
|-------------|----------|---------------------|-----------------|
| 12345 | ABC922-Z | IntelliKidz | Education |

Updated row in Product dimension:

| Product Key | SKU (NK) | Product Description | Department Name | Prior Department Name |
|-------------|----------|---------------------|-----------------|-----------------------|
| 12345 | ABC922-Z | IntelliKidz | Strategy | Education |