

Almacenes de datos y OLAP

Diseño dimensional

Diego Villalba

Universidad de Costa Rica

2012-04-09

Resumen

- Tratamos resaltar una serie de mejores prácticas.

Resumen

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo,

Recapitulación

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.

Resumen

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.

Resumido

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.
- Entre otros aspectos, debemos considerar:

Resumen

- Resaltamos una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.
- Entre otros aspectos, debemos considerar:
 - El valor para el negocio.

Resapitulación

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.
- Entre otros aspectos, debemos considerar:
 - El valor para el negocio.
 - El esfuerzo requerido para construir reportes.

Resumido

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.
- Entre otros aspectos, debemos considerar:
 - El valor para el negocio.
 - El esfuerzo requerido para construir reportes.
 - La complejidad del proceso de carga.

Resumido

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.
- Entre otros aspectos, debemos considerar:
 - El valor para el negocio.
 - El esfuerzo requerido para construir reportes.
 - La complejidad del proceso de carga.
 - El costo.

Resapitulación

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.
- Entre otros aspectos, debemos considerar:
 - El valor para el negocio.
 - El esfuerzo requerido para construir reportes.
 - La complejidad del proceso de carga.
 - El costo.
- Por ejemplo:

Resapitulación

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.
- Entre otros aspectos, debemos considerar:
 - El valor para el negocio.
 - El esfuerzo requerido para construir reportes.
 - La complejidad del proceso de carga.
 - El costo.
- Por ejemplo:
 - El aplanamiento de una jerarquía recursiva simplifica la construcción de reportes y reduce el costo de desarrollo, pero...

Resumido

- Tratamos resaltar una serie de mejores prácticas.
- Sin embargo, no existe una *única forma correcta* de modelar un proceso de negocio particular.
- Entre otros aspectos, debemos considerar:
 - El valor para el negocio.
 - El esfuerzo requerido para construir reportes.
 - La complejidad del proceso de carga.
 - El costo.
- Por ejemplo:
 - El aplanamiento de una jerarquía recursiva simplifica la construcción de reportes y reduce el costo de desarrollo, pero...
 - ...limita el poder de la solución final.

Recapitulación

- Diseñamos bases de datos analíticas.

Resumen

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.

Resumido

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.

Recapitulación

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:

Recapitulación

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:
 - Las mediciones y...

Resumen

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:
 - Las mediciones y...
 - su contexto

Resumen

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:
 - Las mediciones y...
 - su contexto, conocidos como:

Resumido

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:
 - Las mediciones y...
 - su contexto, conocidos como:
 - *hechos y dimensiones*.

Resumen

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:
 - Las mediciones y...
 - su contexto, conocidos como:
 - *hechos y dimensiones*.
- Implementado en una base de datos relacional, al modelo dimensional se le llama:

Resumido

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:
 - Las mediciones y...
 - su contexto, conocidos como:
 - *hechos y dimensiones*.
- Implementado en una base de datos relacional, al modelo dimensional se le llama:
 - *esquema de estrella*.

Resumen

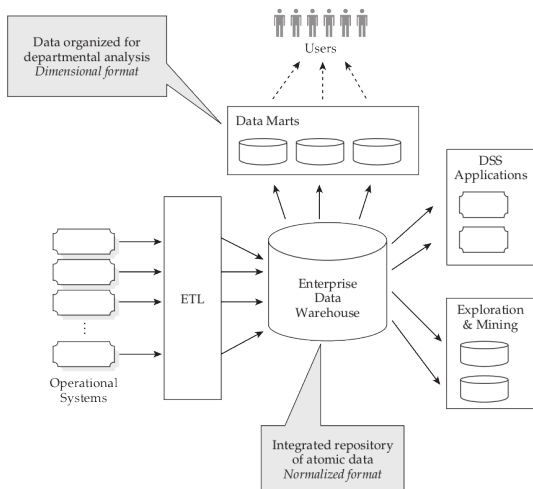
- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:
 - Las mediciones y...
 - su contexto, conocidos como:
 - *hechos y dimensiones*.
- Implementado en una base de datos relacional, al modelo dimensional se le llama:
 - *esquema de estrella*.
- Implementado en una base de datos multidimensional, se lo conoce como:

Resumido

- Diseñamos bases de datos analíticas.
- Modelamos la *medición* de los procesos de negocios.
- El modelo dimensional de un proceso de negocio está hecho de dos componentes:
 - Las mediciones y...
 - su contexto, conocidos como:
 - *hechos y dimensiones*.
- Implementado en una base de datos relacional, al modelo dimensional se le llama:
 - *esquema de estrella*.
- Implementado en una base de datos multidimensional, se lo conoce como:
 - *cubo*.

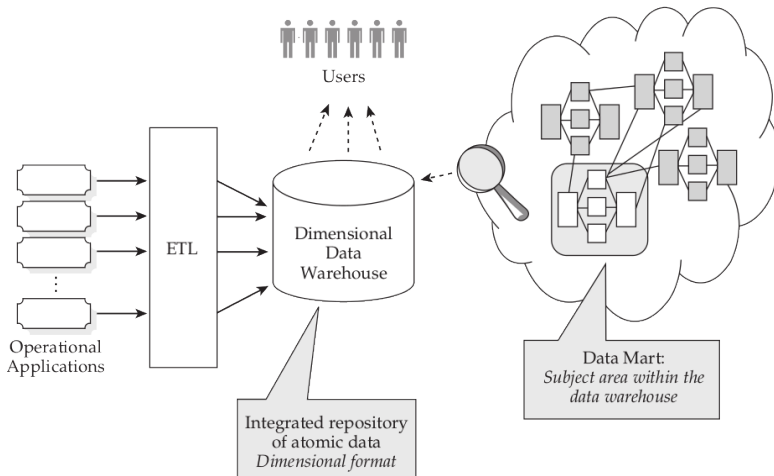
Recapitulación

Arquitecturas - Corporate Information Factory de Inmon



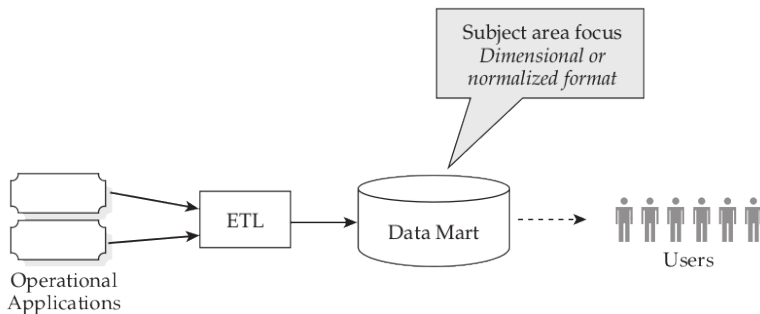
Recapitulación

Arquitecturas - Dimensional Data Warehouse de Kimball



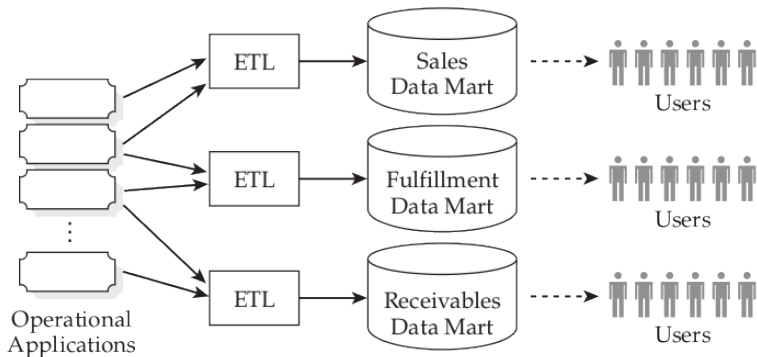
Recapitulación

Arquitecturas - Stand-alone Data Mart



Recapitulación

Arquitecturas - Stand-alone Data Marts Múltiples



Recapitulación

Arquitecturas - Comparación

	Enterprise Level			Subject Area Level		
	Integrated Repository of Atomic Data	Format	Direct Access	Data Marts	Format	Direct Access
Corporate Information Factory	✓	3NF	No	Physical	Dimensional*	Yes
Dimensional Data Warehouse	✓	Dimensional	Yes*	Logical*	Dimensional	Yes
Stand-Alone Data Marts	✗	<i>n/a</i>	<i>n/a</i>	Physical	Dimensional*	Yes

* Optional

Estrellas y cubos

- 1 Características de las tablas de dimensiones
- 2 Características de las tablas de hechos
- 3 Dimensiones lentamente cambiantes
- 4 Cubos

Características de las tablas de dimensiones

- Un conjunto de tablas de dimensiones bien desarrollado provee capacidades analíticas variadas y poderosas.

Características de las tablas de dimensiones

- Un conjunto de tablas de dimensiones bien desarrollado provee capacidades analíticas variadas y poderosas.
- Las dimensiones proveen información contextual.

Características de las tablas de dimensiones

- Un conjunto de tablas de dimensiones bien desarrollado provee capacidades analíticas variadas y poderosas.
- Las dimensiones proveen información contextual.
- Sin ellas, los reportes no tendrían sentido.

Características de las tablas de dimensiones

- Un conjunto de tablas de dimensiones bien desarrollado provee capacidades analíticas variadas y poderosas.
- Las dimensiones proveen información contextual.
- Sin ellas, los reportes no tendrían sentido.
- Un diseño dimensional exitoso depende de:

Características de las tablas de dimensiones

- Un conjunto de tablas de dimensiones bien desarrollado provee capacidades analíticas variadas y poderosas.
- Las dimensiones proveen información contextual.
- Sin ellas, los reportes no tendrían sentido.
- Un diseño dimensional exitoso depende de:
 - 1 El uso apropiado de las llaves.

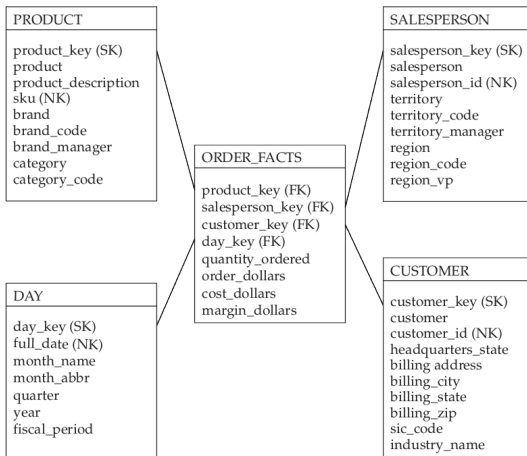
Características de las tablas de dimensiones

- Un conjunto de tablas de dimensiones bien desarrollado provee capacidades analíticas variadas y poderosas.
- Las dimensiones proveen información contextual.
- Sin ellas, los reportes no tendrían sentido.
- Un diseño dimensional exitoso depende de:
 - 1 El uso apropiado de las llaves.
 - 2 El desarrollo de un conjunto ricamente detallado de columnas de dimensiones.

Características de las tablas de dimensiones

- Un conjunto de tablas de dimensiones bien desarrollado provee capacidades analíticas variadas y poderosas.
- Las dimensiones proveen información contextual.
- Sin ellas, los reportes no tendrían sentido.
- Un diseño dimensional exitoso depende de:
 - 1 El uso apropiado de las llaves.
 - 2 El desarrollo de un conjunto ricamente detallado de columnas de dimensiones.
 - 3 El rechazo a la necesidad de ahorrar espacio.

Claves sustitutas y claves naturales



Llaves sustitutas y llaves naturales

- En un esquema de estrella, a cada tabla de dimensión se le asigna una llave sustituta (Surrogate Key - SK)

Llaves sustitutas y llaves naturales

- En un esquema de estrella, a cada tabla de dimensión se le asigna una llave sustituta (Surrogate Key - SK)
- Aparte de las SKs, una o más llaves naturales (NK) aparecen en la mayoría de las tablas de dimensiones.

Llaves sustitutas y llaves naturales

- En un esquema de estrella, a cada tabla de dimensión se le asigna una llave sustituta (Surrogate Key - SK)
- Aparte de las SKs, una o más llaves naturales (NK) aparecen en la mayoría de las tablas de dimensiones.
- Las NKs podrían tener significado para el usuario.

Llaves sustitutas y llaves naturales

- En un esquema de estrella, a cada tabla de dimensión se le asigna una llave sustituta (Surrogate Key - SK)
- Aparte de las SKs, una o más llaves naturales (NK) aparecen en la mayoría de las tablas de dimensiones.
- Las NKs podrían tener significado para el usuario.
- ¿Podrían las NKs necesitar más de una columna?

Llaves sustitutas y llaves naturales

- En un esquema de estrella, a cada tabla de dimensión se le asigna una llave sustituta (Surrogate Key - SK)
- Aparte de las SKs, una o más llaves naturales (NK) aparecen en la mayoría de las tablas de dimensiones.
- Las NKs podrían tener significado para el usuario.
- ¿Podrían las NKs necesitar más de una columna?
- ¿Qué ocurre cuando las NKs provienen de fuentes diferentes?

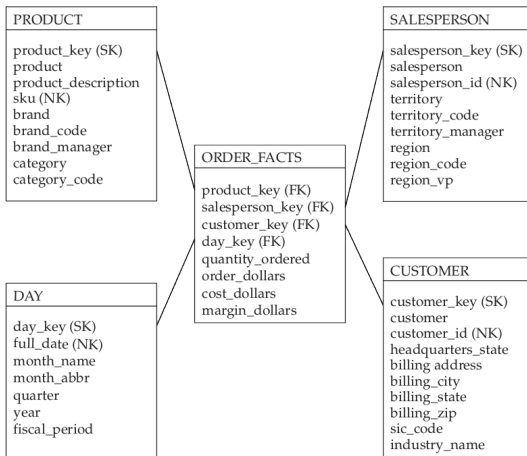
Llaves sustitutas y llaves naturales

- En un esquema de estrella, a cada tabla de dimensión se le asigna una llave sustituta (Surrogate Key - SK)
- Aparte de las SKs, una o más llaves naturales (NK) aparecen en la mayoría de las tablas de dimensiones.
- Las NKs podrían tener significado para el usuario.
- ¿Podrían las NKs necesitar más de una columna?
- ¿Qué ocurre cuando las NKs provienen de fuentes diferentes?
- ¿Qué ocurre si en la tabla de dimensión Cliente utilizamos como PK el ID_Cliente de la tabla fuente y el cliente cambia de dirección?

Llaves sustitutas y llaves naturales

- En un esquema de estrella, a cada tabla de dimensión se le asigna una llave sustituta (Surrogate Key - SK)
- Aparte de las SKs, una o más llaves naturales (NK) aparecen en la mayoría de las tablas de dimensiones.
- Las NKs podrían tener significado para el usuario.
- ¿Podrían las NKs necesitar más de una columna?
- ¿Qué ocurre cuando las NKs provienen de fuentes diferentes?
- ¿Qué ocurre si en la tabla de dimensión Cliente utilizamos como PK el ID_Cliente de la tabla fuente y el cliente cambia de dirección?
- ¿Qué podríamos utilizar como alternativa a una SK?
¿Números de versión? ¿Marcas de tiempo?

Claves sustitutas y claves naturales



Rico conjunto de dimensiones

- Las dimensiones proveen contexto a los hechos.

Rico conjunto de dimensiones

- Las dimensiones proveen contexto a los hechos.
- Sin contexto, los hechos no pueden ser interpretados.

Rico conjunto de dimensiones

- Las dimensiones proveen contexto a los hechos.
- Sin contexto, los hechos no pueden ser interpretados.
- Por ejemplo:
 - “Total facturado: \$10,000”

Rico conjunto de dimensiones

- Las dimensiones proveen contexto a los hechos.
- Sin contexto, los hechos no pueden ser interpretados.
- Por ejemplo:
 - “Total facturado: **\$10,000**”
 - “Total facturado en diciembre de 2011 por concepto de artículos en la categoría Monitores: **\$10,000**”

Rico conjunto de dimensiones

Las dimensiones y sus valores agregan significado de varias formas:

- Se utilizan para filtrar consultas o reportes.

Rico conjunto de dimensiones

Las dimensiones y sus valores agregan significado de varias formas:

- Se utilizan para filtrar consultas o reportes.
- Se utilizan para controlar el ámbito de agregación para los hechos.

Rico conjunto de dimensiones

Las dimensiones y sus valores agregan significado de varias formas:

- Se utilizan para filtrar consultas o reportes.
- Se utilizan para controlar el ámbito de agregación para los hechos.
- Se utilizan para ordenar la información.

Rico conjunto de dimensiones

Las dimensiones y sus valores agregan significado de varias formas:

- Se utilizan para filtrar consultas o reportes.
- Se utilizan para controlar el ámbito de agregación para los hechos.
- Se utilizan para ordenar la información.
- Acompañan a los hechos para proveer contexto en los reportes.

Rico conjunto de dimensiones

Las dimensiones y sus valores agregan significado de varias formas:

- Se utilizan para filtrar consultas o reportes.
- Se utilizan para controlar el ámbito de agregación para los hechos.
- Se utilizan para ordenar la información.
- Acompañan a los hechos para proveer contexto en los reportes.
- Se utilizan para definir organización maestro-detalle, agrupamientos, subtotales y resúmenes.

Rico conjunto de dimensiones

- Cuanto mayor es el conjunto de atributos de dimensiones, hay más formas en que los hechos pueden ser analizados.

Rico conjunto de dimensiones

- Cuanto mayor es el conjunto de atributos de dimensiones, hay más formas en que los hechos pueden ser analizados.
- Cada atributo que agregamos aumenta dramáticamente el número de posibilidades analíticas.

Rico conjunto de dimensiones

- Cuanto mayor es el conjunto de atributos de dimensiones, hay más formas en que los hechos pueden ser analizados.
- Cada atributo que agregamos aumenta dramáticamente el número de posibilidades analíticas.
- Tablas de dimensiones con un gran número de atributos **maximizan el valor analítico.**

Rico conjunto de dimensiones

- Cuanto mayor es el conjunto de atributos de dimensiones, hay más formas en que los hechos pueden ser analizados.
- Cada atributo que agregamos aumenta dramáticamente el número de posibilidades analíticas.
- Tablas de dimensiones con un gran número de atributos **maximizan el valor analítico**.
- Además de atributos comunes, las tablas de dimensiones guardan combinaciones de atributos usados comúnmente.

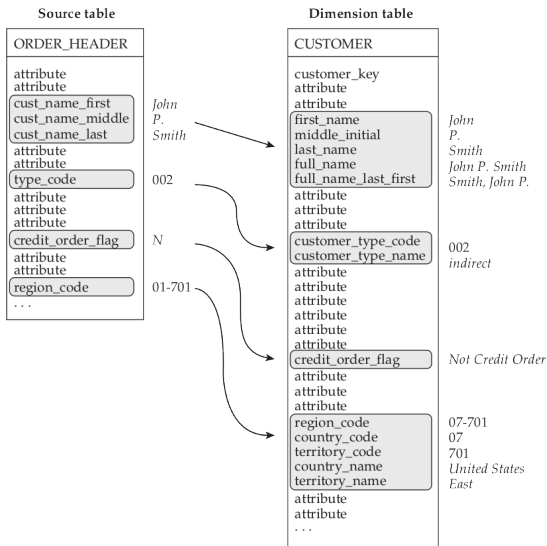
Rico conjunto de dimensiones

- Cuanto mayor es el conjunto de atributos de dimensiones, hay más formas en que los hechos pueden ser analizados.
- Cada atributo que agregamos aumenta dramáticamente el número de posibilidades analíticas.
- Tablas de dimensiones con un gran número de atributos **maximizan el valor analítico**.
- Además de atributos comunes, las tablas de dimensiones guardan combinaciones de atributos usados comúnmente.
- Códigos pueden ser reemplazados por sus descripciones, al igual que banderas booleanas.

Rico conjunto de dimensiones

- Cuanto mayor es el conjunto de atributos de dimensiones, hay más formas en que los hechos pueden ser analizados.
- Cada atributo que agregamos aumenta dramáticamente el número de posibilidades analíticas.
- Tablas de dimensiones con un gran número de atributos **maximizan el valor analítico**.
- Además de atributos comunes, las tablas de dimensiones guardan combinaciones de atributos usados comúnmente.
- Códigos pueden ser reemplazados por sus descripciones, al igual que banderas booleanas.
- Campos multiparte se conservan y además, se descomponen.

Rico conjunto de dimensiones



Dimensiones con valores numéricos

- La mayoría de las dimensiones contienen valores textuales.

Dimensiones con valores numéricos

- La mayoría de las dimensiones contienen valores textuales.
- Los hechos tienden a ser numéricos.

Dimensiones con valores numéricos

- La mayoría de las dimensiones contienen valores textuales.
- Los hechos tienden a ser numéricos.
- El precio de un artículo vendido, ¿es un hecho o una dimensión?

Dimensiones con valores numéricos

- La mayoría de las dimensiones contienen valores textuales.
- Los hechos tienden a ser numéricos.
- El precio de un artículo vendido, ¿es un hecho o una dimensión?
- No siempre está claro si un elemento numérico es un hecho o una dimensión. En caso de duda, ponga atención a cómo será utilizado.

Dimensiones con valores numéricos

- La mayoría de las dimensiones contienen valores textuales.
- Los hechos tienden a ser numéricos.
- El precio de un artículo vendido, ¿es un hecho o una dimensión?
- No siempre está claro si un elemento numérico es un hecho o una dimensión. En caso de duda, ponga atención a cómo será utilizado.
- Si los valores del elemento son usados para filtrar consultas, ordenar datos, controlar agregación o crear relaciones maestro-detalle,

Dimensiones con valores numéricos

- La mayoría de las dimensiones contienen valores textuales.
- Los hechos tienden a ser numéricos.
- El precio de un artículo vendido, ¿es un hecho o una dimensión?
- No siempre está claro si un elemento numérico es un hecho o una dimensión. En caso de duda, ponga atención a cómo será utilizado.
- Si los valores del elemento son usados para filtrar consultas, ordenar datos, controlar agregación o crear relaciones maestro-detalle, **lo más probable es que es una dimensión.**

Dimensiones de comportamiento

- Una técnica analítica poderosa utiliza patrones de comportamiento para analizar hechos.

Dimensiones de comportamiento

- Una técnica analítica poderosa utiliza patrones de comportamiento para analizar hechos.
- Considere esta pregunta:

Dimensiones de comportamiento

- Una técnica analítica poderosa utiliza patrones de comportamiento para analizar hechos.
- Considere esta pregunta: ¿Reciben los clientes que generan más de un millón de dólares en ventas mejores descuentos que los que generan cien mil o menos?

Dimensiones de comportamiento

- Una técnica analítica poderosa utiliza patrones de comportamiento para analizar hechos.
- Considere esta pregunta: **¿Reciben los clientes que generan más de un millón de dólares en ventas mejores descuentos que los que generan cien mil o menos?**
- Las órdenes de compras o facturas son rastreadas por un hecho. Para responder a esta pregunta, debemos agrupar los clientes basados en su actividad de compra pasada.

Dimensiones de comportamiento

- Una técnica analítica poderosa utiliza patrones de comportamiento para analizar hechos.
- Considere esta pregunta: **¿Reciben los clientes que generan más de un millón de dólares en ventas mejores descuentos que los que generan cien mil o menos?**
- Las órdenes de compras o facturas son rastreadas por un hecho. Para responder a esta pregunta, debemos agrupar los clientes basados en su actividad de compra pasada.
- Estos agrupamientos son entonces usados como una dimensión y luego utilizados para estudiar descuentos.

Dimensiones de comportamiento

- Una técnica analítica poderosa utiliza patrones de comportamiento para analizar hechos.
- Considere esta pregunta: **¿Reciben los clientes que generan más de un millón de dólares en ventas mejores descuentos que los que generan cien mil o menos?**
- Las órdenes de compras o facturas son rastreadas por un hecho. Para responder a esta pregunta, debemos agrupar los clientes basados en su actividad de compra pasada.
- Estos agrupamientos son entonces usados como una dimensión y luego utilizados para estudiar descuentos.
- Una dimensión que se calcula basándose en hechos se conoce como **dimensión de comportamiento**.

Agrupando dimensiones

- Los atributos de dimensiones se agrupan en tablas que representan categorías grandes de información de referencia.

Agrupando dimensiones

- Los atributos de dimensiones se agrupan en tablas que representan categorías grandes de información de referencia.
- Cuando principios de normalización son aplicados a una tabla de dimensión, el resultado se conoce como *“snowflake”*.

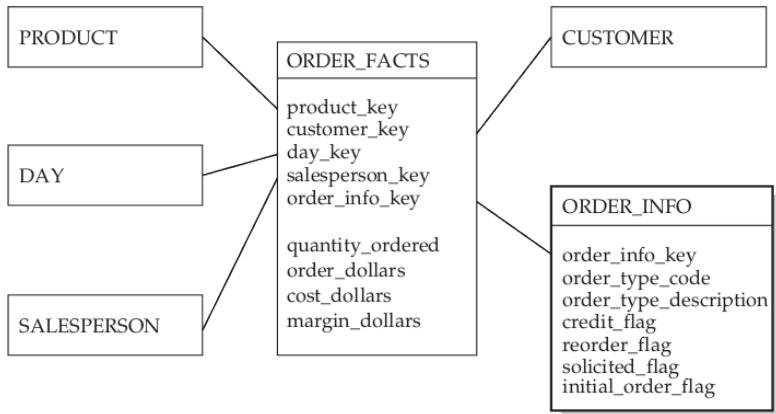
Agrupando dimensiones

- Los atributos de dimensiones se agrupan en tablas que representan categorías grandes de información de referencia.
- Cuando principios de normalización son aplicados a una tabla de dimensión, el resultado se conoce como *“snowflake”*.
- Atributos misceláneos que no comparten una afinidad natural, se agrupan en dimensiones de tipo *“junk”*.

Agrupando dimensiones

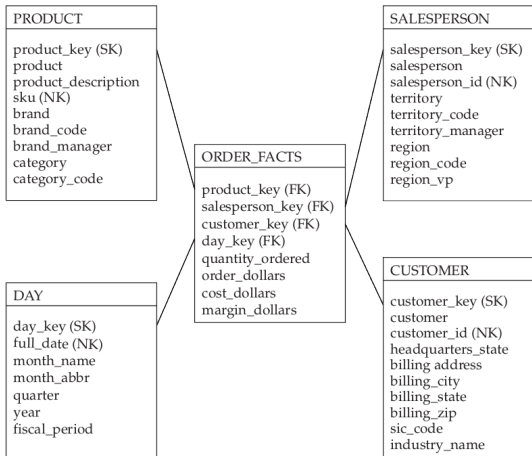
- Los atributos de dimensiones se agrupan en tablas que representan categorías grandes de información de referencia.
- Cuando principios de normalización son aplicados a una tabla de dimensión, el resultado se conoce como *“snowflake”*.
- Atributos misceláneos que no comparten una afinidad natural, se agrupan en dimensiones de tipo *“junk”*.
- A diferencia del diseño en 3NF, el diseño dimensional acepta y alienta el almacenaje redundante de datos.

Tabla de dimensiones tipo *junk*



Snowflakes

¿Por qué no poner las marcas de productos en otra tabla?



Beneficios de la redundancia

Tres principales beneficios:

Beneficios de la redundancia

Tres principales beneficios:

- 1 Rendimiento.

Beneficios de la redundancia

Tres principales beneficios:

- 1 Rendimiento.
- 2 Usabilidad.

Beneficios de la redundancia

Tres principales beneficios:

- 1 Rendimiento.
- 2 Usabilidad.
- 3 Consistencia.

Beneficios de la redundancia

Tres principales beneficios:

- 1 Rendimiento.
- 2 Usabilidad.
- 3 Consistencia.

El almacenaje explícito de cada atributo de dimensión maximiza el rendimiento, ayuda en la usabilidad y garantiza consistencia a través de diferentes aplicaciones. Las reglas de normalización se adaptan mejor a las necesidades de sistemas operacionales que a las de los sistemas analíticos.

Características de las tablas de hechos

Mientras que las tablas de dimensiones son *anchas*, las tablas de hechos son *profundas*.

Características de las tablas de hechos

Mientras que las tablas de dimensiones son *anchas*, las tablas de hechos son *profundas*.

Usualmente, acumulan filas mucho más rápidamente que las tablas de dimensiones asociadas.

Características de las tablas de hechos

Mientras que las tablas de dimensiones son *anchas*, las tablas de hechos son *profundas*.

Usualmente, acumulan filas mucho más rápidamente que las tablas de dimensiones asociadas.

En algunos casos, las tablas de hechos pueden también contener dimensiones.

Características de las tablas de hechos

Mientras que las tablas de dimensiones son *anchas*, las tablas de hechos son *profundas*.

Usualmente, acumulan filas mucho más rápidamente que las tablas de dimensiones asociadas.

En algunos casos, las tablas de hechos pueden también contener dimensiones.

product_key	day_key	salesperson_key	customer_key	quantity_ ordered	order_ dollars	cost_ dollars
102291	3831	2991	240123	882	8822.29	8028.28
194482	2931	1992	572339	249	2499.29	2274.35
183882	2983	2933	937759	3394	33940.29	30885.66
102291	3831	2991	240123	882	8822.29	8028.28

Capturando hechos

- En un diseño dimensional, cada tabla de hechos describe un proceso de negocio.

Capturando hechos

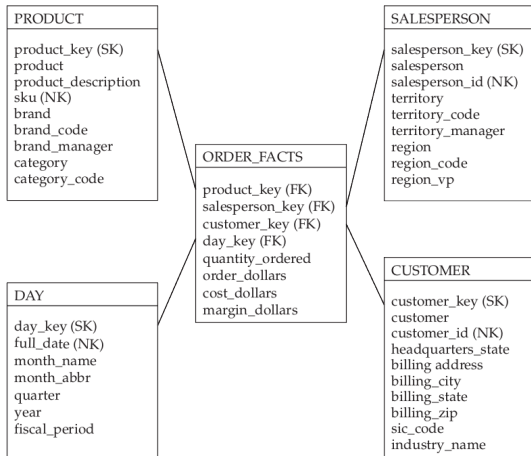
- En un diseño dimensional, cada tabla de hechos describe un proceso de negocio.
- Debe proveer un conjunto de mediciones comprensibles...

Capturando hechos

- En un diseño dimensional, cada tabla de hechos describe un proceso de negocio.
- Debe proveer un conjunto de mediciones comprensibles...
- ...incluso aunque algunas de ellas sean redundantes.

Capturando hechos

¿Redundancia otra vez? ¿Calcular utilidad durante ETL?



Hechos no aditivos

Day	Salesperson	Product	Customer	Margin Rate
1/1/2009	Jones	Gel Pen Red	Balter Inc.	3.02%
1/1/2009	Jones	Gel Pen Red	Raytech	3.02%
1/1/2009	Baldwin	Gel Pen Red	Venerable Holdings	3.02%
1/1/2009	Baldwin	Gel Pen Red	eMart LLC	3.02%
1/1/2009	Baldwin	Gel Pen Red	Shatter & Lose	3.02%
1/1/2009	Sebenik	Gel Pen Red	Comstock Realty	3.02%
1/1/2009	Sebenik	Gel Pen Red	RizSpace	3.02%
1/1/2009	Sebenik	Gel Pen Red	StarComp	3.02%
1/1/2009	Sgamma	Gel Pen Red	Implosion Town	3.02%
			Sum:	27.18%

Hechos no aditivos

Margin Report				
Date: January 1, 2009 Product: Gel Pen Red				
Salesperson	Customer	Margin Dollars	Order Dollars	Margin Rate
Jones	Balter Inc.	192.74	6,382.21	3.02%
	Raytech	39.05	1,293.11	3.02%
	<i>Subtotal:</i>	231.79	7,675.32	3.02%
Baldwin	Venerable Holdings	121.50	4,023.22	3.02%
	eMart LLC	253.44	8,392.00	3.02%
	Shatter & Lose	8.74	289.54	3.02%
	<i>Subtotal:</i>	383.68	12,704.76	3.02%
Sebenik	Comstock Realty	12.06	399.29	3.02%
	RizSpace	58.10	1,923.93	3.02%
	Starcomp	90.36	2,992.11	3.02%
	<i>Subtotal:</i>	160.52	5,315.33	3.02%
Sgamma	Implosion Town	213.88	7,082.22	3.02%
	DemiSpace	113.92	3,772.11	3.02%
	<i>Subtotal:</i>	327.80	10,854.33	3.02%
Grand Total:		1,103.80	36,549.74	3.02%

Margin Rate is a nonadditive fact.

Summary row is computed as a ratio of the subtotals, not by summing margin rates for the salesperson.

Grano

- Nivel de detalle representado por una fila de la tabla de hechos.

Grano

- Nivel de detalle representado por una fila de la tabla de hechos.
- Declarar el grano es parte importante en el proceso de diseño:

Grano

- Nivel de detalle representado por una fila de la tabla de hechos.
- Declarar el grano es parte importante en el proceso de diseño:
 - Asegura que no habrá confusión sobre el significado de cada fila de la tabla de hechos.

Grano

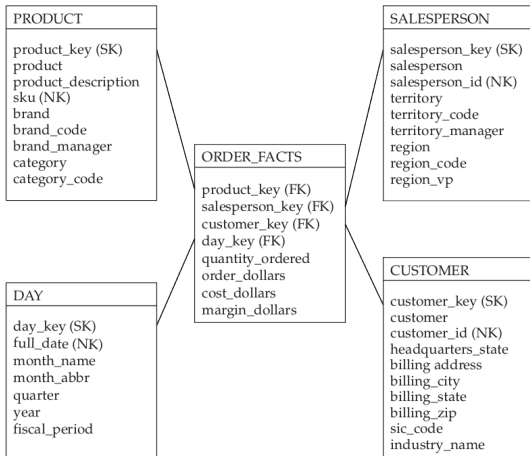
- Nivel de detalle representado por una fila de la tabla de hechos.
- Declarar el grano es parte importante en el proceso de diseño:
 - Asegura que no habrá confusión sobre el significado de cada fila de la tabla de hechos.
 - Garantiza que todos los hechos serán registrados al mismo nivel de detalle.

Grano

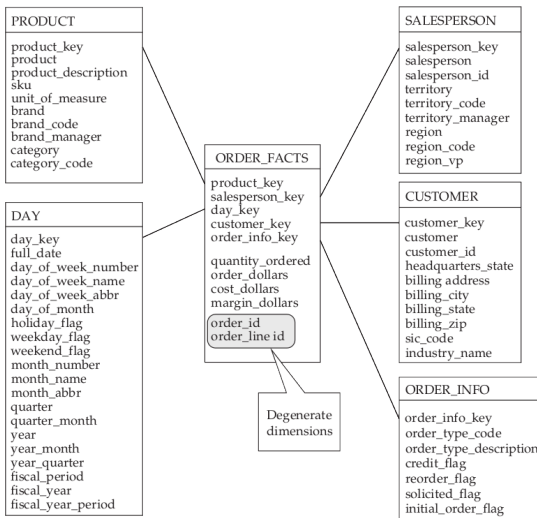
- Nivel de detalle representado por una fila de la tabla de hechos.
- Declarar el grano es parte importante en el proceso de diseño:
 - Asegura que no habrá confusión sobre el significado de cada fila de la tabla de hechos.
 - Garantiza que todos los hechos serán registrados al mismo nivel de detalle.
- En el ejemplo anterior, puede ser descrito como:
"Facturas por Día, Vendedor, Producto y Cliente".

Grano

“Facturas por Día, Vendedor, Producto y Cliente”



Dimensiones degeneradas



Dimensiones lentamente cambiantes

- Los datos de las tablas de dimensiones se originan en los sistemas operacionales.

Dimensiones lentamente cambiantes

- Los datos de las tablas de dimensiones se originan en los sistemas operacionales.
- Luego de ser insertados en las tablas de dimensiones, podrían cambiar en la fuente.

Dimensiones lentamente cambiantes

- Los datos de las tablas de dimensiones se originan en los sistemas operacionales.
- Luego de ser insertados en las tablas de dimensiones, podrían cambiar en la fuente.
- En un diseño dimensional, es crucial identificar cómo representar los cambios en los datos fuente en las tablas de dimensiones.

Dimensiones lentamente cambiantes

January 1, 2007

```
customer_id: 9900011
cust_name: Johnson, Sue
cust_birth_date: 3/2/1961
address_state: AZ
(more attributes) ...
```

January 31, 2007

```
customer_id: 9900011
cust_name: Johnson, Sue
cust_birth_date: 3/2/1971
address_state: AZ
(more attributes) ...
```

Sue's date of birth has changed.

May 5, 2009

```
customer_id: 9900011
cust_name: Johnson, Sue
cust_birth_date: 3/2/1971
address_state: CA
(more attributes) ...
```

Sue's state has changed.

Dimensiones lentamente cambiantes

- Tres tipos diferentes de respuestas:

Dimensiones lentamente cambiantes

- Tres tipos diferentes de respuestas:
 - 1 Tipo 1: sobrescribir historia transaccional.

Dimensiones lentamente cambiantes

- Tres tipos diferentes de respuestas:
 - 1 Tipo 1: sobrescribir historia transaccional.
 - 2 Tipo 2: conservar historia transaccional.

Dimensiones lentamente cambiantes

- Tres tipos diferentes de respuestas:
 - 1 Tipo 1: sobrescribir historia transaccional.
 - 2 Tipo 2: conservar historia transaccional.
 - 3 Tipo 3: no se requiere historia transaccional pero se conservan valores *antes y después*.

Cambios de tipo 1

- Sobreescribir el valor.

Cambios de tipo 1

- Sobreescribir el valor.
- Hechos preexistentes tienen nuevo contexto.

Cambios de tipo 1

- Sobreescribir el valor.
- Hechos preexistentes tienen nuevo contexto.
- Historia de la dimensión no se conserva.

Cambios de tipo 1

Customer dimension table

customer_key	customer_id	customer_name	date_of_birth	state	
1011	1140400	Davis, Robert	7/8/1932	NY	
1022	3305300	Nguyen, Tamara	1/2/1965	DC	
1302	7733300	Rodriguez, Jason	4/5/1972	MA	
1499	9900011	Johnson, Sue	3/2/1961	AZ	

More attributes...

Orders fact table

customer_key	day_key	product_key	quantity_ordered	
1499	2322	10119	5	

Record for Sue in Dimension table

BEFORE

Cambios de tipo 1

AFTER

Customer dimension table

customer_key	customer_id	customer_name	date_of_birth	state	
1011	1140400	Davis, Robert	7/8/1932	NY	
1022	3305300	Nguyen, Tamara	1/2/1965	DC	
1302	7733300	Rodriguez, Jason	4/5/1972	MA	
1499	9900011	Johnson, Sue	3/2/1971	AZ	

More attributes...

Record overwritten with new date of birth

Orders fact table

customer_key	day_key	product_key	quantity_ordered	
1499	2322	10119	5	

Old fact references updated dimension row

Cambios de tipo 2

- Metodología más utilizada: insertar una nueva fila.

Cambios de tipo 2

- Metodología más utilizada: insertar una nueva fila.
- Contexto histórico se conserva.

Cambios de tipo 2

- Metodología más utilizada: insertar una nueva fila.
- Contexto histórico se conserva.
- Historia de la dimensión se conserva parcialmente.

Cambios de tipo 2

Customer dimension table

customer_key	customer_id	customer_name	date_of_birth	state	
1011	1140400	Davis, Robert	7/8/1932	NY	
1022	3305300	Nguyen, Tamara	1/2/1965	DC	
1302	7733300	Rodriguez, Jason	4/5/1972	MA	
1499	9900011	Johnson, Sue	3/2/1971	AZ	

→ More attributes...

Orders fact table

customer_key	day_key	product_key	quantity_ordered	
1499	2322	10119	5	

BEFORE

Cambios de tipo 2

AFTER

Customer dimension table

customer_key	customer_id	customer_name	date_of_birth	state	
1011	1140400	Davis, Robert	7/8/1932	NY	→ More attributes...
1022	3305300	Nguyen, Tamara	1/2/1965	DC	
1302	7733300	Rodriguez, Jason	4/5/1972	MA	
1499	9900011	Johnson, Sue	3/2/1971	AZ	
2507	9900011	Johnson, Sue	3/2/1971	CA	

Row added for new version of Sue

Orders fact table

customer_key	day_key	product_key	quantity_ordered	
1499	2322	10119	5	Old fact references historic version of Sue
2507	4722	20112	1	New fact references revised version of Sue

Implementación de los tipos de respuesta

- Parte importante del diseño: identificar reglas para el procesamiento de cambio lento en dimensiones.

Implementación de los tipos de respuesta

- Parte importante del diseño: identificar reglas para el procesamiento de cambio lento en dimensiones.
- Dada una fuente de cambio, la respuesta correcta depende de los requerimientos analíticos.

Implementación de los tipos de respuesta

- Parte importante del diseño: identificar reglas para el procesamiento de cambio lento en dimensiones.
- Dada una fuente de cambio, la respuesta correcta depende de los requerimientos analíticos.
- Responsabilidad de los desarrolladores de ETL.

Implementación de los tipos de respuesta

Documentando las reglas de cambio lento para una tabla de dimensiones.

CUSTOMER	
customer_key	SK
customer_id	NK
first_name	2
middle_initial	2
last_name	2
full_name	2
full_name_last_first	2
gender	1
data_of_birth	1
address	2
city	2
state	2
zip_code	2
marital_status	2
daytime_phone	2
evening_phone	2
...	

Legend:

SK Surrogate Key

NK Natural Key

1 Type 1

2 Type 2

Implementación de los tipos de respuesta

Resumen de técnicas para diseño e implementación de dimensiones lentamente cambiantes

	Action	Effect on Facts
Type 1	Update Dimension	Restates History
Type 2	Insert New Row in Dimension Table	Preserves History

Cubos

- Los modelos dimensionales no siempre se implementan en bases de datos relacionales.

Cubos

- Los modelos dimensionales no siempre se implementan en bases de datos relacionales.
- Las MDB guardan la información dimensional en un formato llamado *cubo*.

Cubos

- Los modelos dimensionales no siempre se implementan en bases de datos relacionales.
- Las MDB guardan la información dimensional en un formato llamado *cubo*.
- El concepto básico detrás del cubo es la computación *a priori* de varias combinaciones de valores de hechos y dimensiones para que puedan ser estudiadas interactivamente.

Almacenamiento multidimensional vs. relacional

- La principal ventaja de una MDB es su velocidad.

Almacenamiento multidimensional vs. relacional



- La principal ventaja de una MDB es su velocidad.
- El cubo permite a los usuarios cambiar su perspectiva de los datos interactivamente, agregando o quitando atributos en su vista y recibiendo respuesta instantánea.

Almacenamiento multidimensional vs. relacional

- La principal ventaja de una MDB es su velocidad.
- El cubo permite a los usuarios cambiar su perspectiva de los datos interactivamente, agregando o quitando atributos en su vista y recibiendo respuesta instantánea.
- Este proceso es conocido como OLAP.

Almacenamiento multidimensional vs. relacional

Tecnologías alternativas para el almacenamiento de datos dimensionales

	Data Structure	Access Language	Style of Interaction	Advantages
Relational Database	 Star Schema	Structured Query Language (SQL)	Query and response	Scalable Widely understood access language
Multi-dimensional Database	 Cube	Proprietary API or MDX	Interactive (OLAP)	Fast Expressive access language

These distinctions are beginning to fade

Referencias



Adamson, C.

Star Schema - The Complete Reference.

McGraw-Hill, 2010.