TEMA 5

Casos de diseño

José Samos Jiménez

2020 jsamos (lsi-ugr)
Departamento de Lenguajes y Sistemas Informáticos
Universidad de Granada

Curso 2019-20

Tipos de hechos y de jerarquías

Tipos de hechos [KR13]

- Transacciones: datos dispersos.
- Instantánea: estados, datos no dispersos.
- Instantánea acumulada: estados que se actualizan.

Tipos de jerarquías [MZ08]

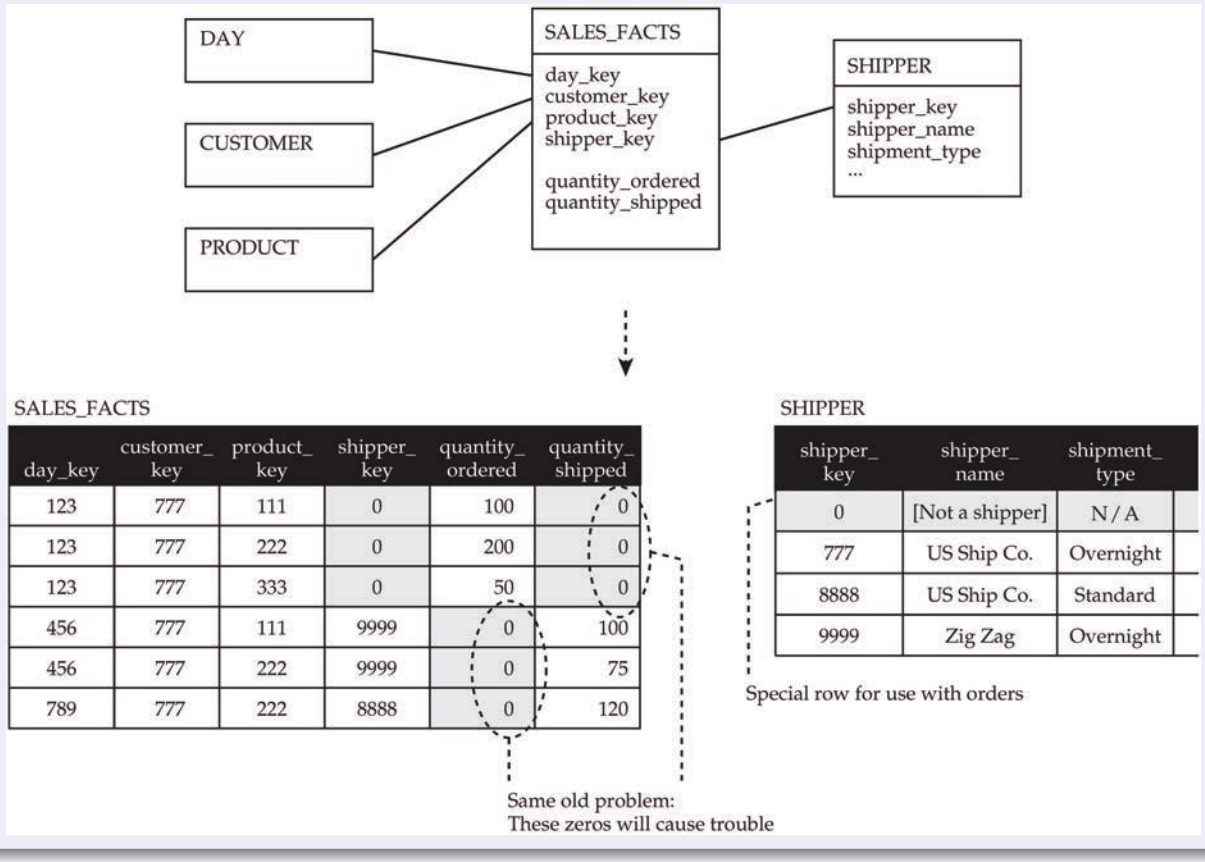
- Balanceada / no-balanceada (p.e., subordinado-jefe).
- Estricta / no-estricta (p.e., autor-libro).
- Regulares / irregulares (p.e., provincia-CA con Ceuta o Melilla).

Contenido

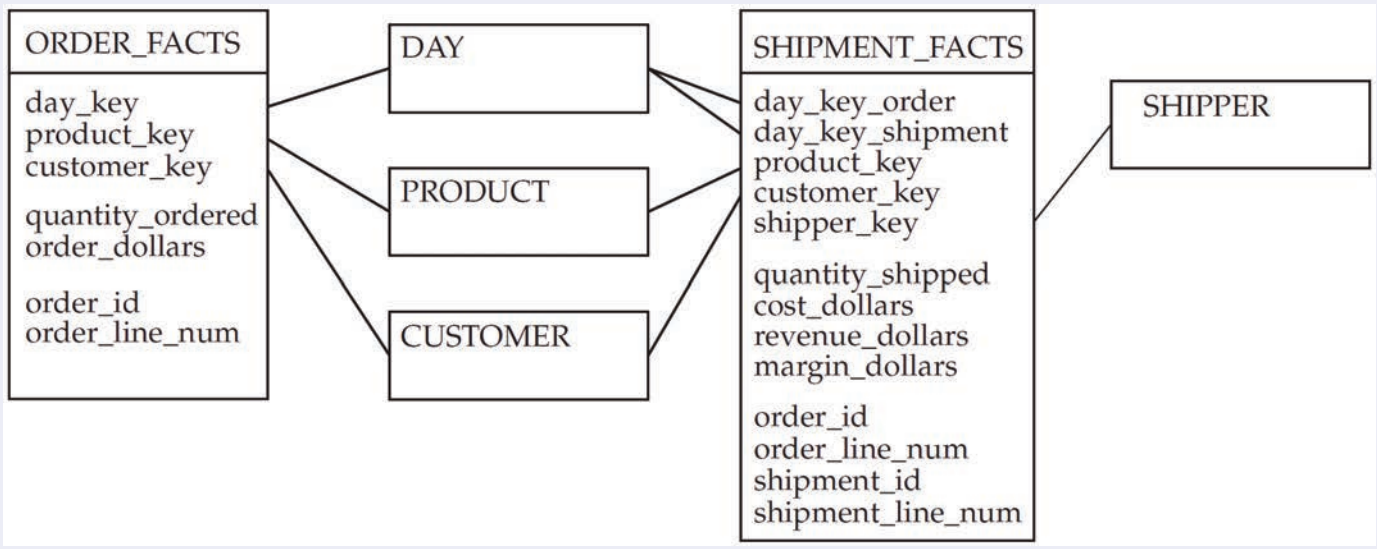
Dimensiones

Hechos con diferente granularidad

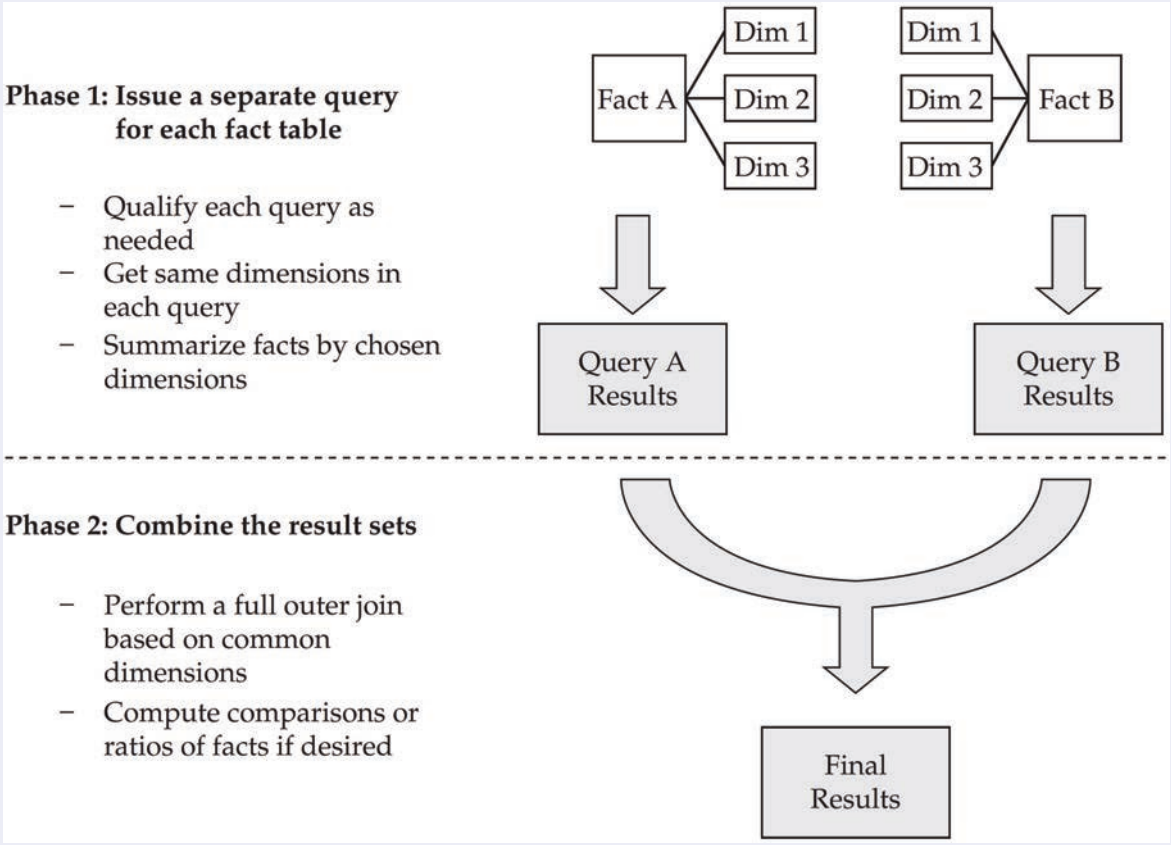
En el mismo cubo



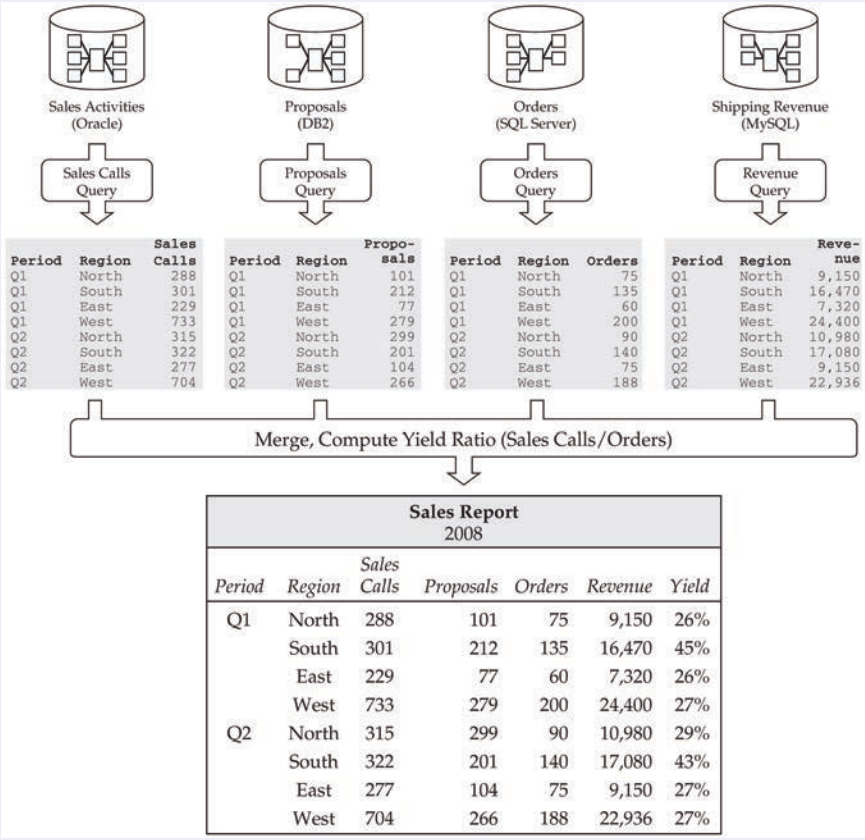
En cubos distintos



Operación *Drill-Across* (i)



Operación *Drill-Across* (y ii)



Period	Region	Sales Calls
Q1	North	288
Q1	South	301
Q1	East	229
Q1	West	733
Q2	North	315
Q2	South	322
Q2	East	277
Q2	West	704

Period	Region	Proposals
Q1	North	101
Q1	South	212
Q1	East	77
Q1	West	279
Q2	North	299
Q2	South	201
Q2	East	104
Q2	West	266

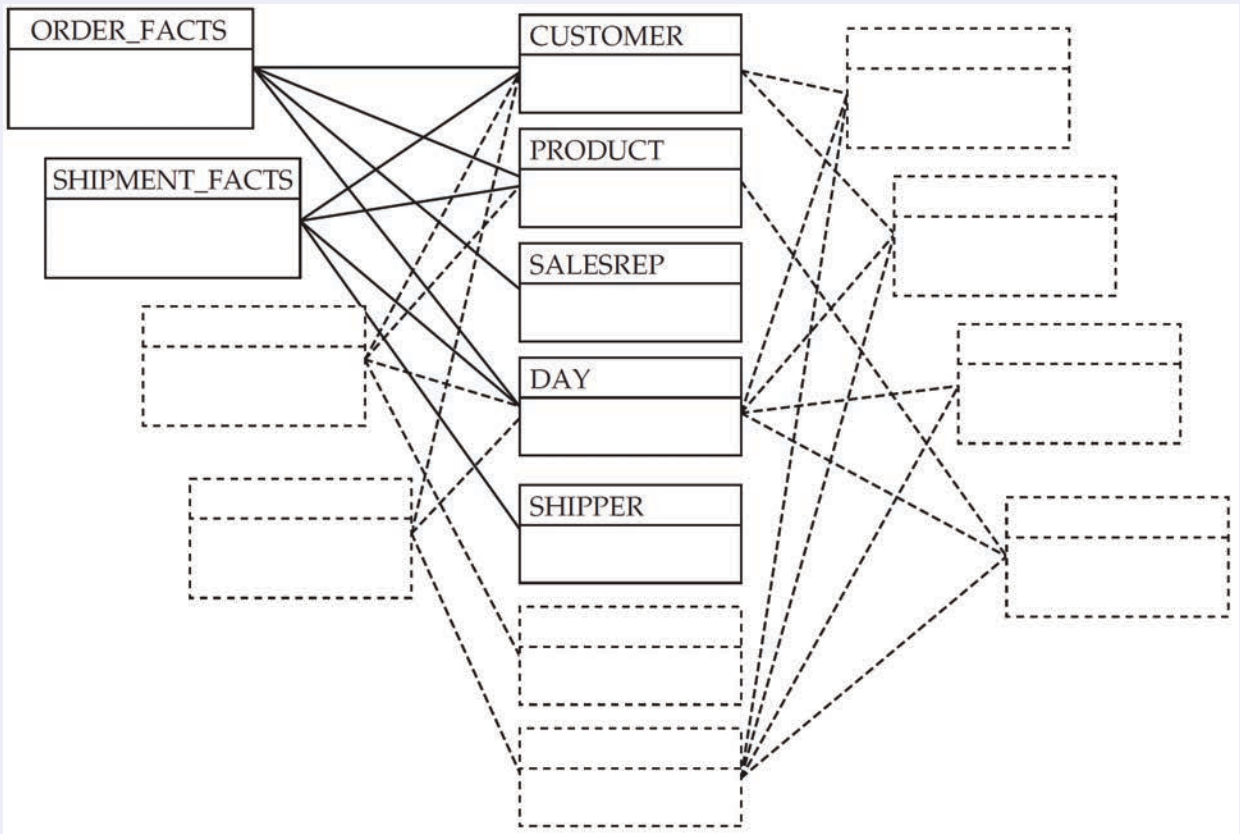
Period	Region	Orders
Q1	North	75
Q1	South	135
Q1	East	60
Q1	West	200
Q2	North	90
Q2	South	140
Q2	East	75
Q2	West	188

Period	Region	Revenue
Q1	North	9,150
Q1	South	16,470
Q1	East	7,320
Q1	West	24,400
Q2	North	10,980
Q2	South	17,080
Q2	East	9,150
Q2	West	22,936

Sales Report
2008

Period	*Region*	*Sales Calls*	*Proposals*	*Orders*	*Revenue*	*Yield*
Q1	North	288	101	75	9,150	26%
	South	301	212	135	16,470	45%
	East	229	77	60	7,320	26%
	West	733	279	200	24,400	27%
Q2	North	315	299	90	10,980	29%
	South	322	201	140	17,080	43%
	East	277	104	75	9,150	27%
	West	704	266	188	22,936	27%

Dimensiones conformadas

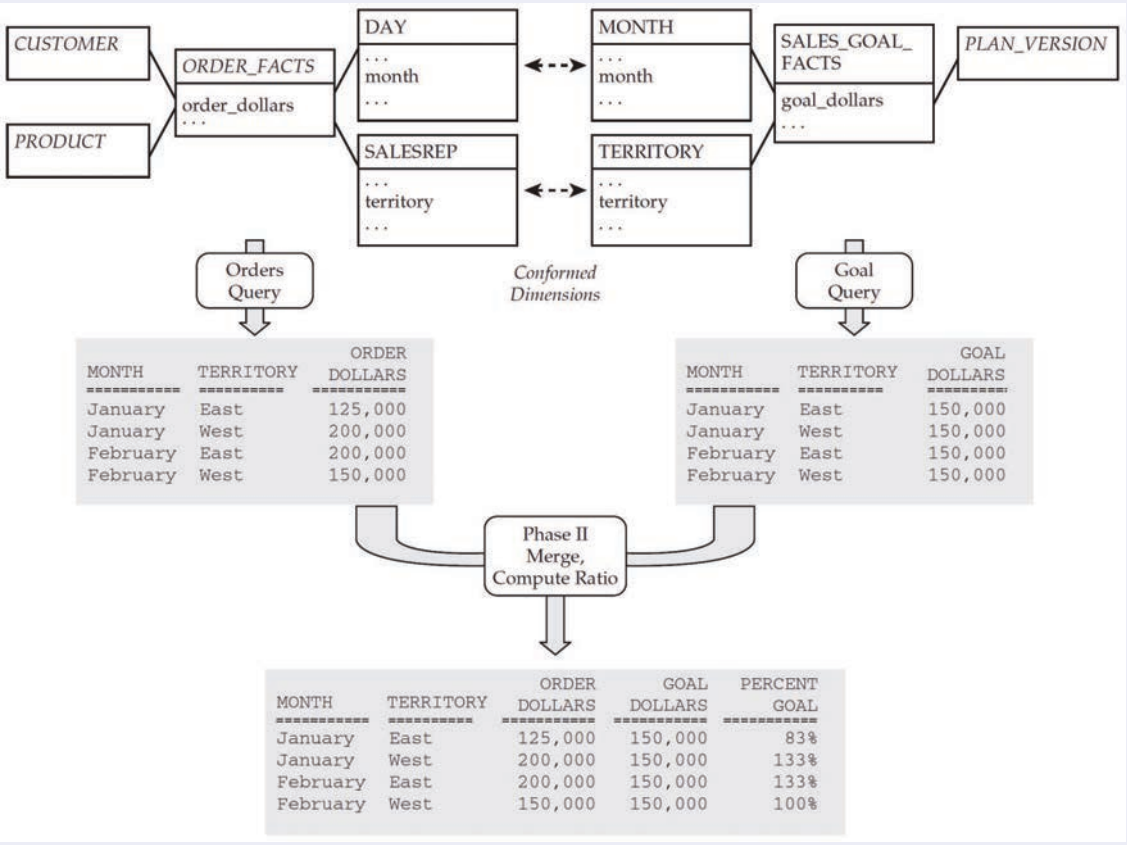


Matriz de conformación

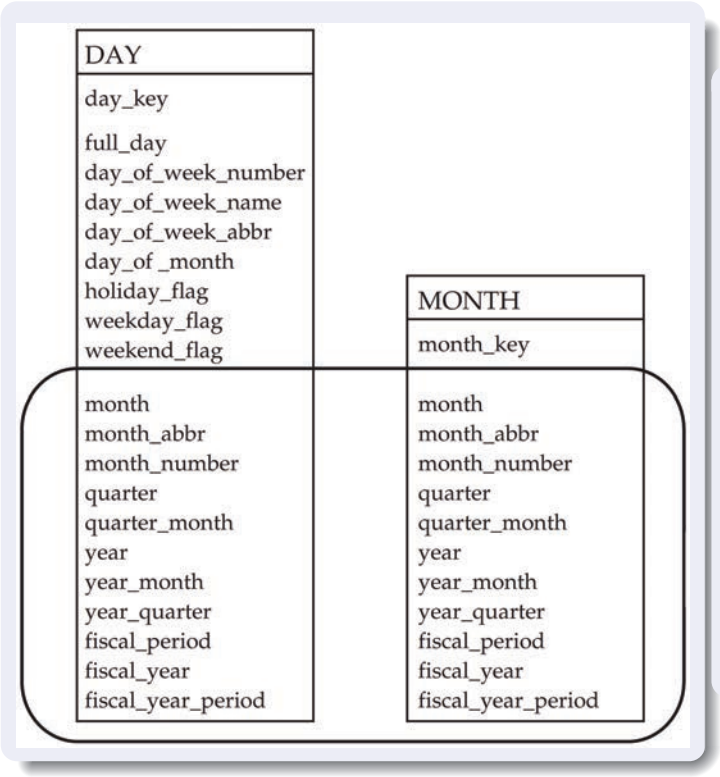
Data Warehouse Bus Architecture

	Time			Product			Sales Org.							
	Date	Month	Quarter	Product	Brand	Category	Salesperson	Sales Region	Promotion	Customer	Order Type	Shipment Method	Return Reason	Warehouse
Order Facts	✓			✓			✓		✓	✓	✓			
Promotion Facts	✓			✓					✓					
Shipment Facts	✓			✓			✓		✓	✓	✓	✓		✓
Return Facts	✓			✓			✓		✓	✓	✓		✓	✓
Inventory Snapshot Facts	✓			✓										✓
Annual Plan Facts			✓			✓		✓						
Sales Forecast Facts		✓			✓			✓		✓				

Drill-Across con niveles comunes



Definición de niveles comunes



DAY

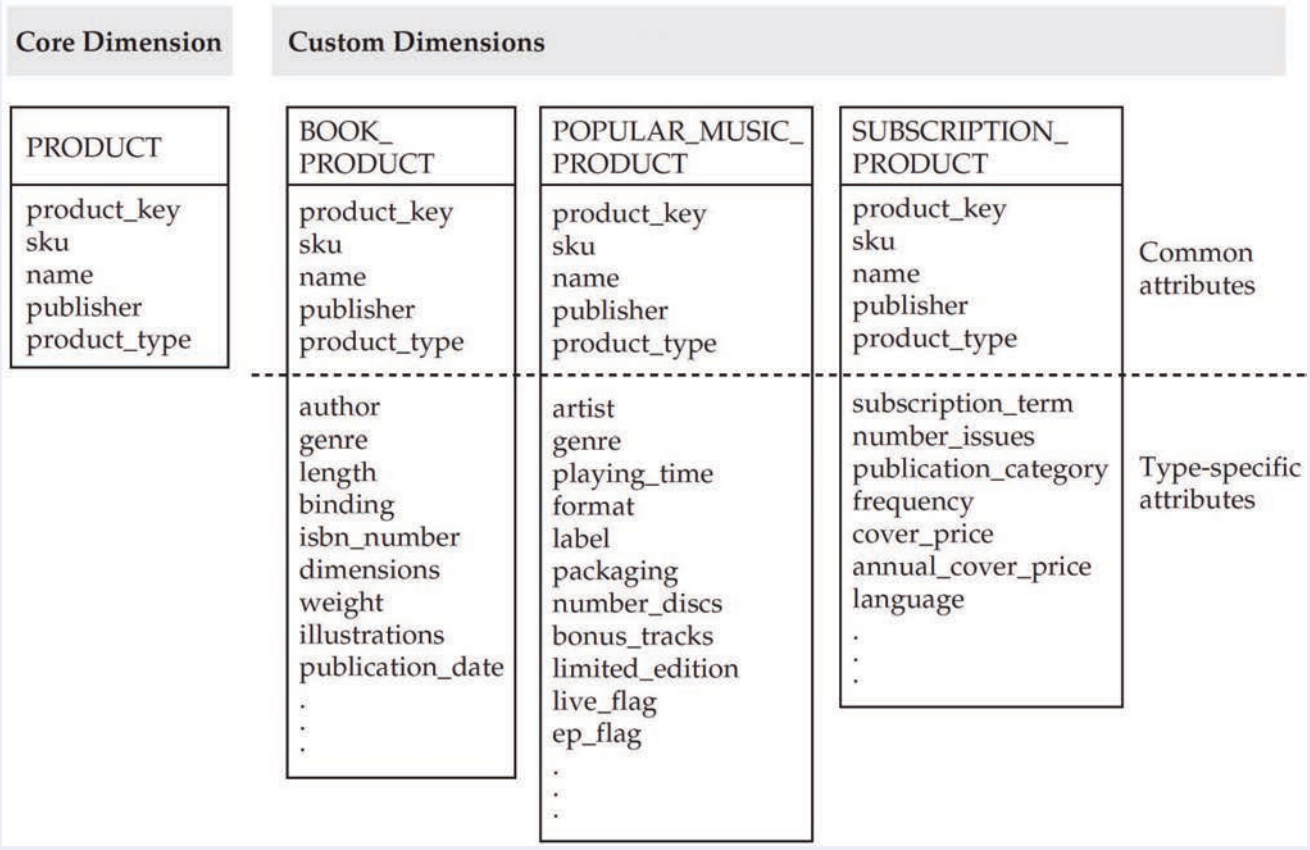
day_key	full_date	month	year	
755	1/30/2007	January	2007	
766	1/31/2007	January	2007	
767	2/1/2007	February	2007	
768	2/2/2007	February	2007	
769	2/3/2007	February	2007	

MONTH

month_key	month	year	
10	January	2007	
11	February	2007	

Especialización de dimensiones

Atributos específicos según el tipo



Correspondencia entre instancias de las dimensiones (i)

PRODUCT

product_key	sku	name	product_type
100	B221101	Summer in New York	Book
101	N22911	Music Criticism Journal	Subscription
102	B33921	Gadget Review Weekly	Subscription
103	A220022	The Contrarian Pathway	Book
104	Q27822	Cooking and Decorating	Subscription
105	C770077	Havana, 3AM	Book

Same attribute content
Same key values

BOOK_PRODUCT

product_key	sku	name	product_type	author	genre	cover_type
100	B221101	Summer in New York	Book	Michael Jones	Travel	Paperback
103	A220022	The Contrarian Pathway	Book	John Mellor	Lifestyle	Paperback
105	C770077	Havana, 3AM	Book	P. Gustave Simonon	Fiction	Hardcover

Correspondencia entre instancias de las dimensiones (y ii)

PRODUCT

product_ key	sku	name	product_ type
100	B221101	Summer in New York	Book
101	N22911	Music Criticism Journal	Subscription
102	B33921	Gadget Review Weekly	Subscription
103	A220022	The Contrarian Pathway	Book
104	Q27822	Cooking and Decorating	Subscription
105	C770077	Havana, 3AM	Book
201	C770077	Havana, 3AM	Book

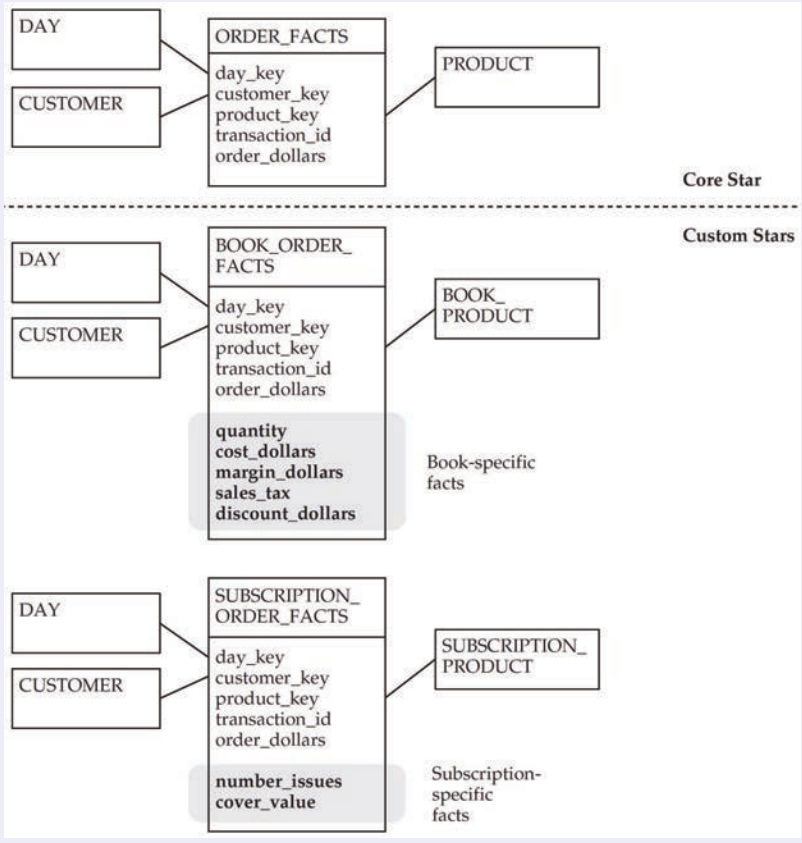
These rows are identical in the core dimension table, save for their surrogate keys.

They were precipitated by a type 2 change to the product's genre.

BOOK_PRODUCT

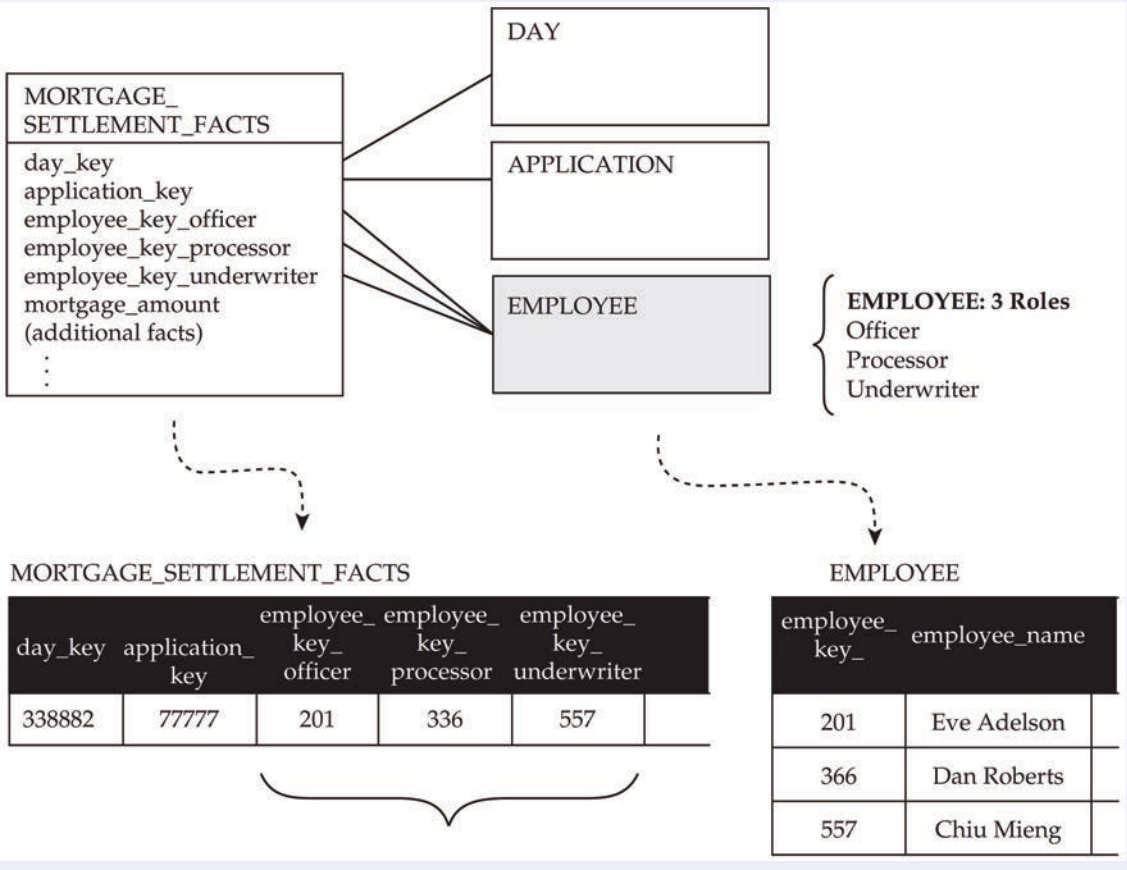
product_ key	sku	name	product_ type	author	genre	cover_ type	
100	B221101	Summer in New York	Book	Michael Jones	Travel	Paperback	
103	A220022	The Contrarian Pathway	Book	John Mellor	Lifestyle	Paperback	
105	C770077	Havana, 3AM	Book	P. Gustave Simonon	Fiction	Hardcover	
201	C770077	Havana, 3AM	Book	P. Gustave Simonon	Classics	Hardcover	

Diseños común y personalizados

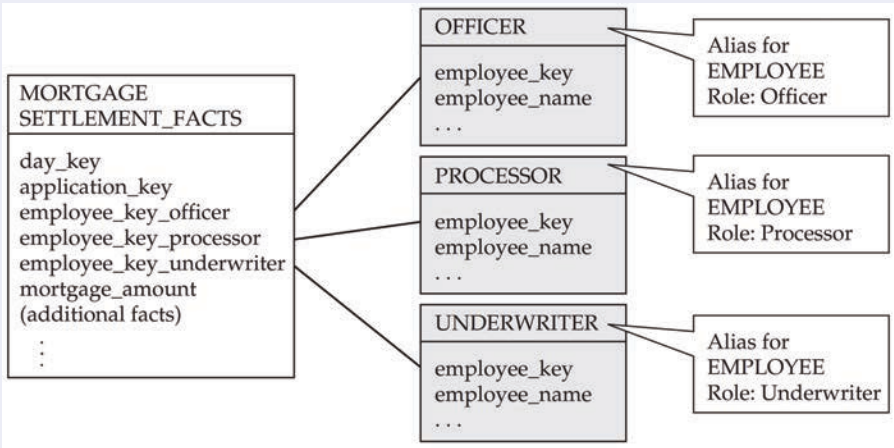


Dimensiones con varios papeles

Una sola tabla para varias dimensiones



Uso de alias



SQL Query

```
SELECT
   officer.employee_name AS officer_name,
   processor.employee_name AS processor_name,
   underwriter.employee_name AS underwriter_name
FROM
   -- Alias the employee table 3 times:
   --
   employee ALIAS officer,
   employee ALIAS processor,
   employee ALIAS underwriter,
   --
   --
   mortgage_closing_facts
WHERE
   --
```


Dimensión *Cuándo*

Tabla a nivel de Fecha

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
...

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

- Literales en lugar de Sí/No, Cierto/Falso o códigos.

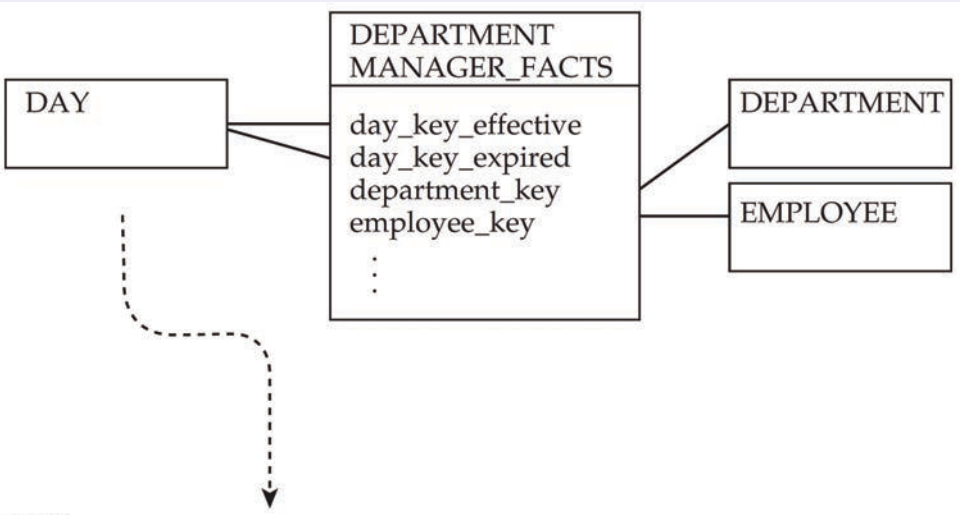
Fecha y hora

- *Fecha* (p.e., 20 de Marzo de 2020),
- *Hora* (p.e., 10:15).
- Dimensión *Cuándo* a nivel de hora (minuto): $24 \times 60 = 1440$ minutos; en un año: $356 \times 1440 = 525600$ instancias.
- Para 5 años...
- A nivel de segundo...
- Dimensión *Fecha* y dimensión *Hora*.

Dimensión *Hora*

- Subparte del día, parte del día, ...

Eventos futuros

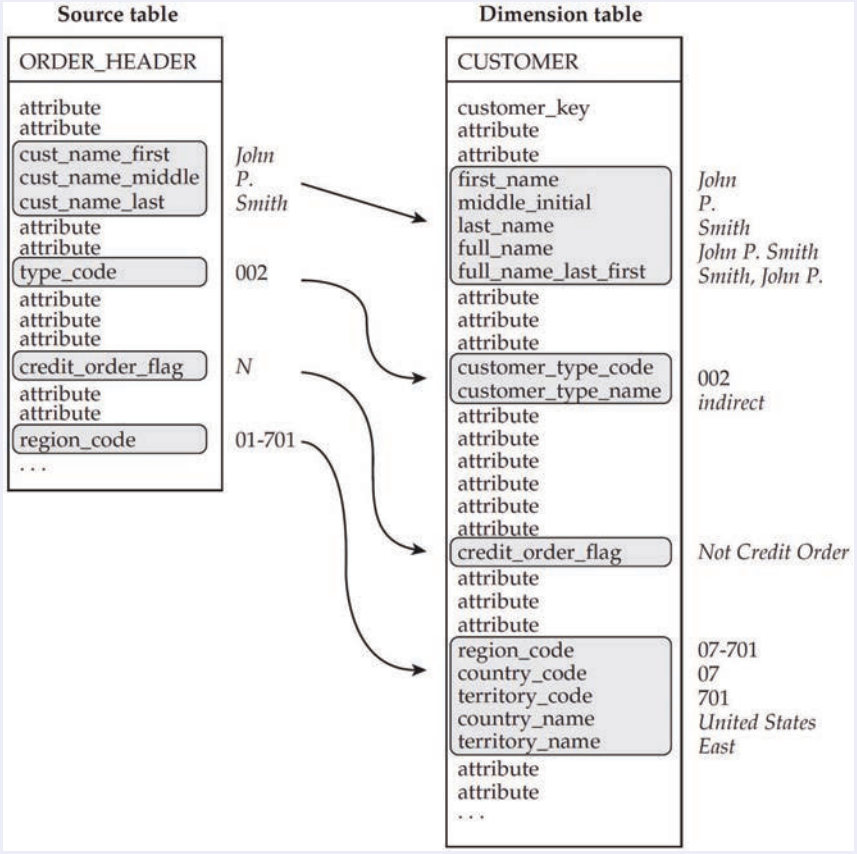


DAY

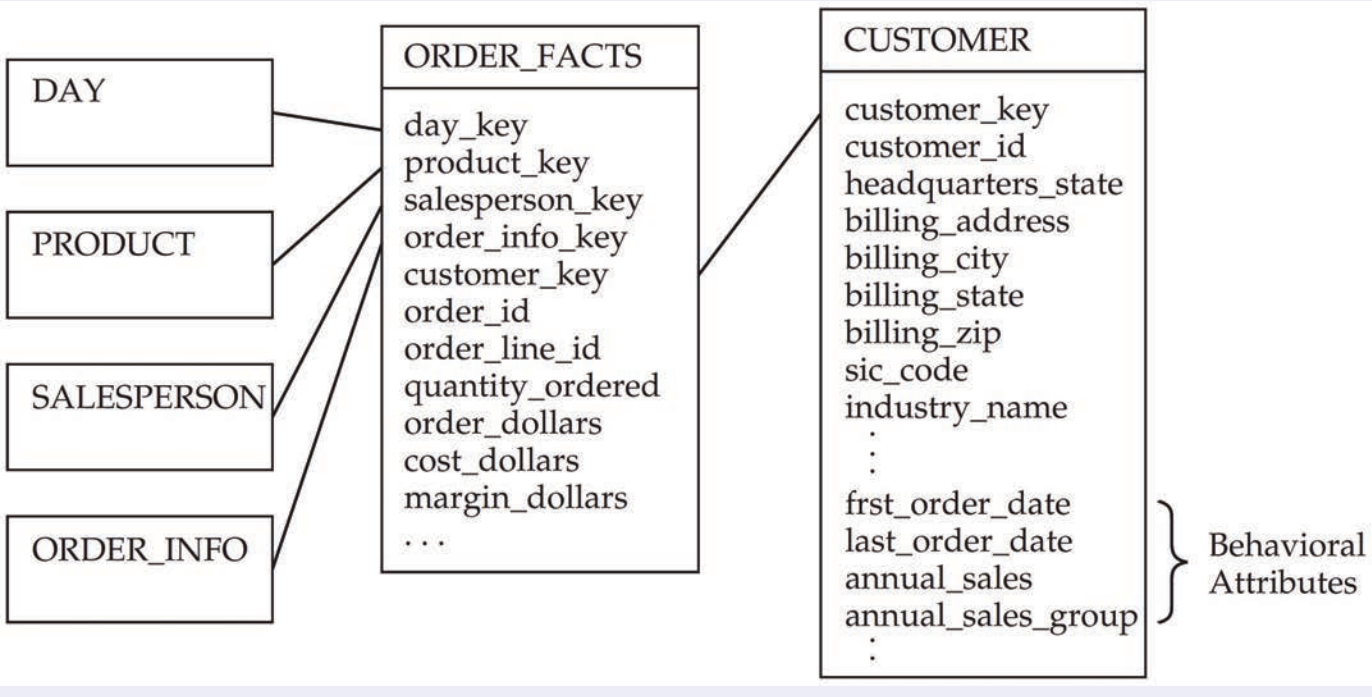
day_key	row_type	full_date	month	
0	Future date	12/31/9999	n/a	
101	Day	1/1/2004	January	
102	Day	1/2/2004	January	
103	Day	1/3/2004	January	

Enriquecimiento de dimensiones

Atributos descriptivos

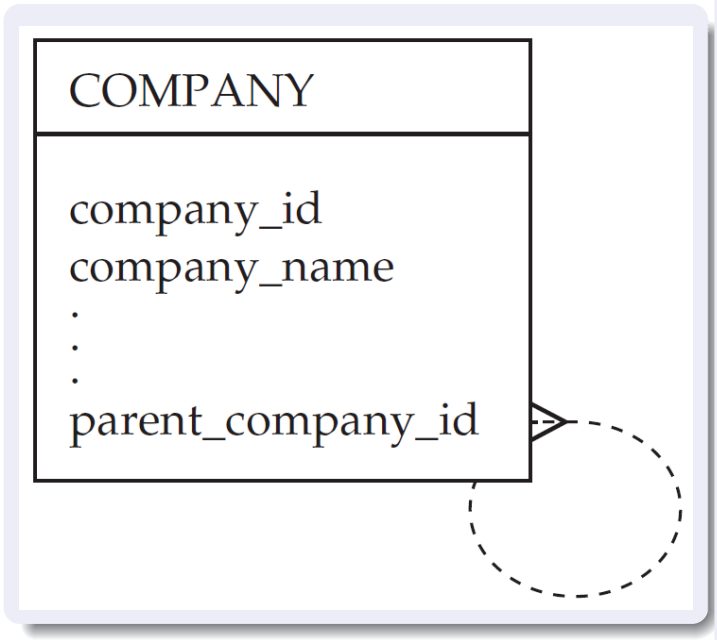


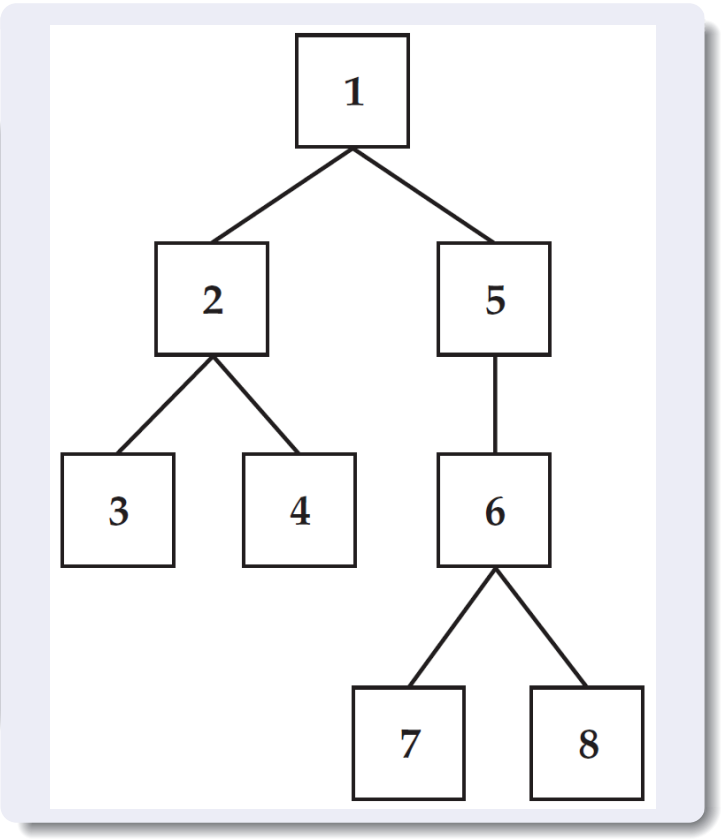
Dimensiones con datos obtenidos a partir de los hechos



Jerarquías no-balanceadas

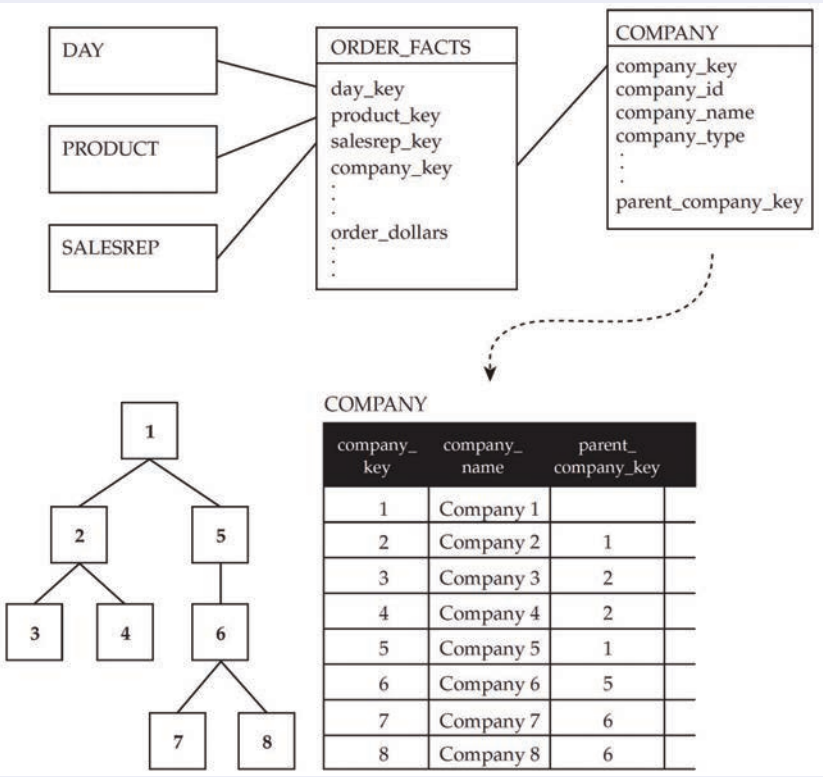
Relación y ejemplo de instancias





Instancias en la tabla de la dimensión

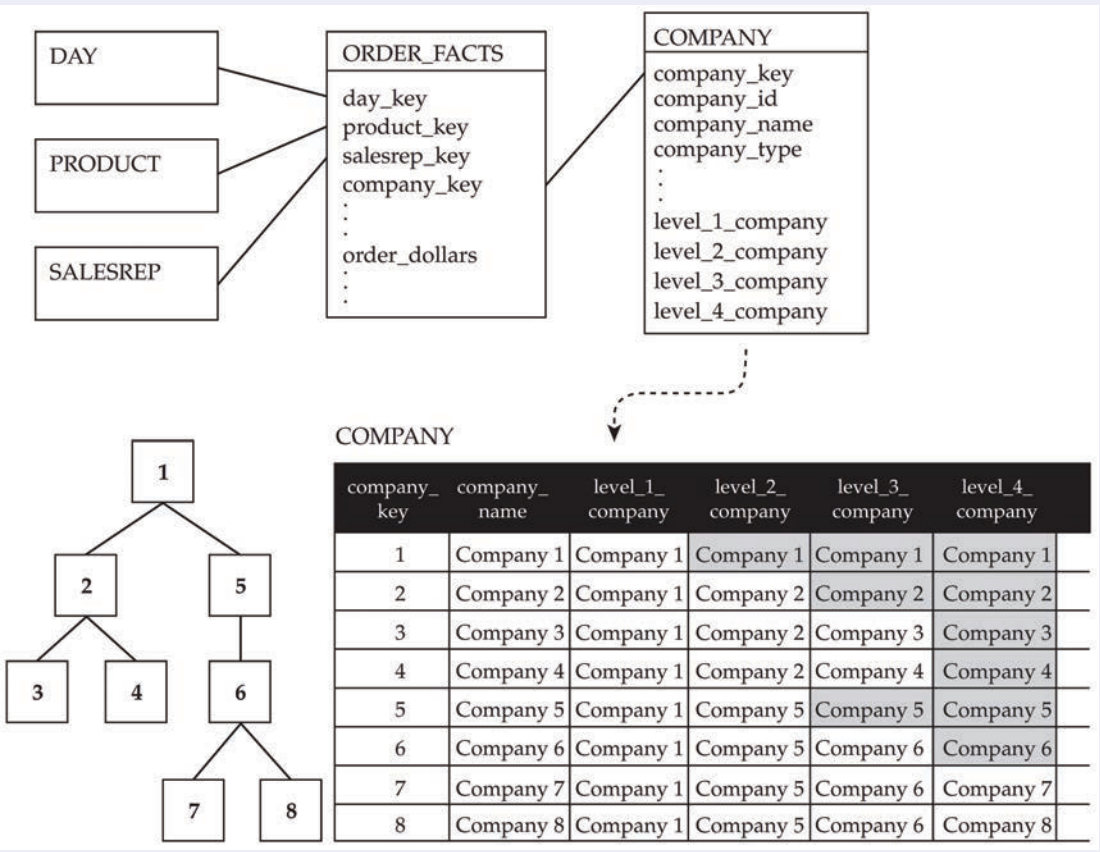
Consultas recursivas en SQL



COMPANY

company_ key	company_ name	parent_ company_key
1	Company 1	
2	Company 2	1
3	Company 3	2
4	Company 4	2
5	Company 5	1
6	Company 6	5
7	Company 7	6
8	Company 8	6

Dimensión aplanada

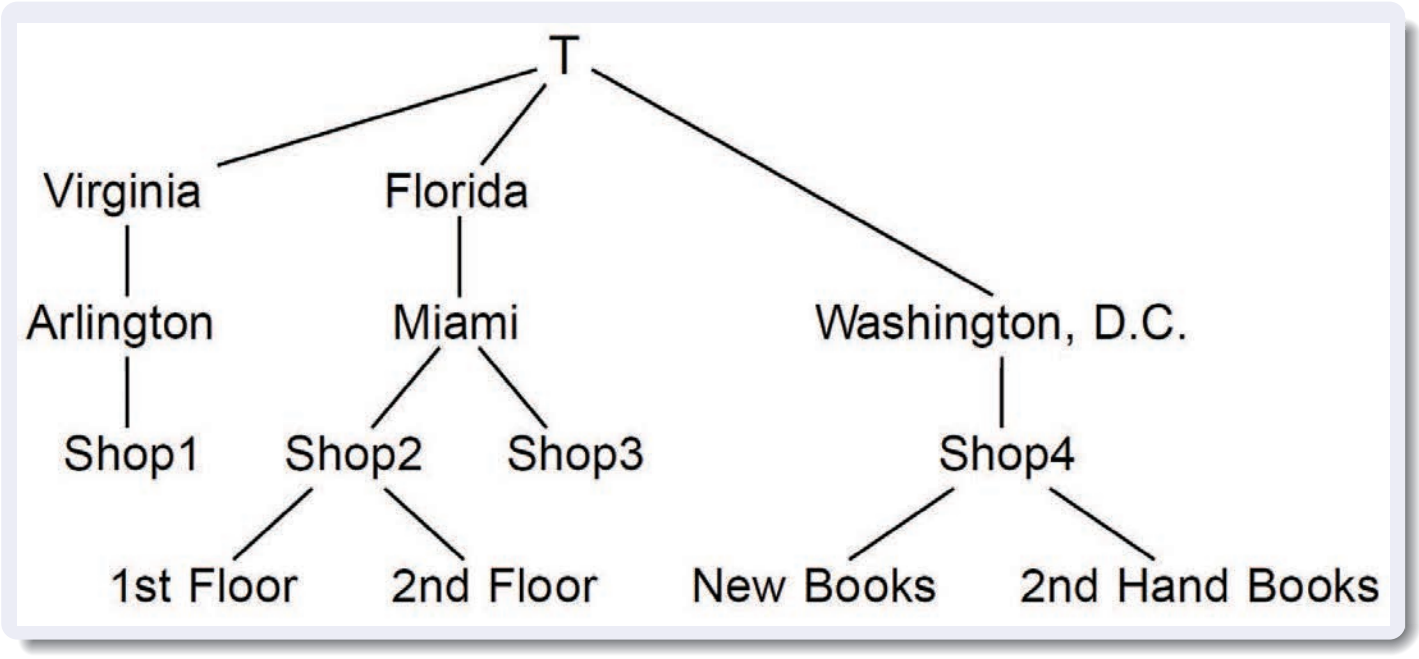


COMPANY

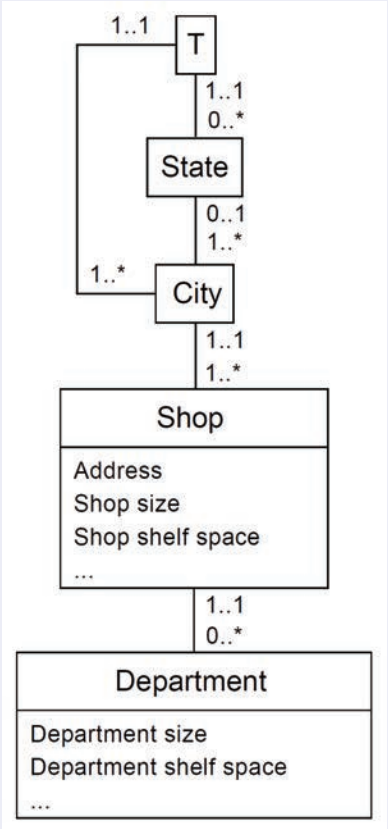
company_ key	company_ name	level_1_ company	level_2_ company	level_3_ company	level_4_ company
1	Company 1	Company 1	Company 1	Company 1	Company 1
2	Company 2	Company 1	Company 2	Company 2	Company 2
3	Company 3	Company 1	Company 2	Company 3	Company 3
4	Company 4	Company 1	Company 2	Company 4	Company 4
5	Company 5	Company 1	Company 5	Company 5	Company 5
6	Company 6	Company 1	Company 5	Company 6	Company 6
7	Company 7	Company 1	Company 5	Company 6	Company 7
8	Company 8	Company 1	Company 5	Company 6	Company 8

Jerarquías irregulares

Ejemplo de instancias en una jerarquía irregular

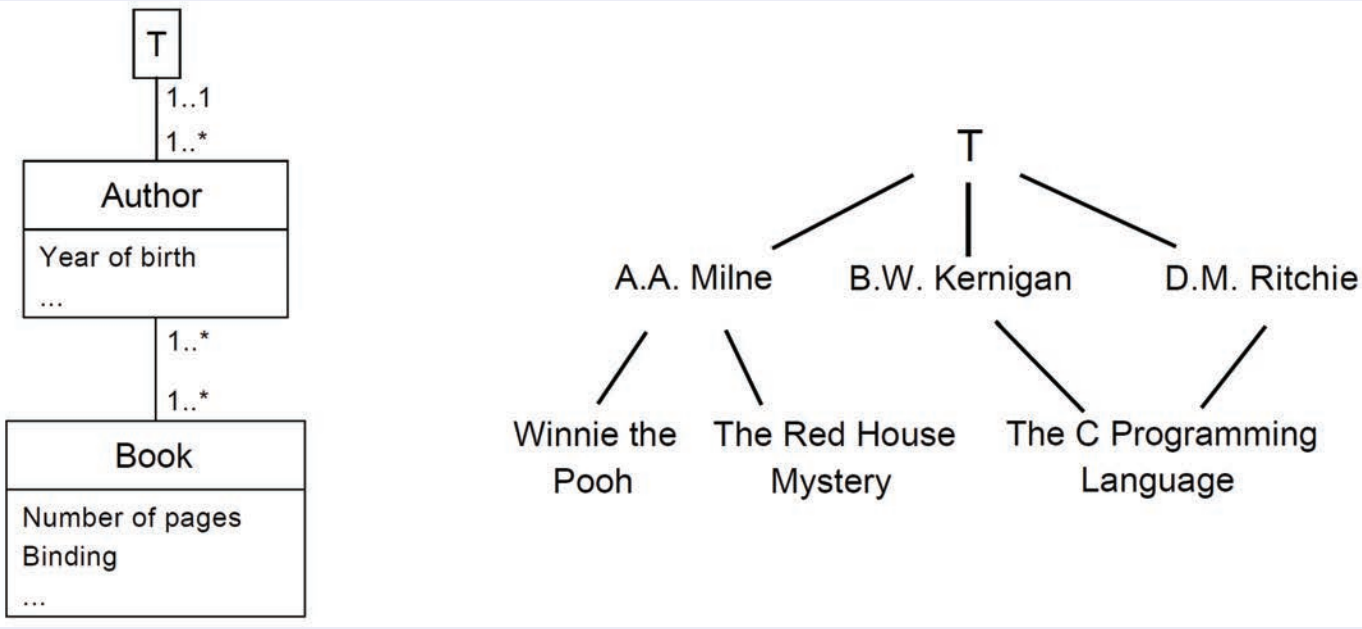


Jerarquía irregular

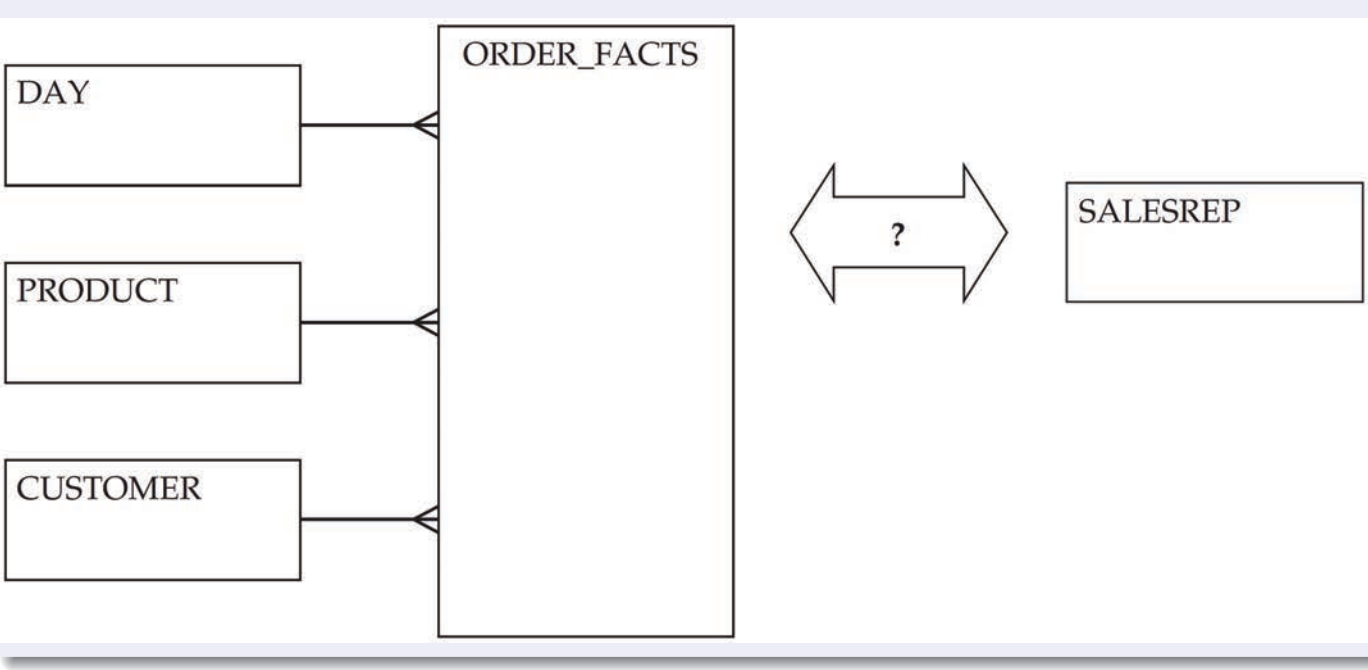


Jerarquías no-estrictas

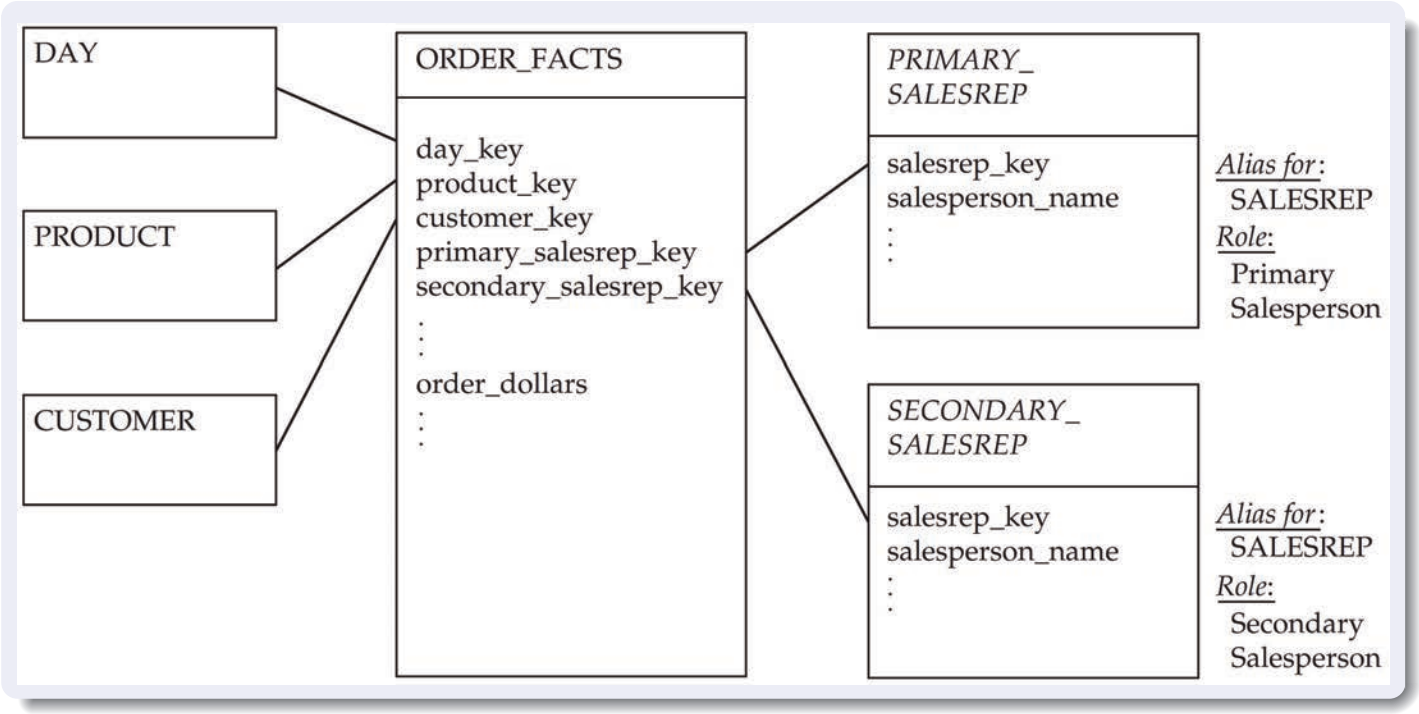
Ejemplo de jerarquía no-estricta



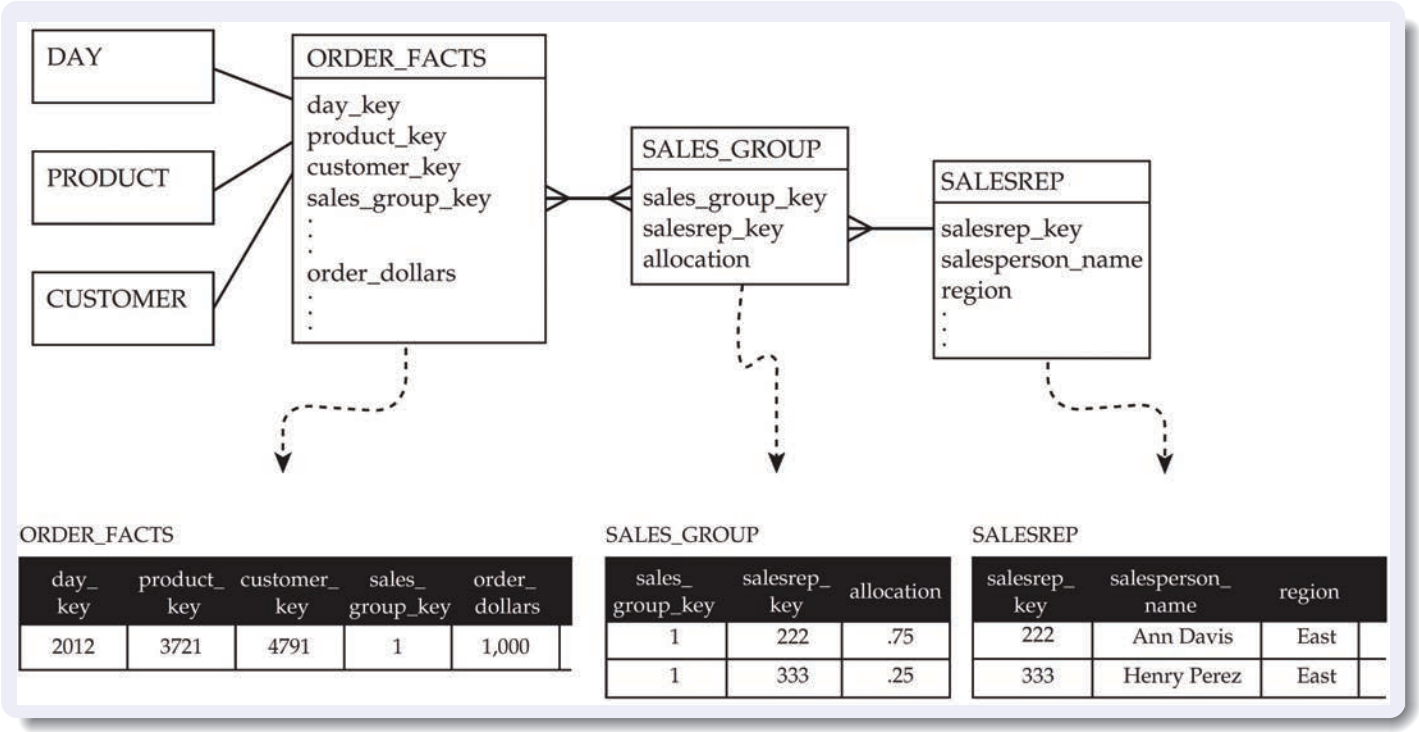
Relación N:M



Simplificando la relación



Dimensiones multivaluadas y tablas puente (i)



ORDER_FACTS

day_ key	product_ key	customer_ key	sales_ group_key	order_ dollars
2012	3721	4791	1	1,000

SALES_GROUP

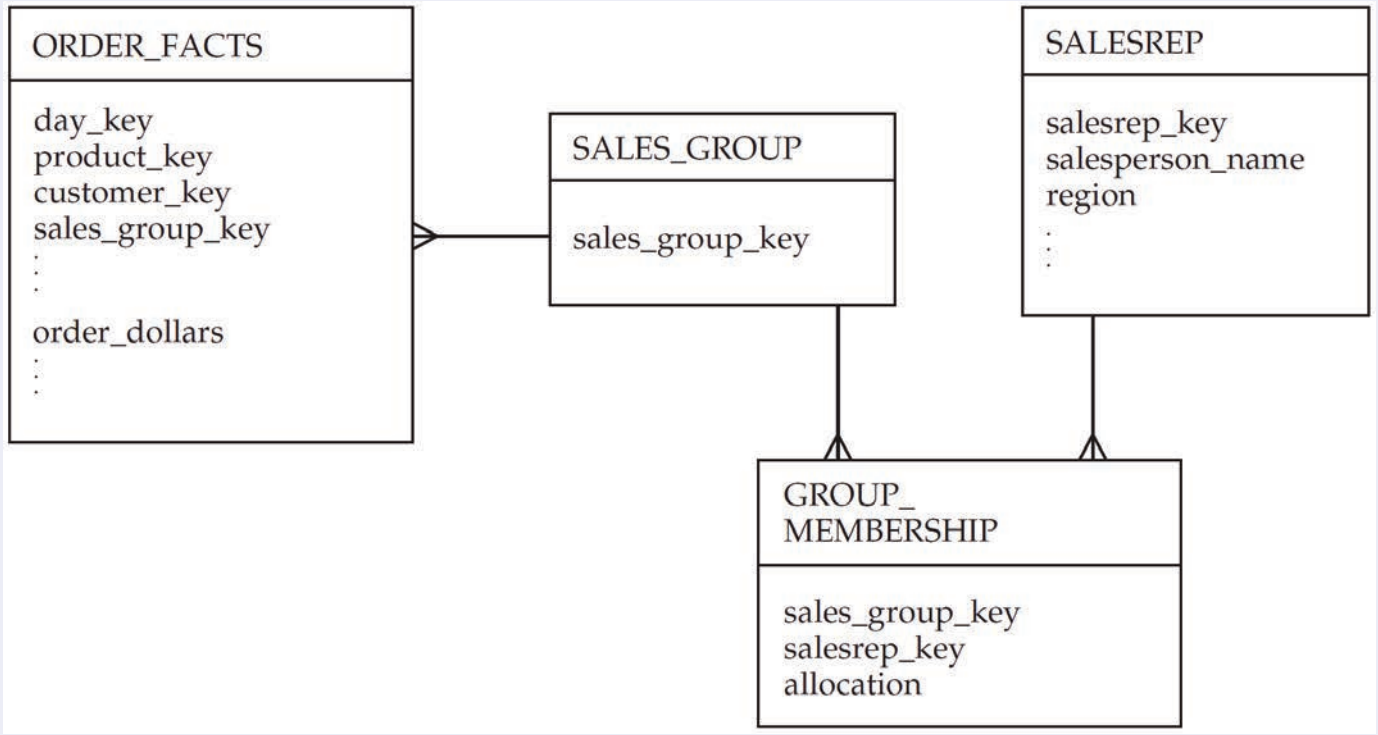
sales_ group_key	salesrep_ key	allocation
1	222	.75
1	333	.25

SALESREP

salesrep_ key	salesperson_ name	region
222	Ann Davis	East
333	Henry Perez	East

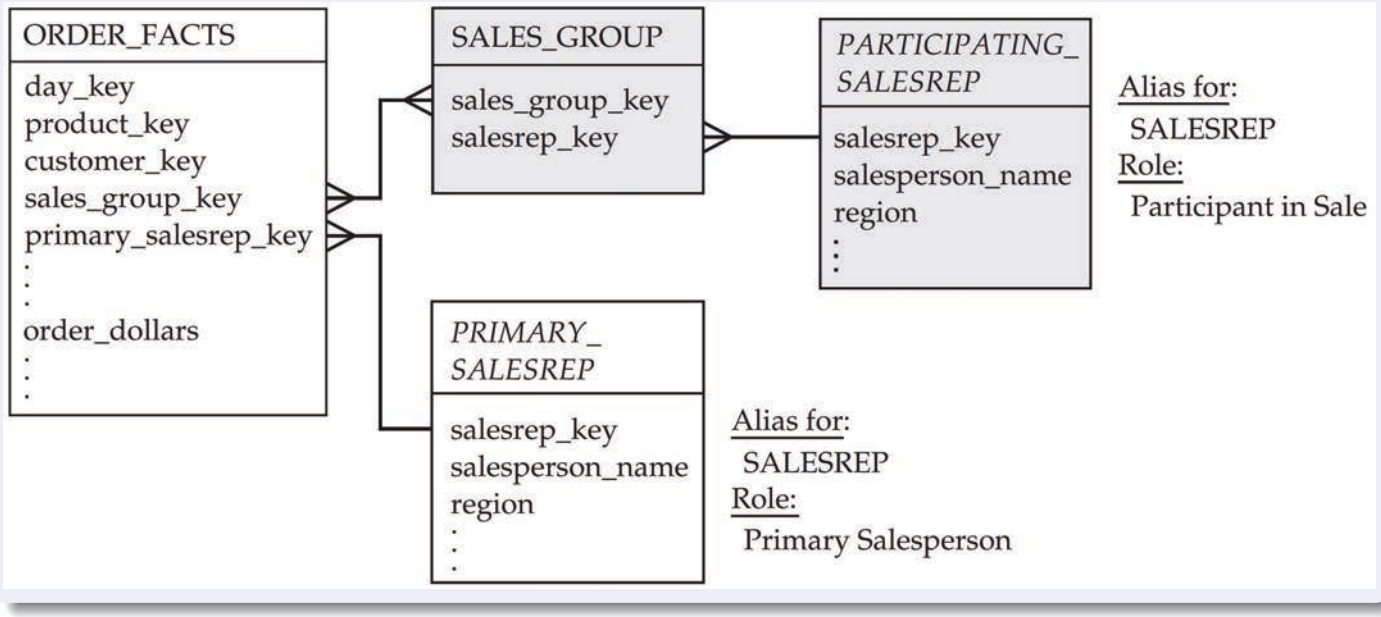
Dimensiones multivaluadas y tablas puente (y ii)

Resolviendo la relación N:M

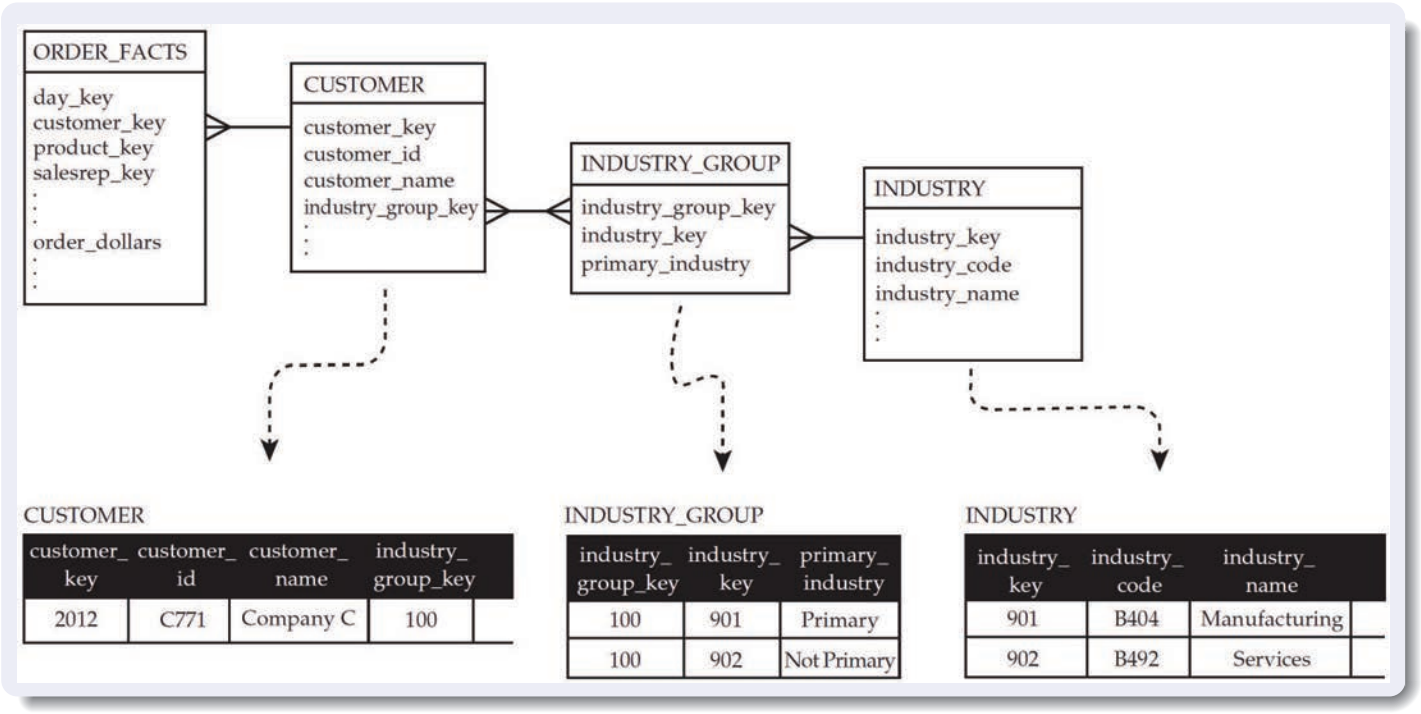


Solución combinada

Para análisis sencillo y sofisticado



Tablas puente a otros niveles en las jerarquías



CUSTOMER

customer_ key	customer_ id	customer_ name	industry_ group_key
2012	C771	Company C	100

INDUSTRY_GROUP

industry_ group_key	industry_ key	primary_ industry
100	901	Primary
100	902	Not Primary

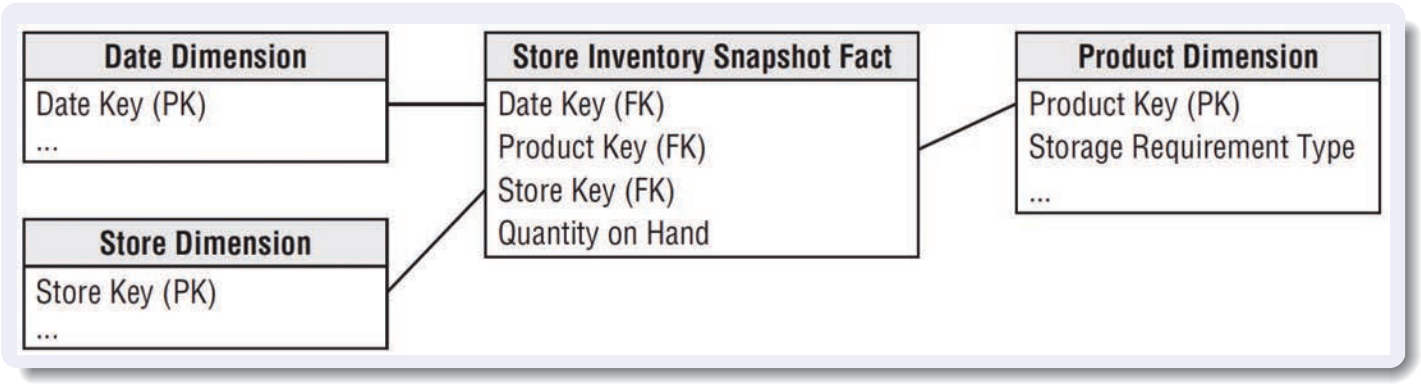
INDUSTRY

industry_ key	industry_ code	industry_ name
901	B404	Manufacturing
902	B492	Services

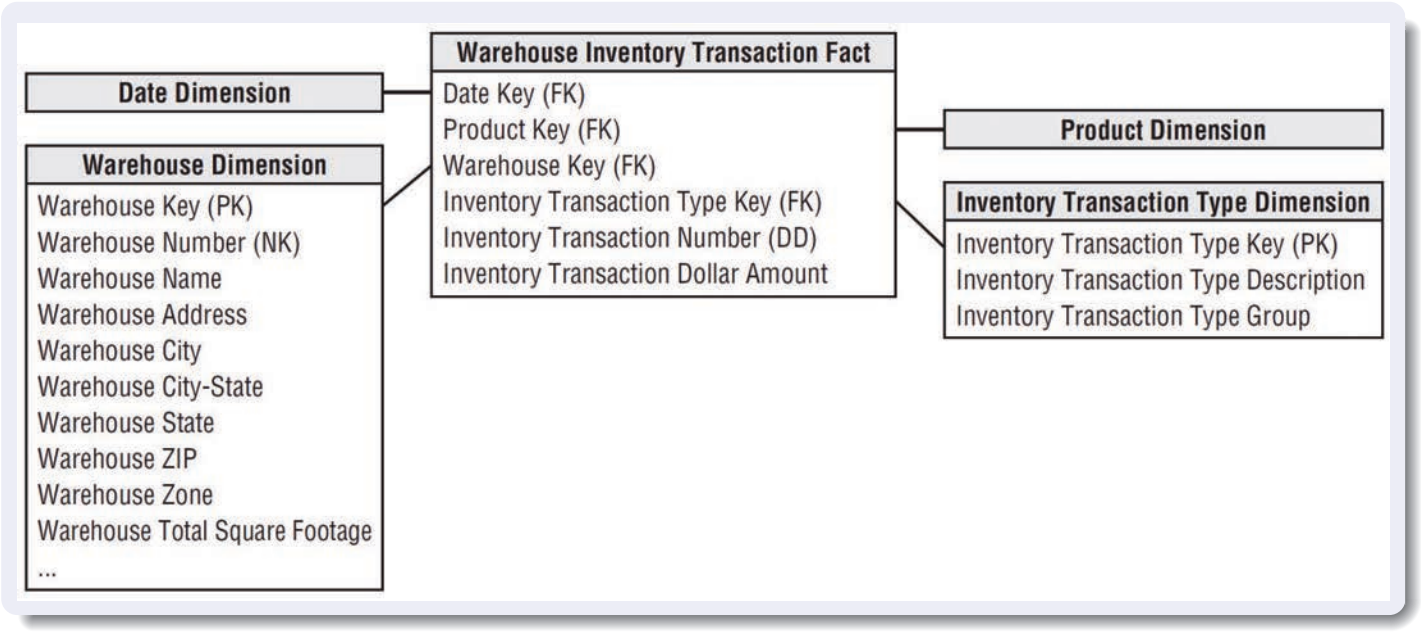
Hechos

Tipos de hechos

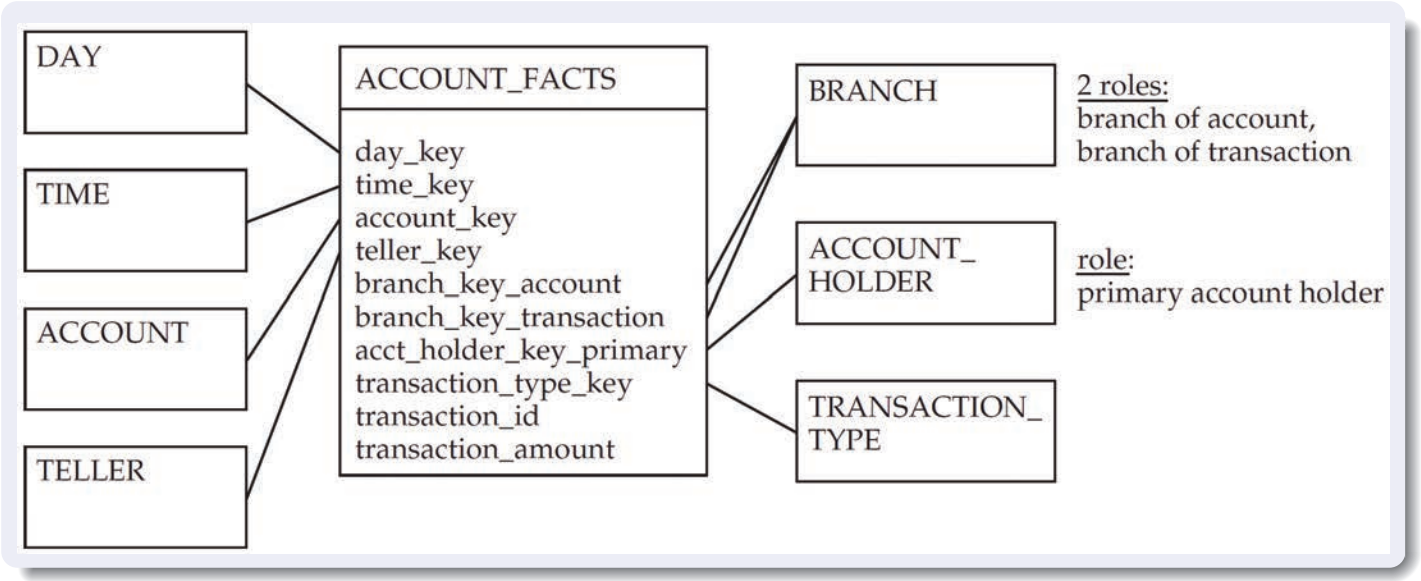
Inventario (i)



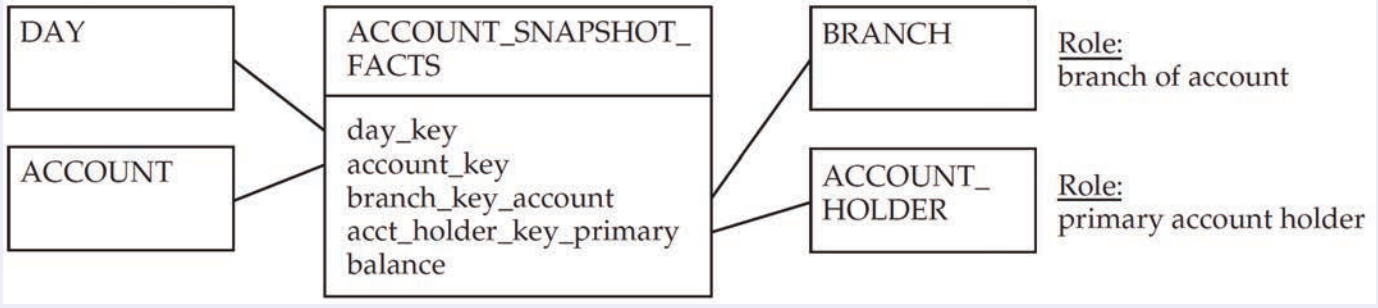
Inventario (y ii)



Contabilidad (i)



Contabilidad (y ii)



Transacciones vs. instantánea

Account: 7922-3002
Period: 2/1/2009 – 2/14/2009

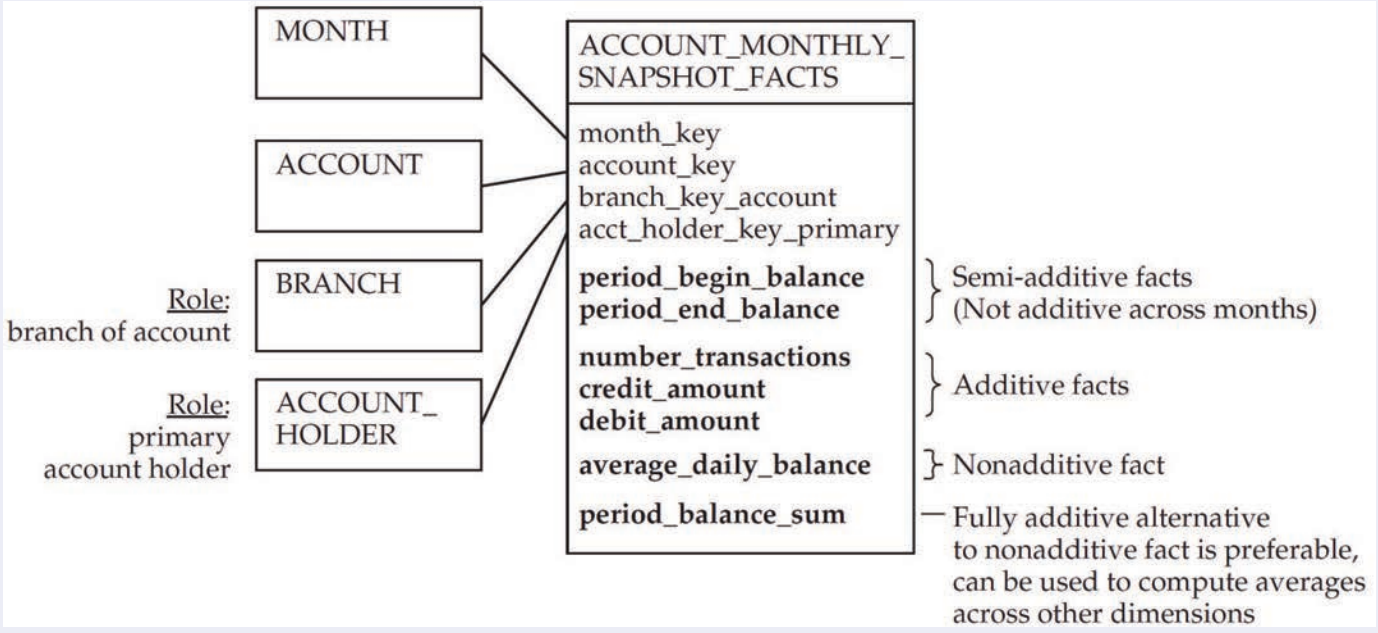
Granular transaction data stored in account_facts:

Day	Transaction Type	Transaction Amount
2/1/2009	Initial Deposit	2000.00
2/2/2009	Withdrawal	(20.00)
2/3/2009	Check	(35.50)
2/3/2009	Check	(17.02)
2/6/2009	Check	(75.00)
2/6/2009	Deposit	75.00
2/7/2009	Check	(800.00)
2/10/2009	Check	(68.29)
2/14/2009	Withdrawal	(100.00)

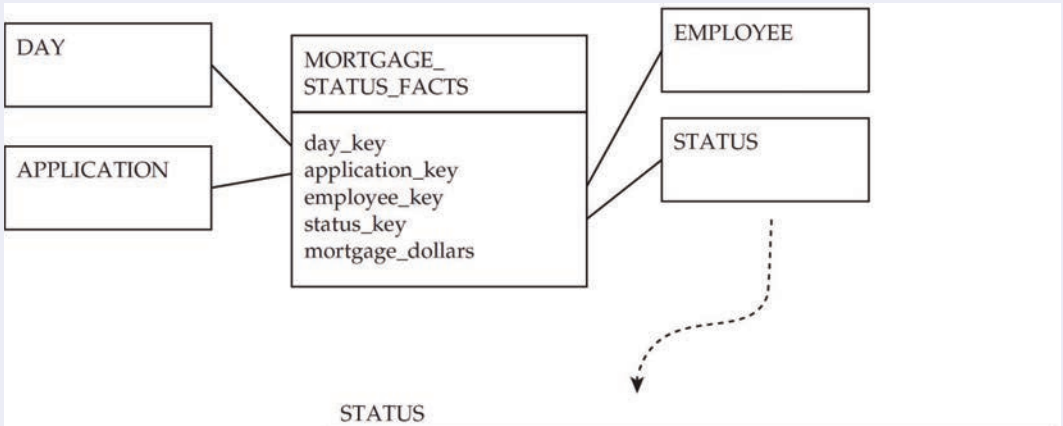
Periodic status data stored in account_snapshot_facts:

Day	Balance	
2/1/2009	2000.00	
2/2/2009	1980.00	Two transactions occurred on this day
2/3/2009	1927.48	
2/4/2009	1927.48	No transactions occurred on these days
2/5/2009	1927.48	
2/6/2009	1927.48	Offsetting transactions occurred on this day
2/7/2009	1127.48	
2/8/2009	1127.48	
2/9/2009	1127.48	
2/10/2009	1059.19	
2/11/2009	1059.19	
2/12/2009	1059.19	
2/13/2009	1059.19	
2/14/2009	959.19	

Resumen transaccional en instantánea



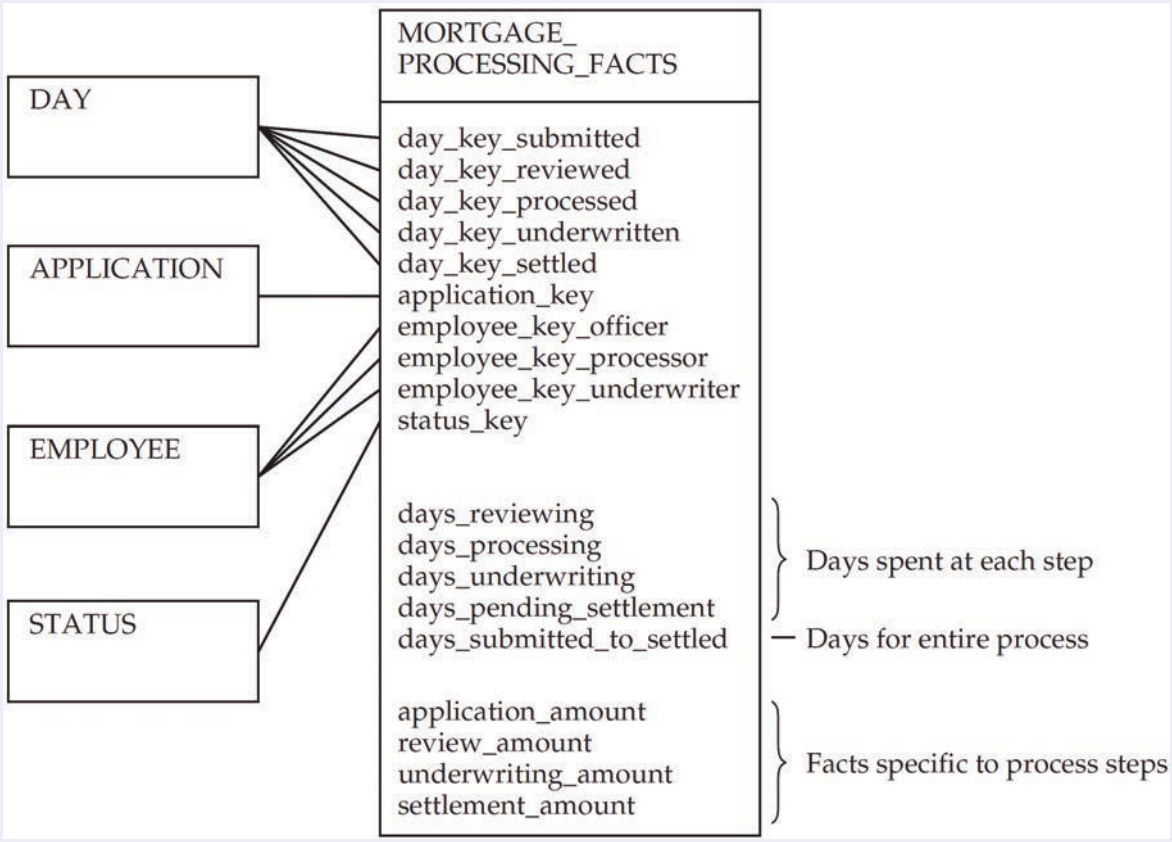
Posibles estados y cambios de estado



STATUS

status_ key	status_ code	status	status_description
1000	S1	Submitted	Submitted. Under review by mortgage officer.
1001	A2	Reviewed	Reviewed. Documentation being gathered by processor.
1002	P2	Processed	Processed. Under examination by underwriter.
1003	U2	Underwritten	Underwritten. Awaiting settlement.
1004	EX	Settled	Settled.

Instantánea acumulada



Instancias de los hechos

On Day 1 (Submitted; under review by officer):

application_ key	day_key_ submitted	day_key_ reviewed	day_key_ processed	day_key_ underwritten	day_key_ closing	application_ amount	review_ amount	underwriting_ amount	days_ reviewing	days_ processing	...
1011	**1021**	0000	0000	0000	0000	**100,000**	0	0	0	0	

Day 2 (No status change):

application_ key	day_key_ submitted	day_key_ reviewed	day_key_ processed	day_key_ underwritten	day_key_ closing	application_ amount	review_ amount	underwriting_ amount	days_ reviewing	days_ processing	...
1011	1021	0000	0000	0000	0000	100,000	0	0	**1**	0	

Days 3–9 (not shown)...

Day 10 (Reviewed; documents being gathered by processor):

application_ key	day_key_ submitted	day_key_ reviewed	day_key_ processed	day_key_ underwritten	day_key_ closing	application_ amount	review_ amount	underwriting_ amount	days_ reviewing	days_ processing	...
1011	1021	**1031**	0000	0000	0000	100,000	**90,000**	0	**9**	0	

Day 11 (No status change):

application_ key	day_key_ submitted	day_key_ reviewed	day_key_ processed	day_key_ underwritten	day_key_ closing	application_ amount	review_ amount	underwriting_ amount	days_ reviewing	days_ processing	...
1011	1021	1031	0000	0000	0000	100,000	90,000	0	9	**1**	

Remaining steps...

Simplificando los posibles estados

Status Codes from Operational System

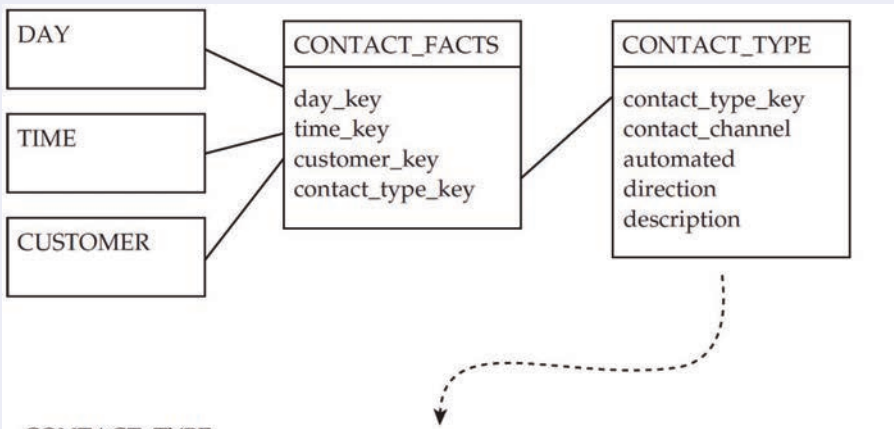
status_ code	category	status	
S1	Submitted	Submitted	Review activity
S2	Submitted	Under review by officer	
S3	Submitted	Awaiting customer response	
R1	Reviewed	Under review by processor	Processing activity
R2	Reviewed	Taxes requested	
R3	Reviewed	Paystubs requested	
R4	Reviewed	Credit fee requested	
R5	Reviewed	Taxes received	
R6	Reviewed	Paystubs received	
R7	Reviewed	Credit fee received	
R8	Reviewed	Credit report ordered	
U1	Processed	Forwarded to underwriter	Underwriting activity

Comparación de los tipos de hechos

	Transaction	Periodic Snapshot	Accumulating Snapshot
Periodicity	Discrete transaction point in time	Recurring snapshots at regular, predictable intervals	Indeterminate time span for evolving pipeline/workflow
Grain	1 row per transaction or transaction line	1 row per snapshot period plus other dimensions	1 row per pipeline occurrence
Date dimension(s)	Transaction date	Snapshot date	Multiple dates for pipeline's key milestones
Facts	Transaction performance	Cumulative performance for time interval	Performance for pipeline occurrence
Fact table sparsity	Sparse or dense, depending on activity	Predictably dense	Sparse or dense, depending on pipeline occurrence
Fact table updates	No updates, unless error correction	No updates, unless error correction	Updated whenever pipeline activity occurs

Hechos sin mediciones (*factless fact*)

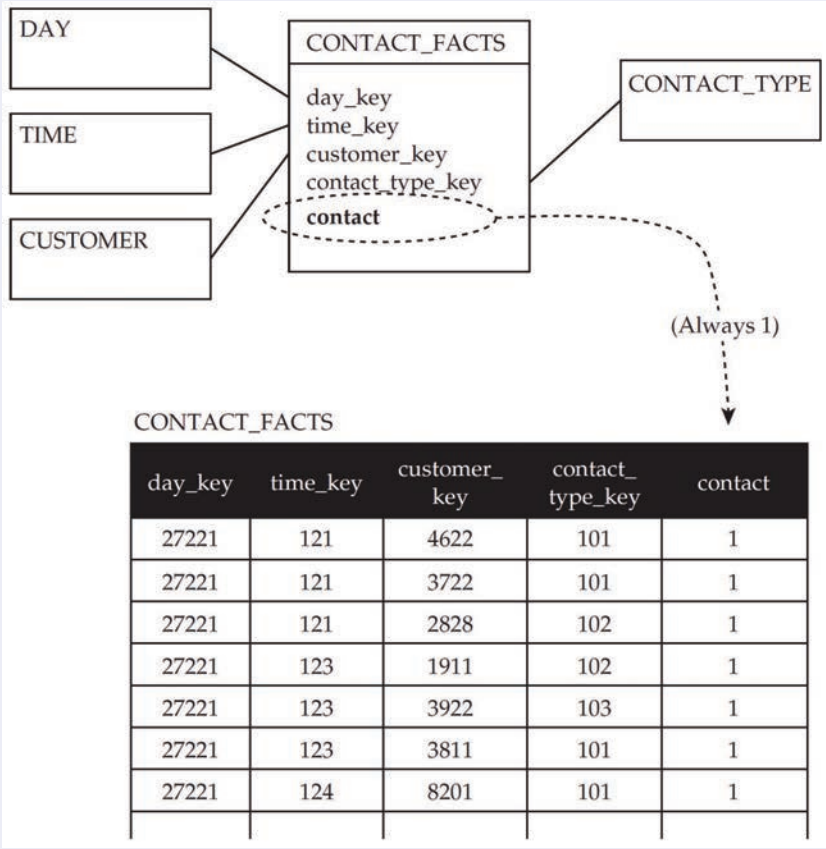
Eventos sin mediciones



CONTACT_TYPE

contact_ type_key	contact_ channel	automated	direction	description
100	E-Mail	Automated	Outbound	Weekly Coupons
101	E-Mail	Automated	Outbound	eNews Mailing
102	E-Mail	Not Automated	Inbound	Tech Support
103	E-Mail	Not Automated	Outbound	Tech Support
104	Telephone	Not Automated	Outbound	Promotional Offer
105	Telephone	Not Automated	Inbound	Tech Support
106	Telephone	Not Automated	Outbound	Tech Support

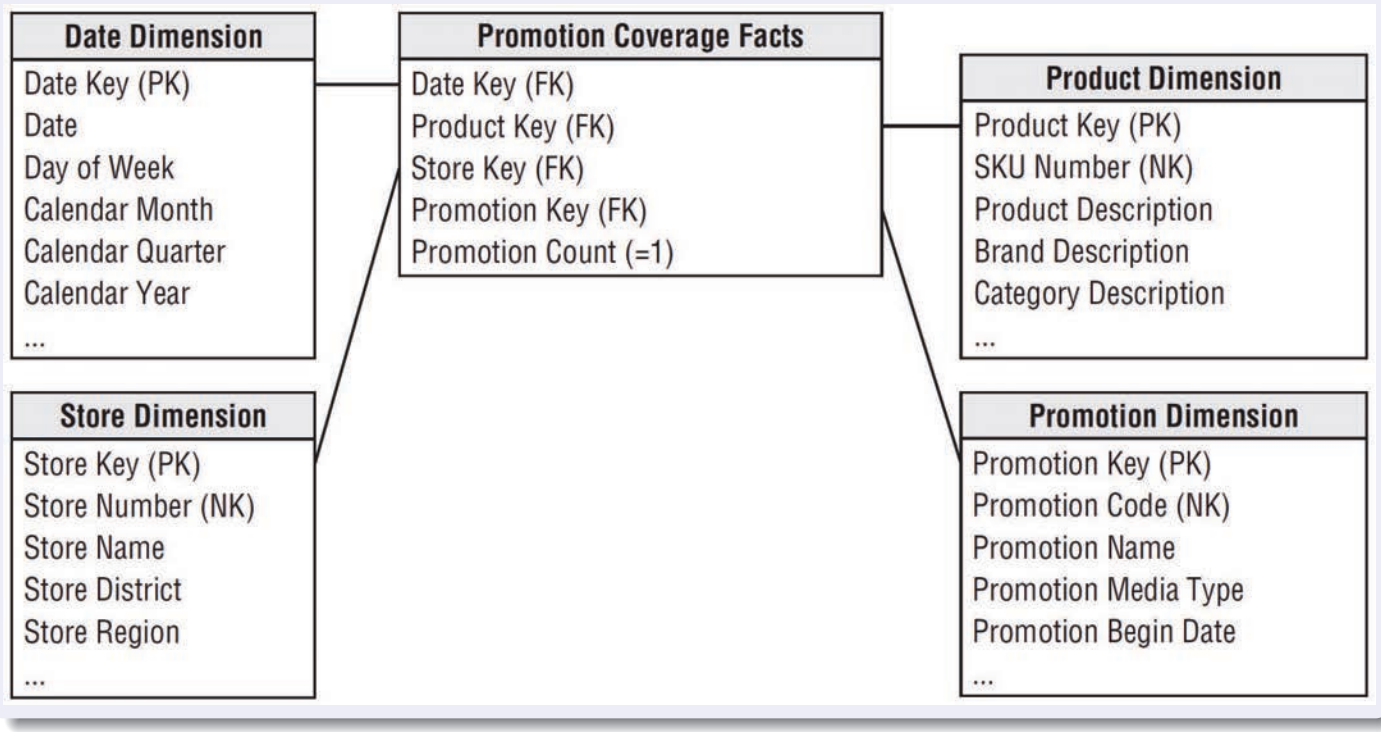
Medición añadida al evento



CONTACT_FACTS

day_key	time_key	customer_key	contact_type_key	contact
27221	121	4622	101	1
27221	121	3722	101	1
27221	121	2828	102	1
27221	123	1911	102	1
27221	123	3922	103	1
27221	123	3811	101	1
27221	124	8201	101	1

Promociones



Condiciones sin mediciones

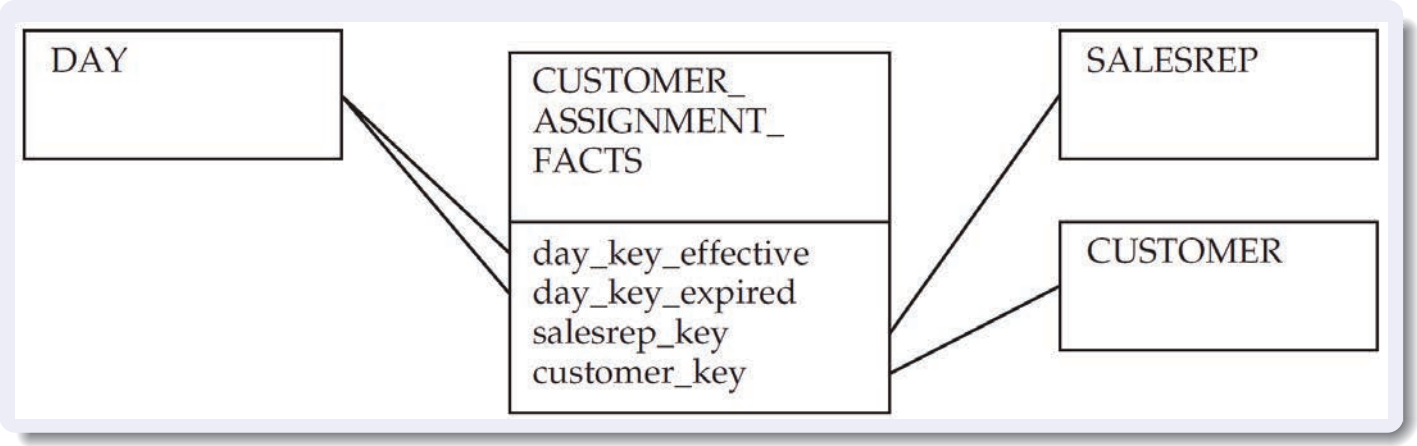
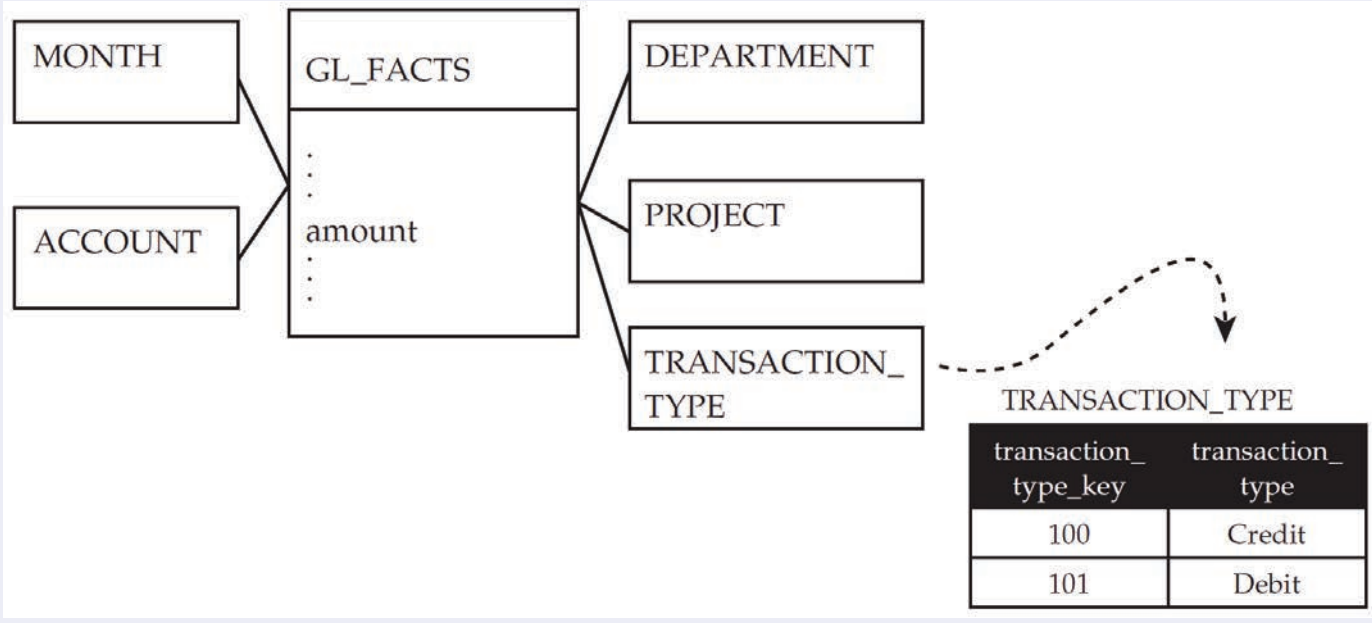
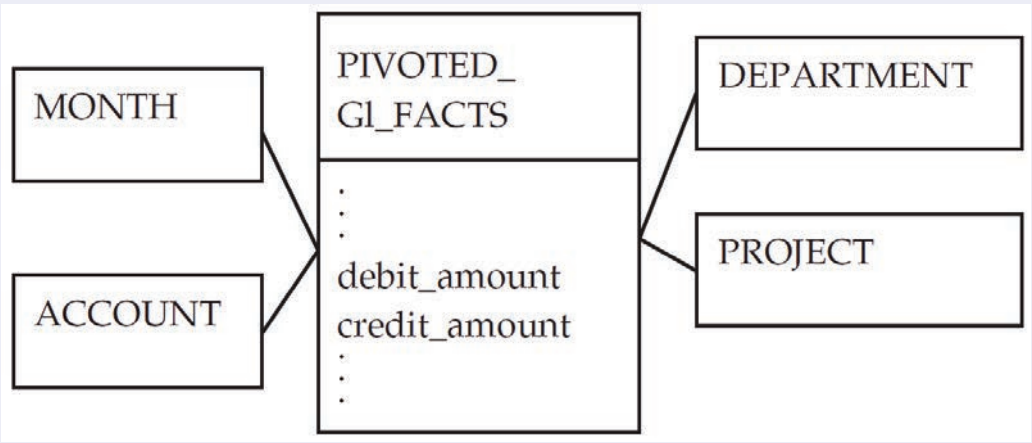


Tabla de hechos dinámica

Medición genérica



Mediciones para cada tipo

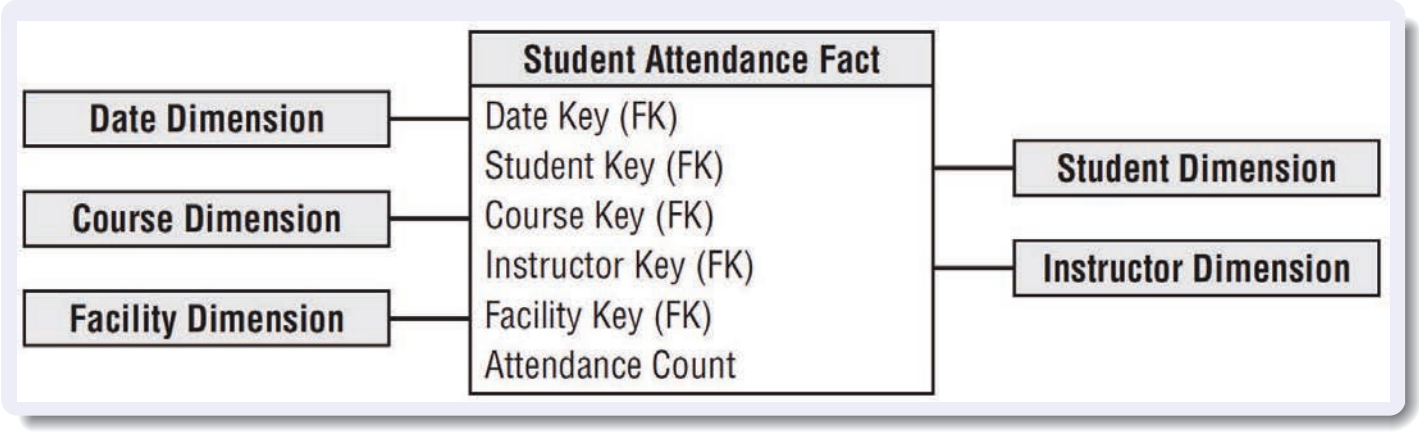


Ejemplos

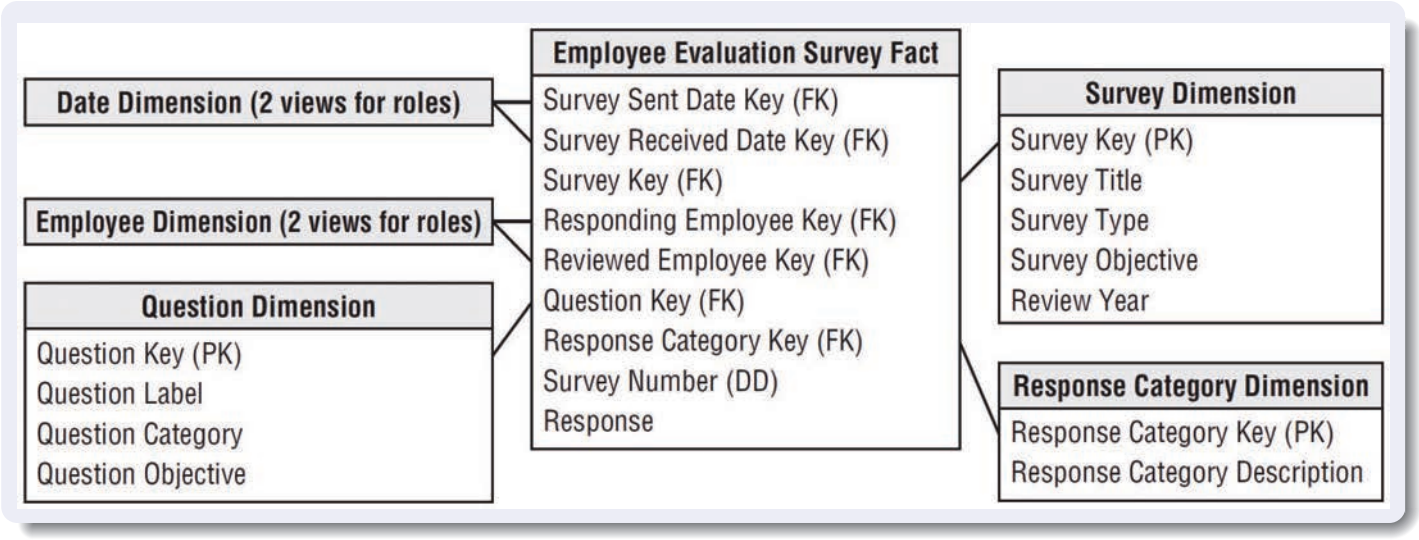
Ejemplos

- Resumen de ventas
- Ventas a nivel línea de ticket
- Asistencia a clase
- Encuestas de opinión
- ...

Asistencia



Encuestas



Bibliografía

Bibliografía

Ada10 C. Adamson: *Star Schema: The Complete Reference*. McGraw-Hill, 2010.

JPT10 C. Jensen, T. Pedersen, C. Thomsen: *Multidimensional Databases and Data Warehousing*. Morgan & Claypool, 2010.

KR13 R. Kimball, M. Ross: *The Data Warehouse Toolkit (3rd Edition)*. Wiley, 2013.

MZ08 E. Malinowski, E. Zimányi: *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications*. Springer (2008).