

Diseño Multidimensional

José Samos Jiménez

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

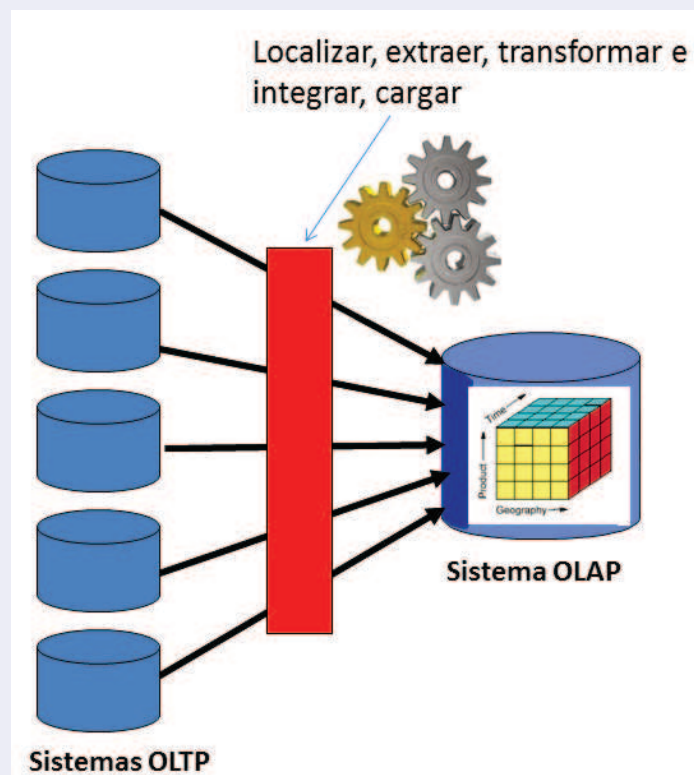
Curso 2019-20

Contenido

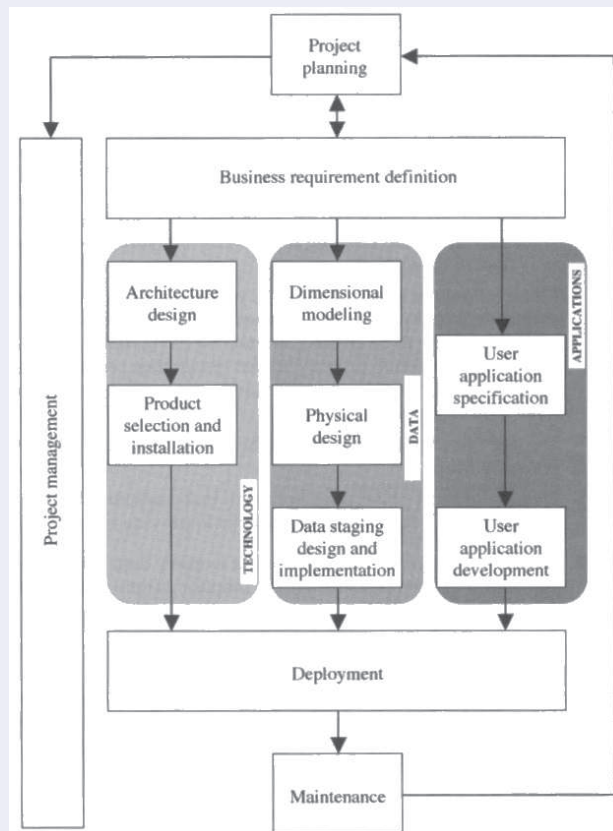
- 1 Ciclo de vida
- 2 Diseño conceptual
 - Fases de diseño
 - Ejemplo de esquemas
- 3 Diseño lógico
 - Regla general
 - Dimensiones *degeneradas*
 - Dimensiones *cajón de sastre*
 - SCD (dimensiones lentamente cambiantes)
 - Desdoblamiento de dimensiones
 - Dimensión de mediciones
- 4 Diseño físico
 - Partición de la tabla de hechos
 - Índices de mapa de bits y de join
- 5 Bibliografía

Ciclo de vida

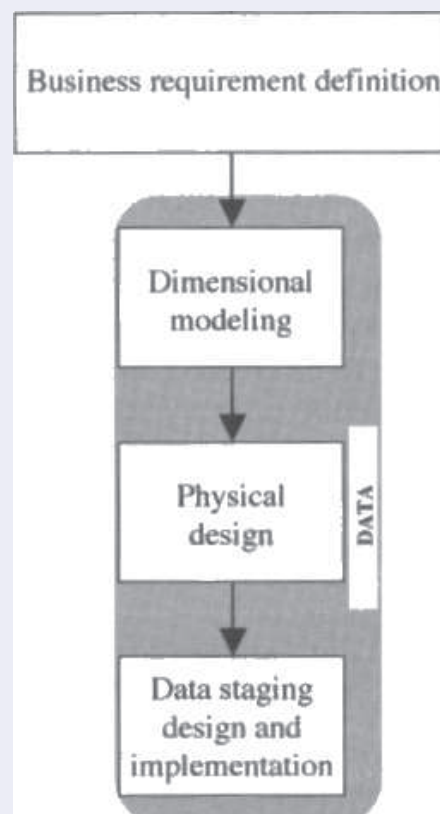
Componente ETL



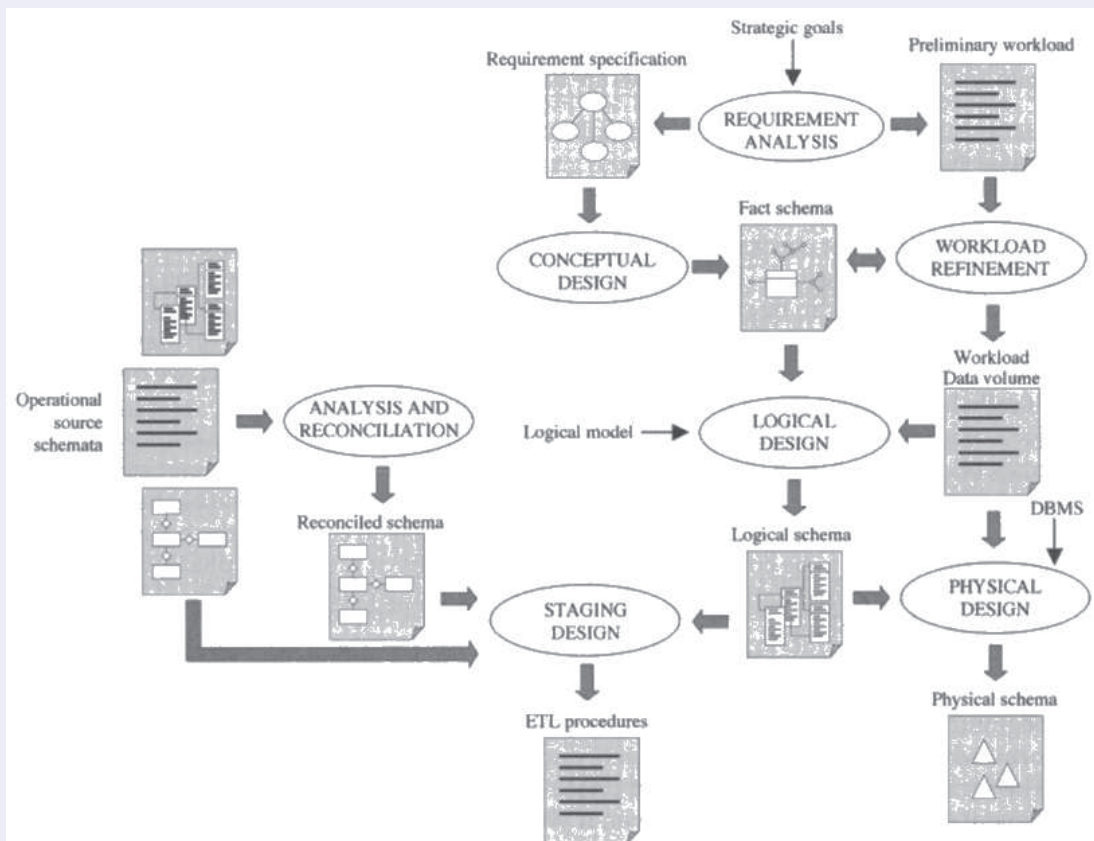
Ciclo de vida



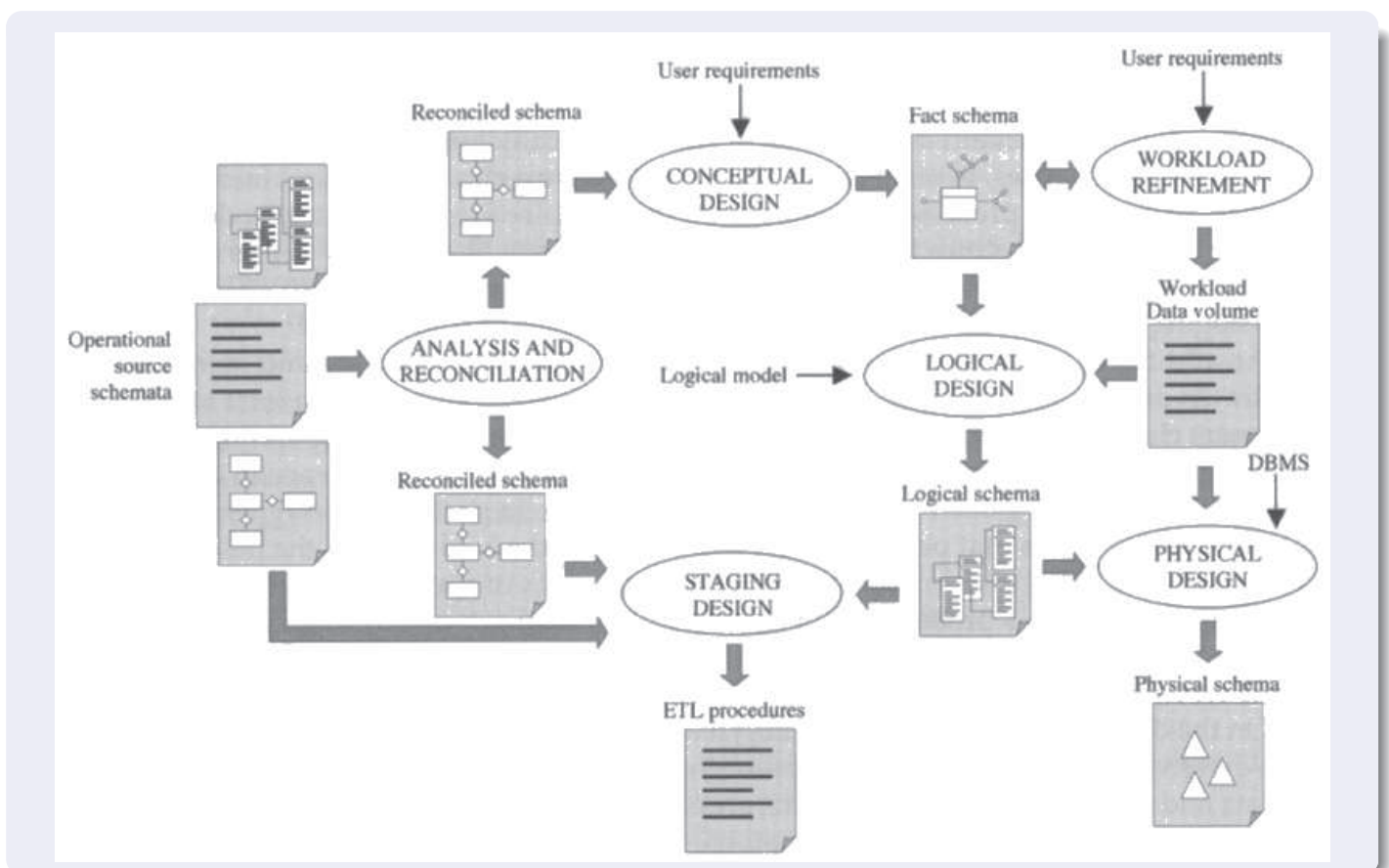
Ciclo de vida sobre los datos



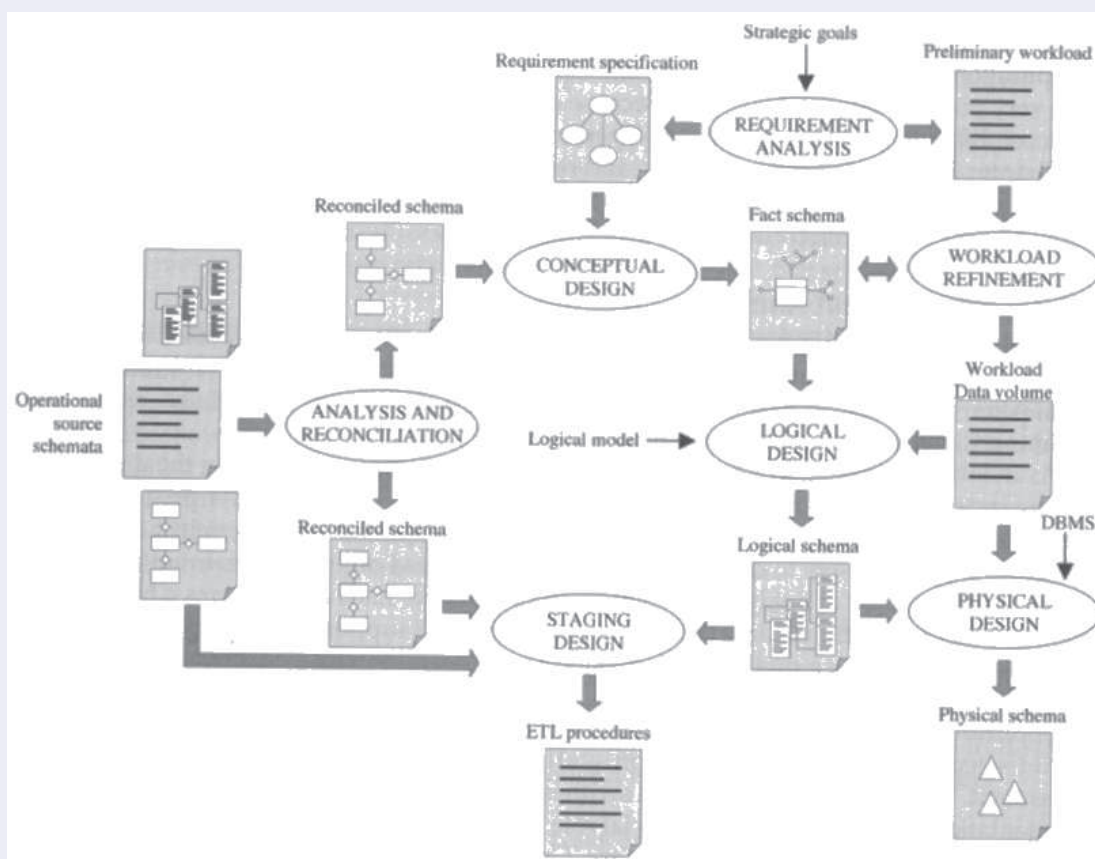
Enfoque dirigido por los requerimientos [GR09]



Enfoque dirigido por los datos [GR09]



Enfoque mixto [GR09]

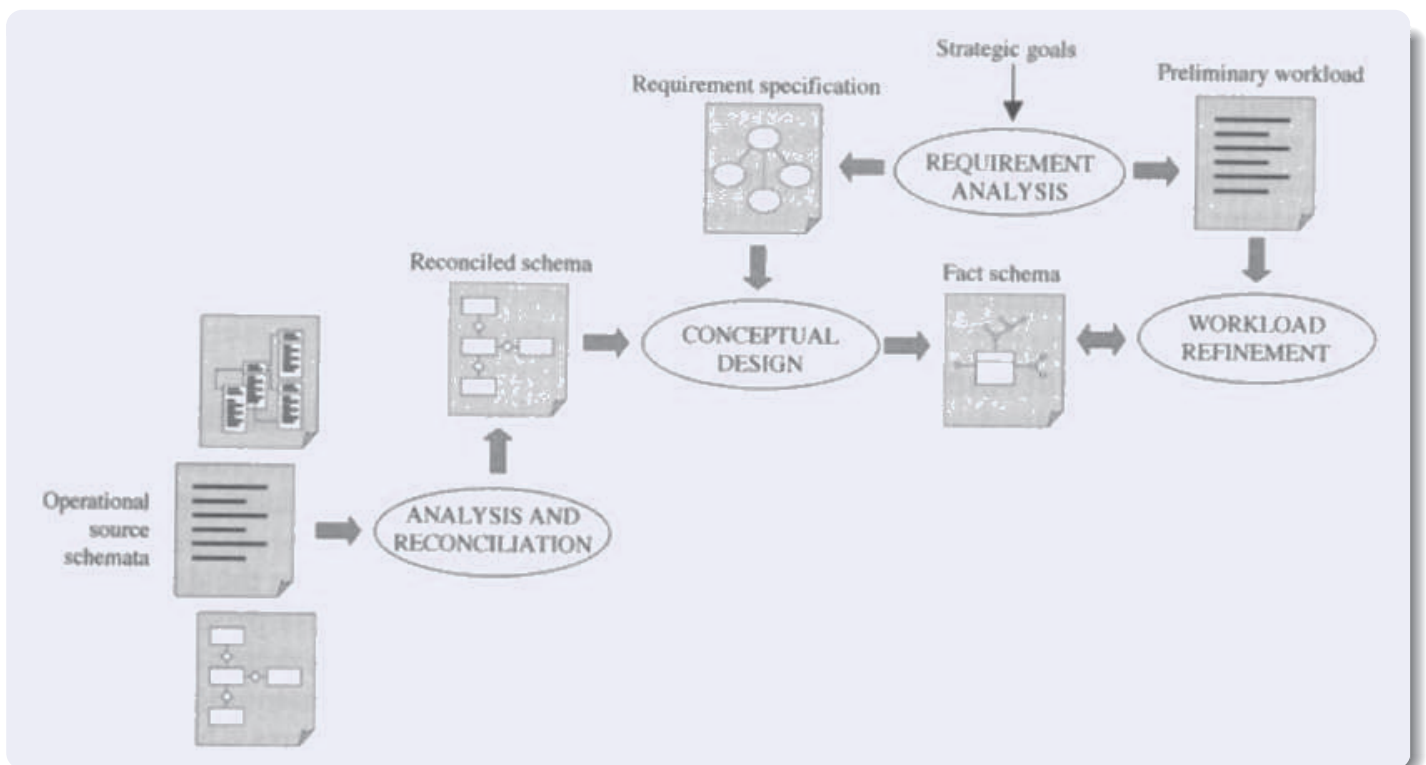


Descripción de las fases [GR09]

Phase	Input	Output	People Involved
Analysis and reconciliation of data sources	Operational source schemata	Reconciled schema	Designer, data processing center staff
Requirement analysis	Strategic goals	Requirement specifications, preliminary workload	Designer, end users
Conceptual design	Reconciled schema, requirement specification	Fact schemata	Designer, end users
Workload refinement, validation of conceptual schemata	Fact schemata, preliminary workload	Workload, data volume, validated fact schemata	Designer, end users
Logical design	Fact schemata, target logical model, workload	Logical data mart schema	Designer
Data-staging design	Source schemata, reconciled schema, logical data mart schema	ETL procedures	Designer, database administrators
Physical design	Logical data mart schema, target DBMS, workload	Physical data mart schema	Designer

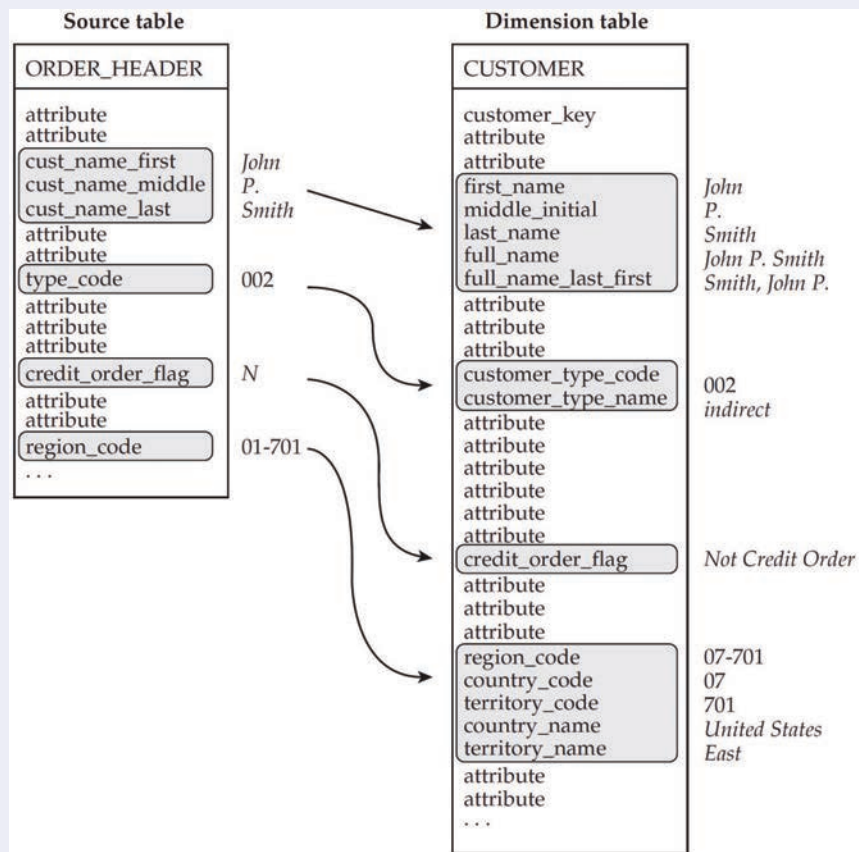
Diseño conceptual

Diseño conceptual



- Atención a la «calidad» de las instancias (no solo al esquema) OLTP.

Enriquecimiento de dimensiones



Fases de diseño

Fases

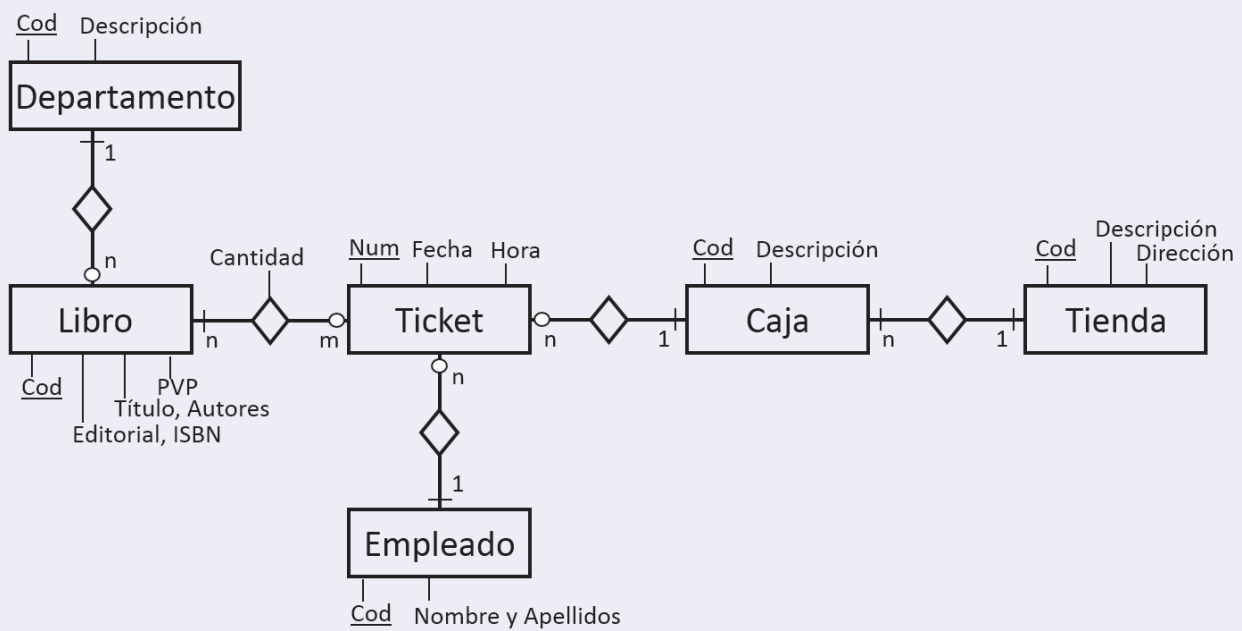
- Seleccionar el *proceso de negocio* a modelar.
 - ▶ Determinar el de mayor potencial de beneficios.
 - 1 Seleccionar la *granularidad* del proceso de negocio.
 - ▶ Determina el significado de los hechos: expresarlo.
 - ▶ Bases.
 - 2 Diseñar las *dimensiones*.
 - ▶ Niveles y jerarquías.
 - 3 Seleccionar las *mediciones* (determinadas por las dimensiones).
 - ▶ Aditividad.
 - ▶ Mediciones calculadas.
 - 4 Estimar el número de instancias (dimensiones y hechos).
- Si se dispone de otros diseños multidimensionales, *conformar* dimensiones y mediciones.

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ↺ ↻ ↻

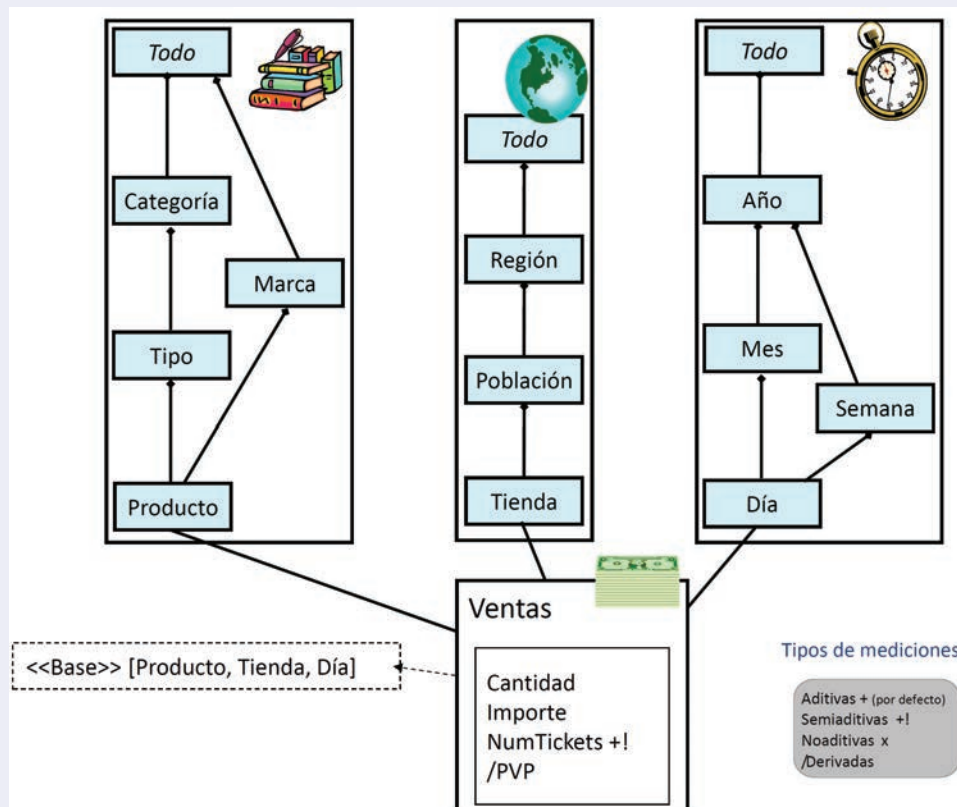
Ejemplo de esquemas

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ▶ ↺ ↻ ↻

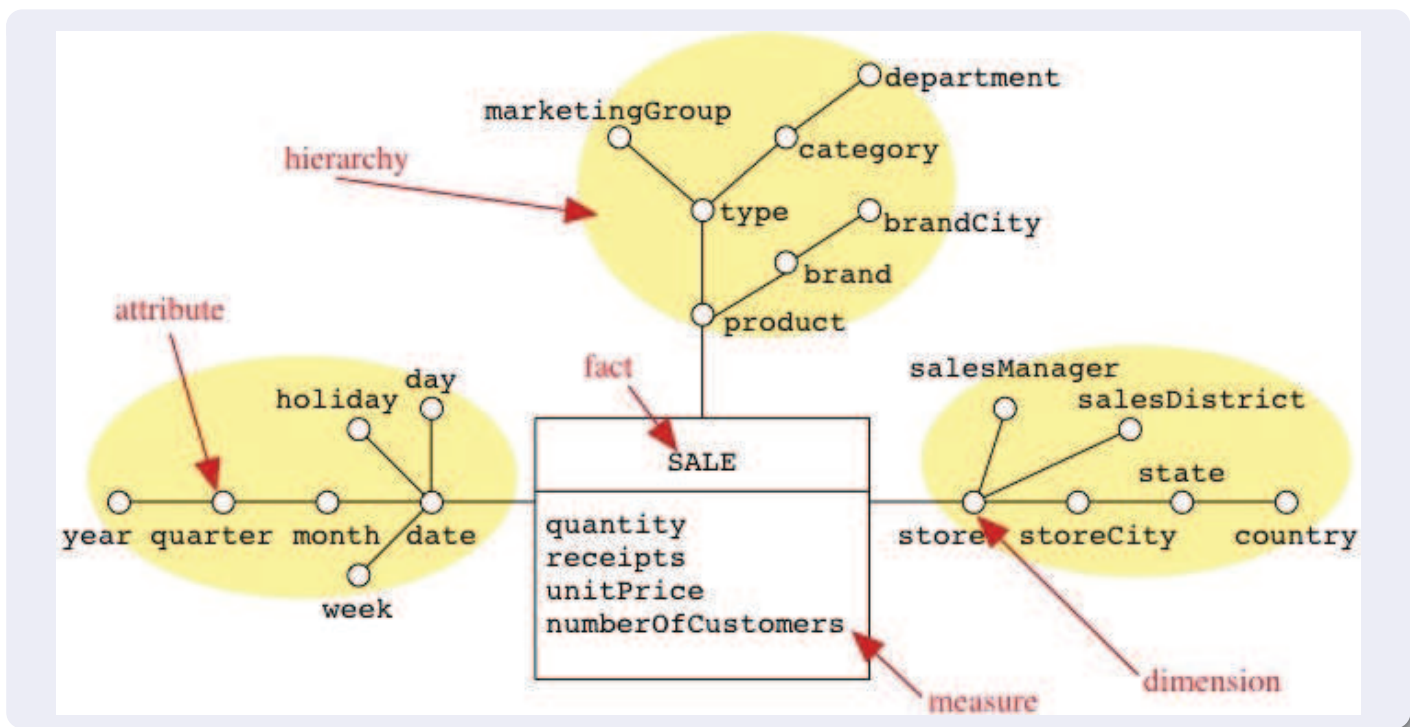
Sistema OLTP: esquema E/R



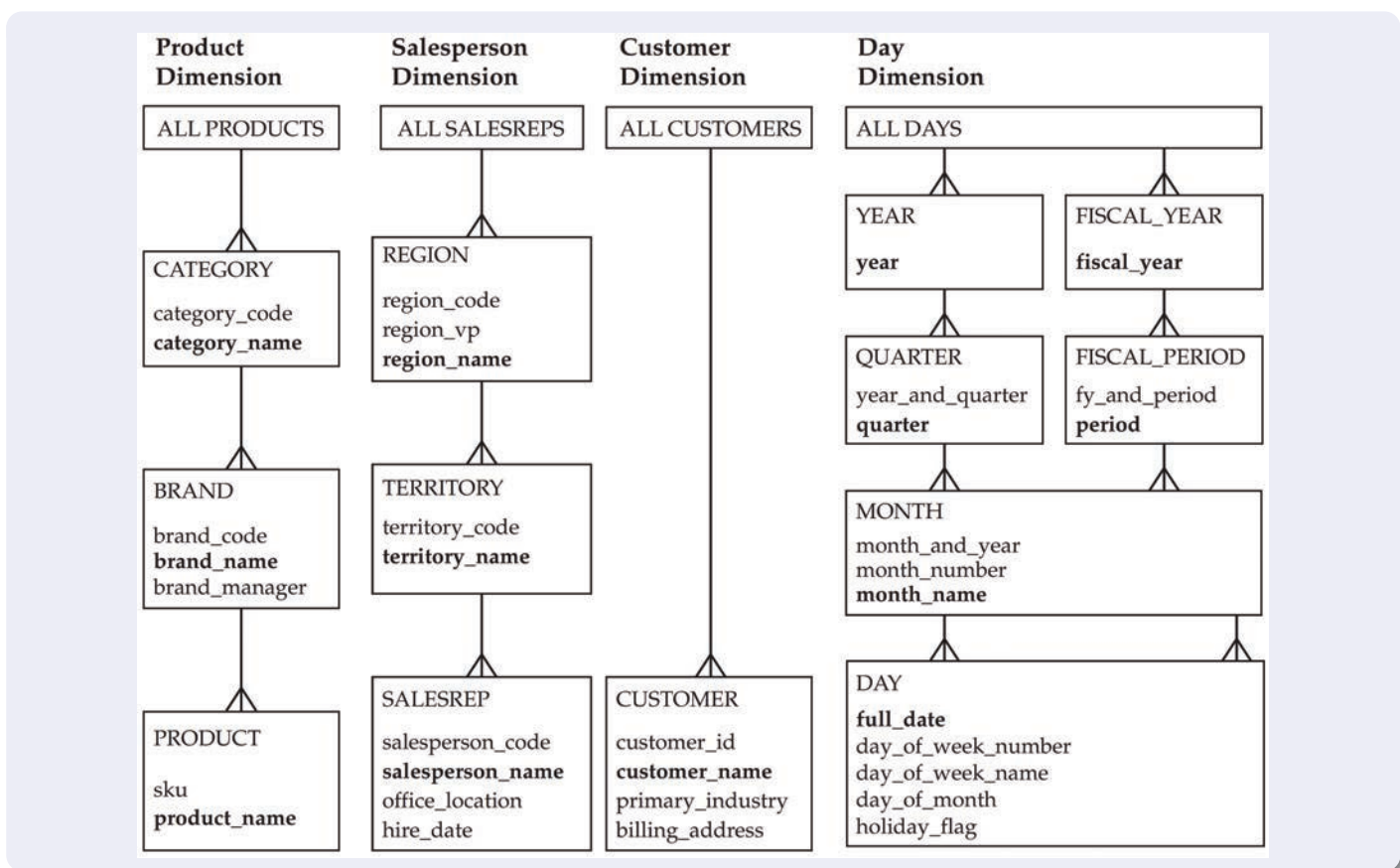
Esquema multidimensional



Representación de las dimensiones (i)

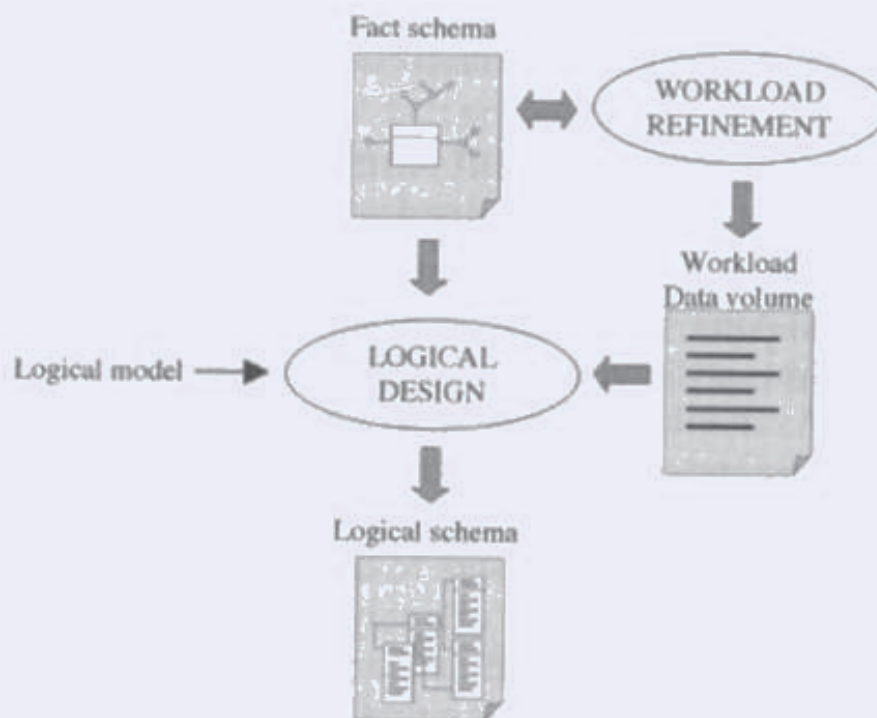


Representación de las dimensiones (y ii)



Diseño lógico

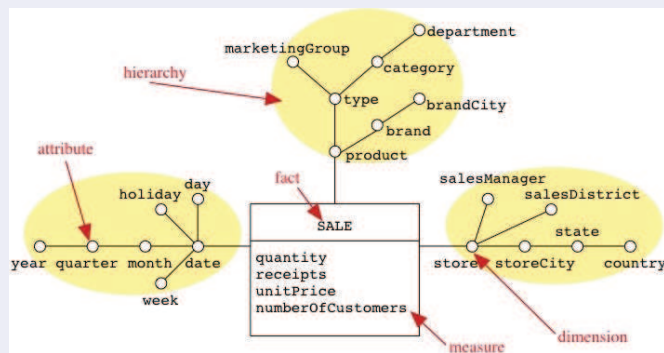
Diseño lógico



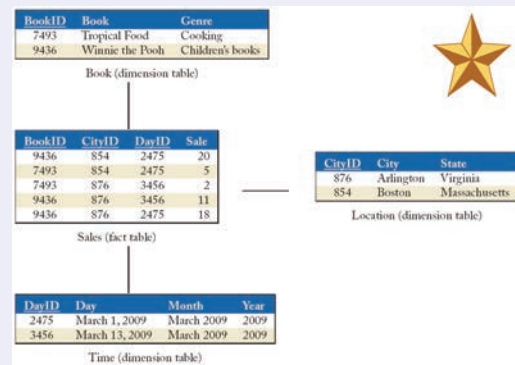
Regla general

Esquema conceptual y esquema lógico

Conceptual

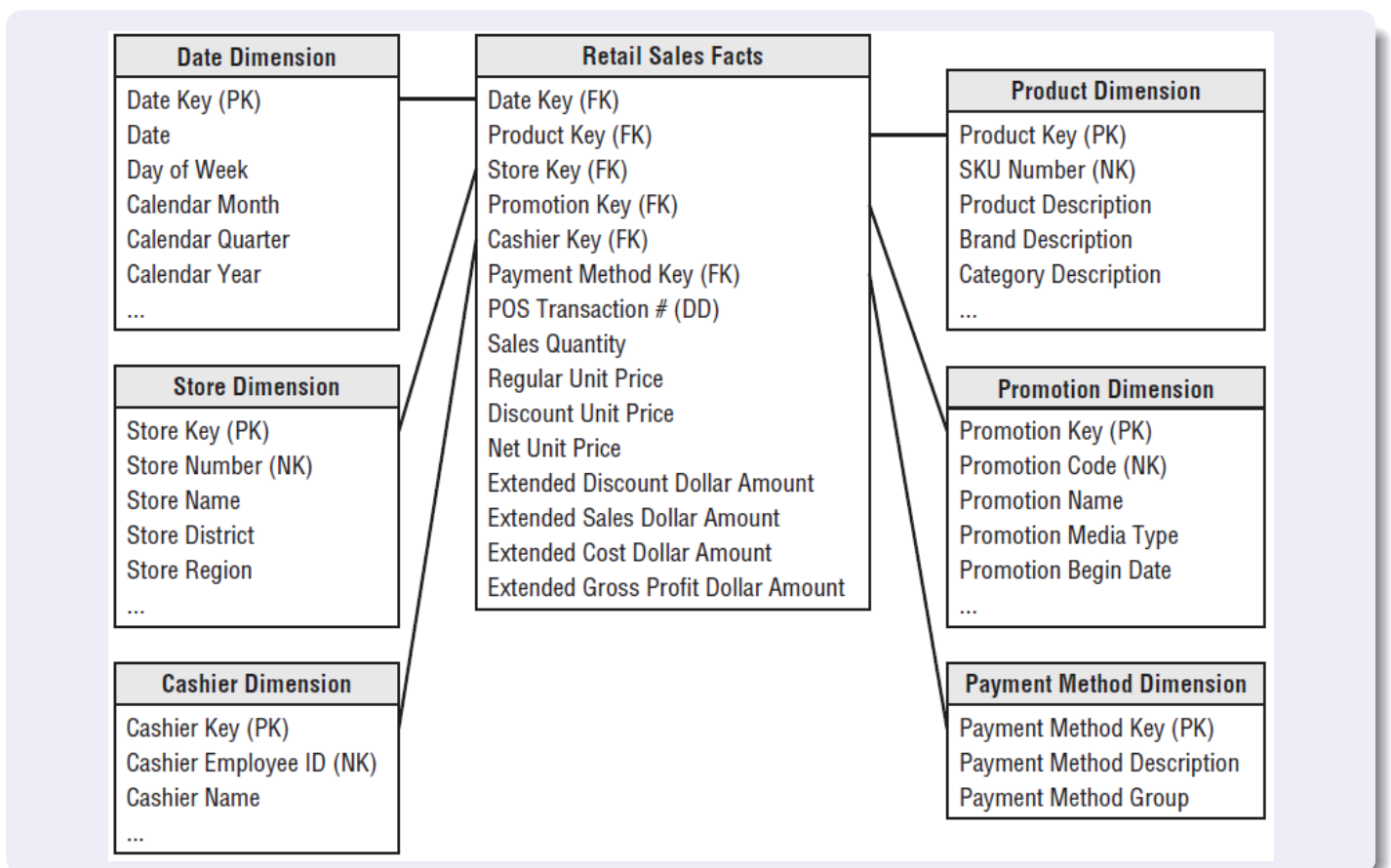


Lógico



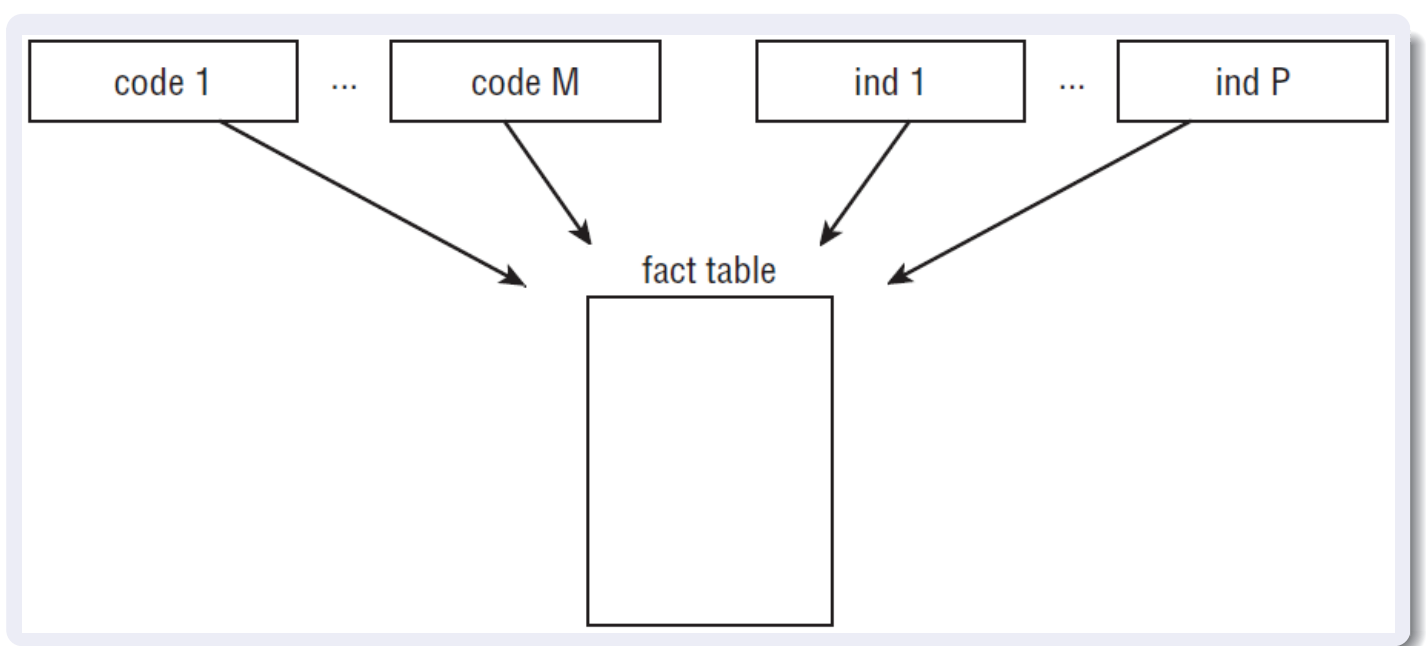
Dimensiones *degeneradas*

Dimensión degenerada

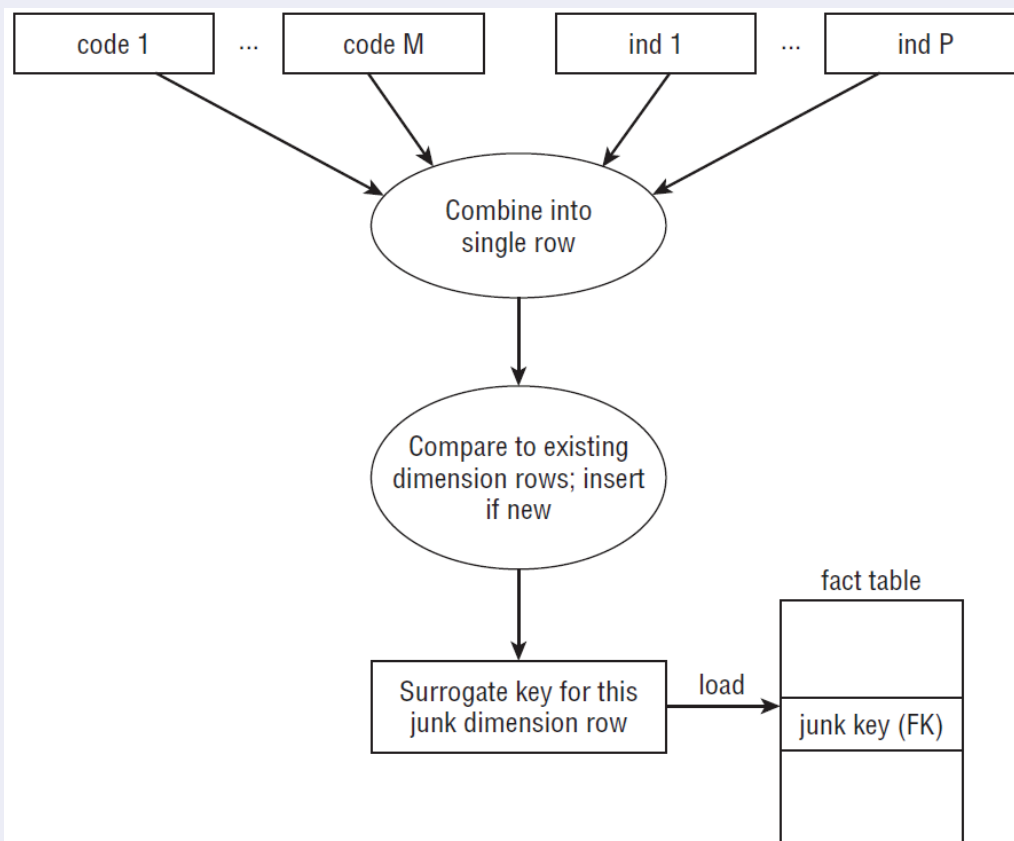


Dimensiones *cajón de sastre*

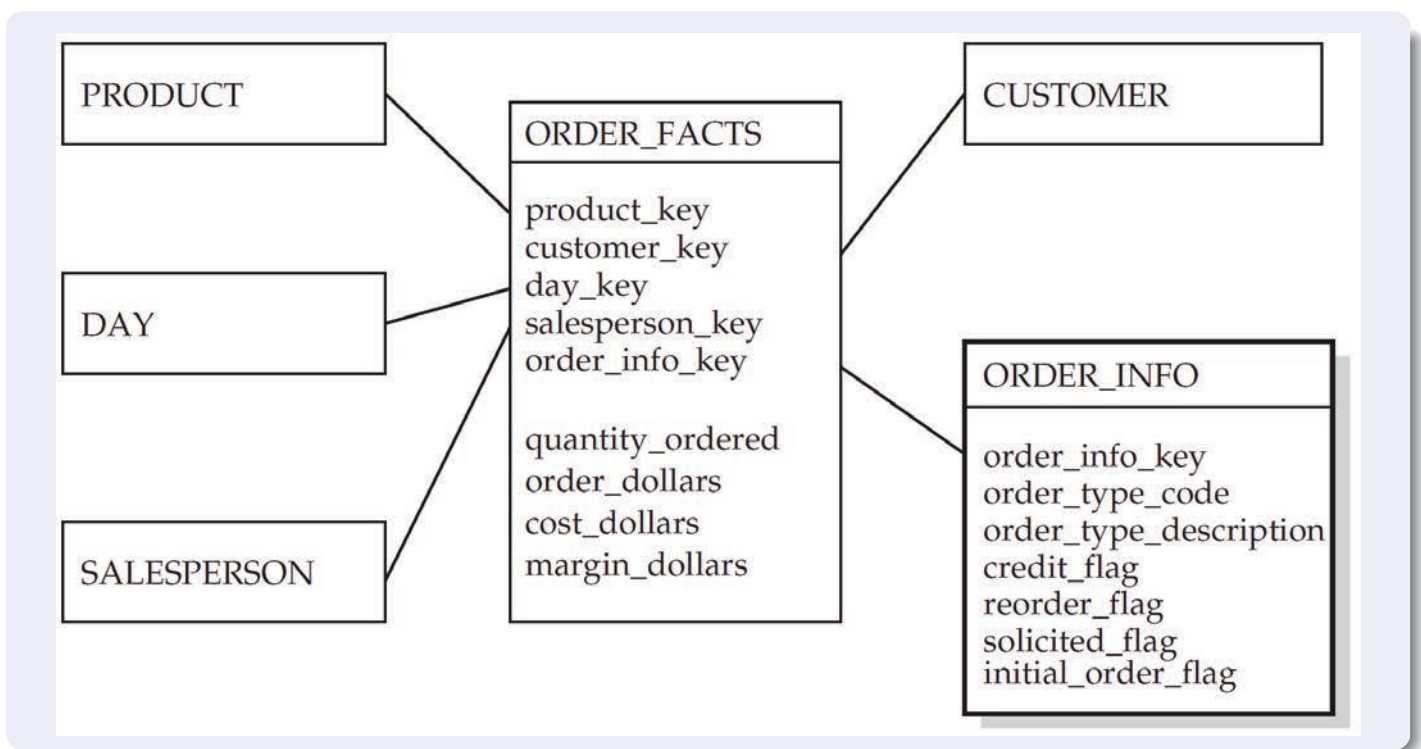
Flags como dimensiones



Proceso de generación de la dimensión



Dimensión cajón de sastre / de trastos / *junk dimension*



Ejemplo de instancias de una dimensión cajón de trastos

Order Indicator Key	Payment Type Description	Payment Type Group	Order Type	Commission Credit Indicator
1	Cash	Cash	Inbound	Commissionable
2	Cash	Cash	Inbound	Non-Commissionable
3	Cash	Cash	Outbound	Commissionable
4	Cash	Cash	Outbound	Non-Commissionable
5	Visa	Credit	Inbound	Commissionable
6	Visa	Credit	Inbound	Non-Commissionable
7	Visa	Credit	Outbound	Commissionable
8	Visa	Credit	Outbound	Non-Commissionable
9	MasterCard	Credit	Inbound	Commissionable
10	MasterCard	Credit	Inbound	Non-Commissionable
11	MasterCard	Credit	Outbound	Non-Commissionable
12	MasterCard	Credit	Outbound	Commissionable

SCD (dimensiones lentamente cambiantes)

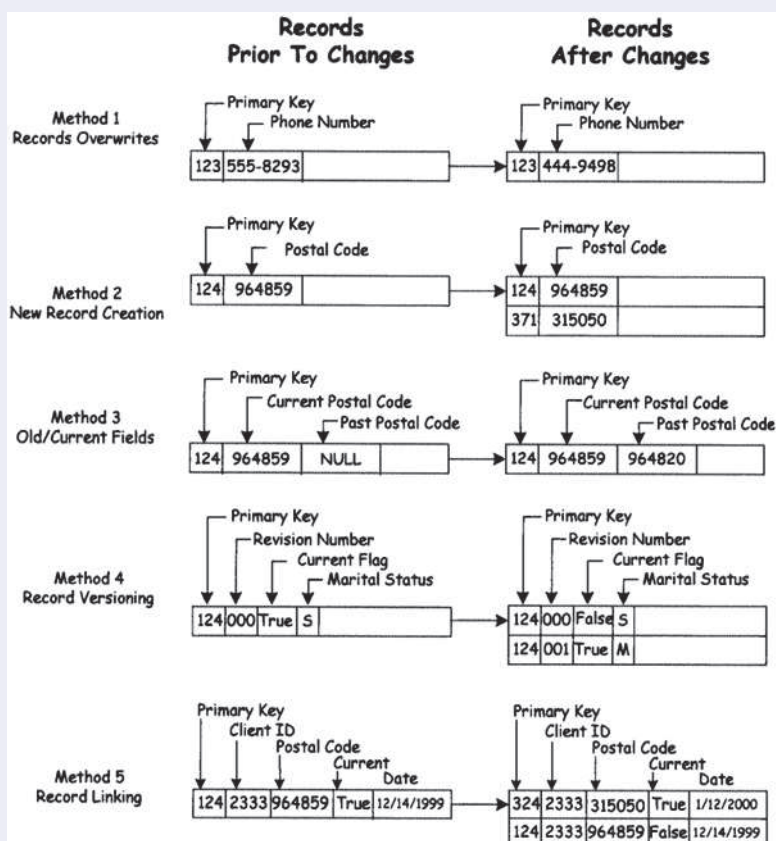
Estabilidad de las dimensiones

Empleado
<u>IdEmpleado</u>
DNI
Nombre y apellidos
Domicilio
Fecha de nacimiento
Sexo
Estado civil
Nº de hijos
Nivel de estudios
Categoría profesional
Antigüedad de categoría
Antigüedad en la empresa
Tipo de contrato
Sueldo
Nivel de Sueldo
Tramo de Edad
Población
Región

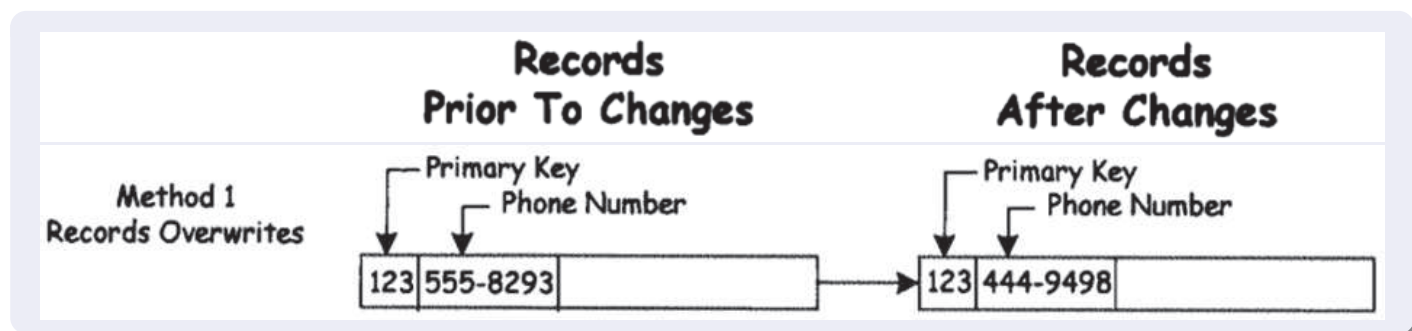
Métodos de solución al problema SCD

Dimensiones lentamente cambiantes

Empleado
<u>IdEmpleado</u>
DNI
Nombre y apellidos
Domicilio
Fecha de nacimiento
Sexo
Estado civil
Nº de hijos
Nivel de estudios
Categoría profesional
Antigüedad de categoría
Antigüedad en la empresa
Tipo de contrato
Sueldo
Nivel de Sueldo
Tramo de Edad
Población
Región



SCD M1: sobrecribir (*Tipo 1*)



SCD M2: nuevo registro (*Tipo 2*)



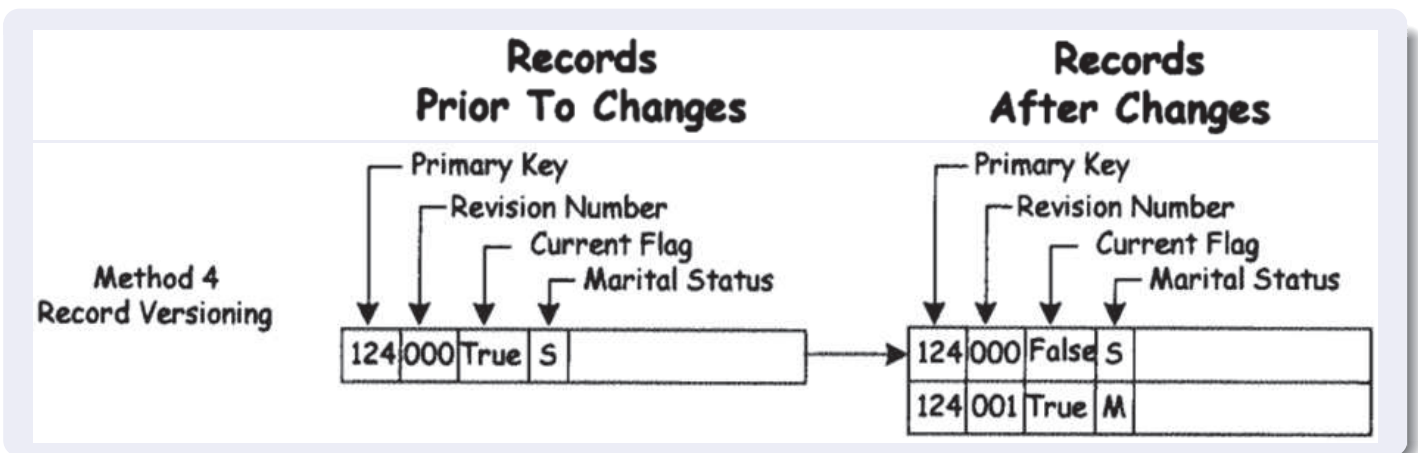
SCD M3: nuevo campo (*Tipo 3*)



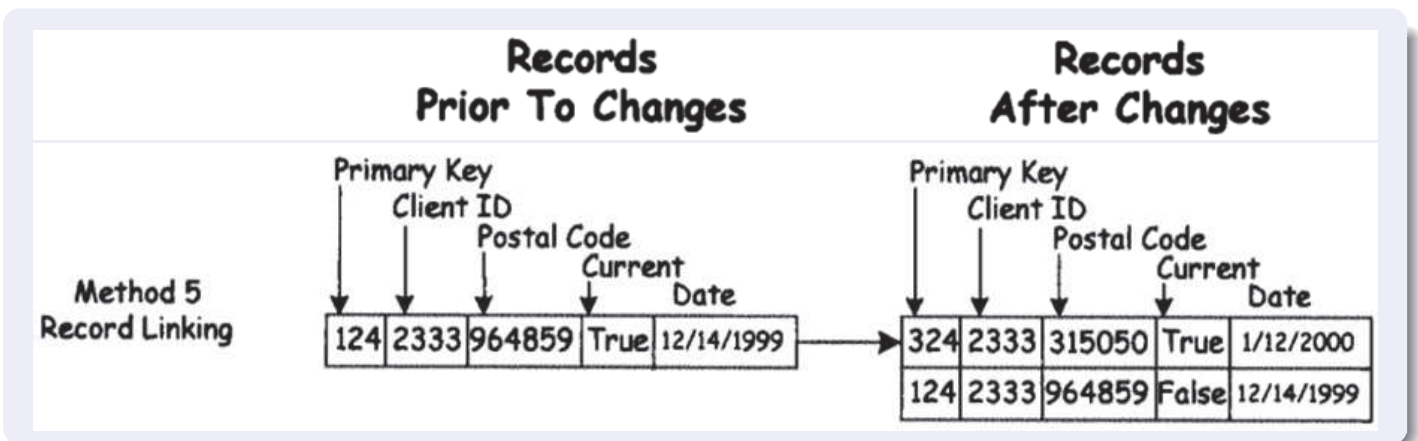
Ejemplo

- Productos en *Departamento*.

SCD M4: versiones de registros (*Tipo 2*)

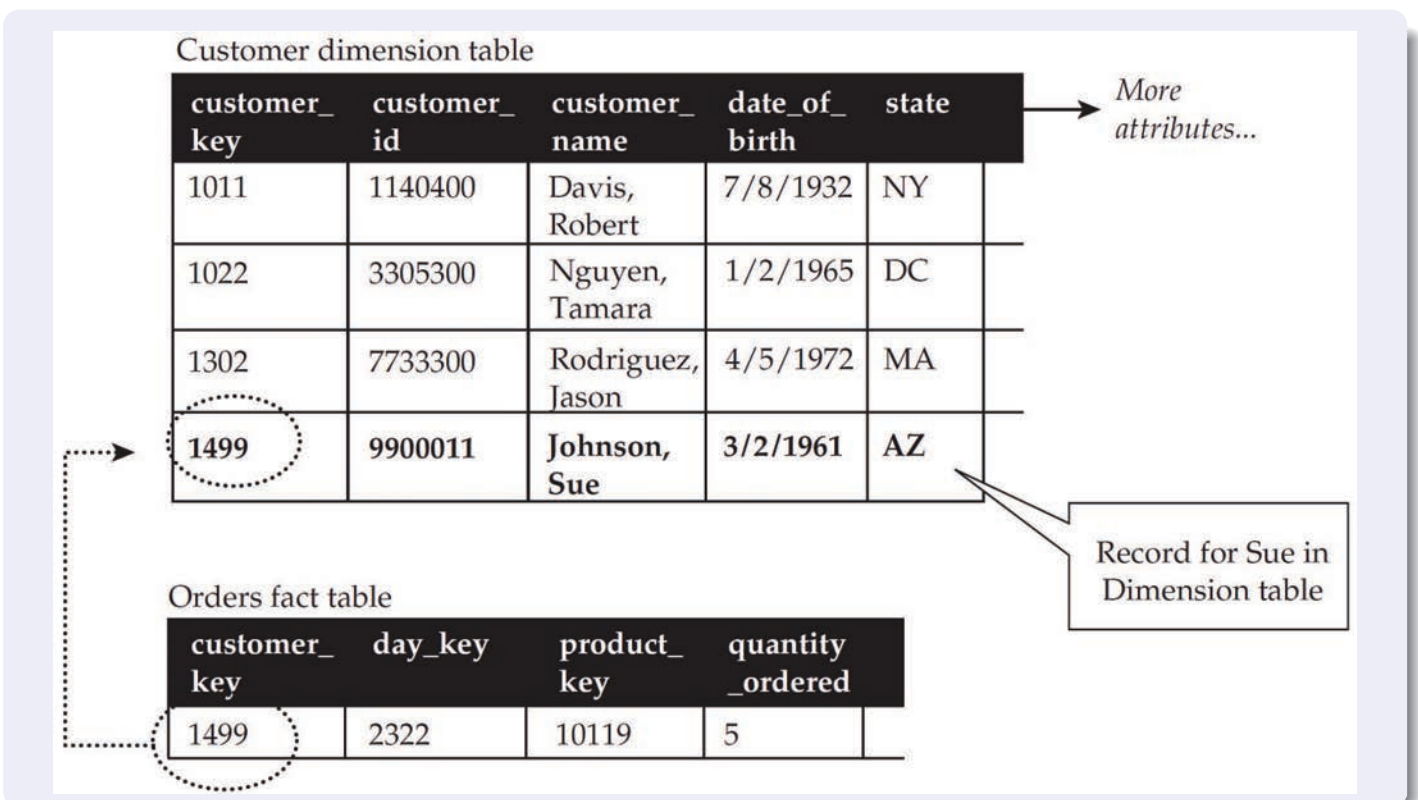


SCD M5: nuevo registro vinculado (Tipo 2)

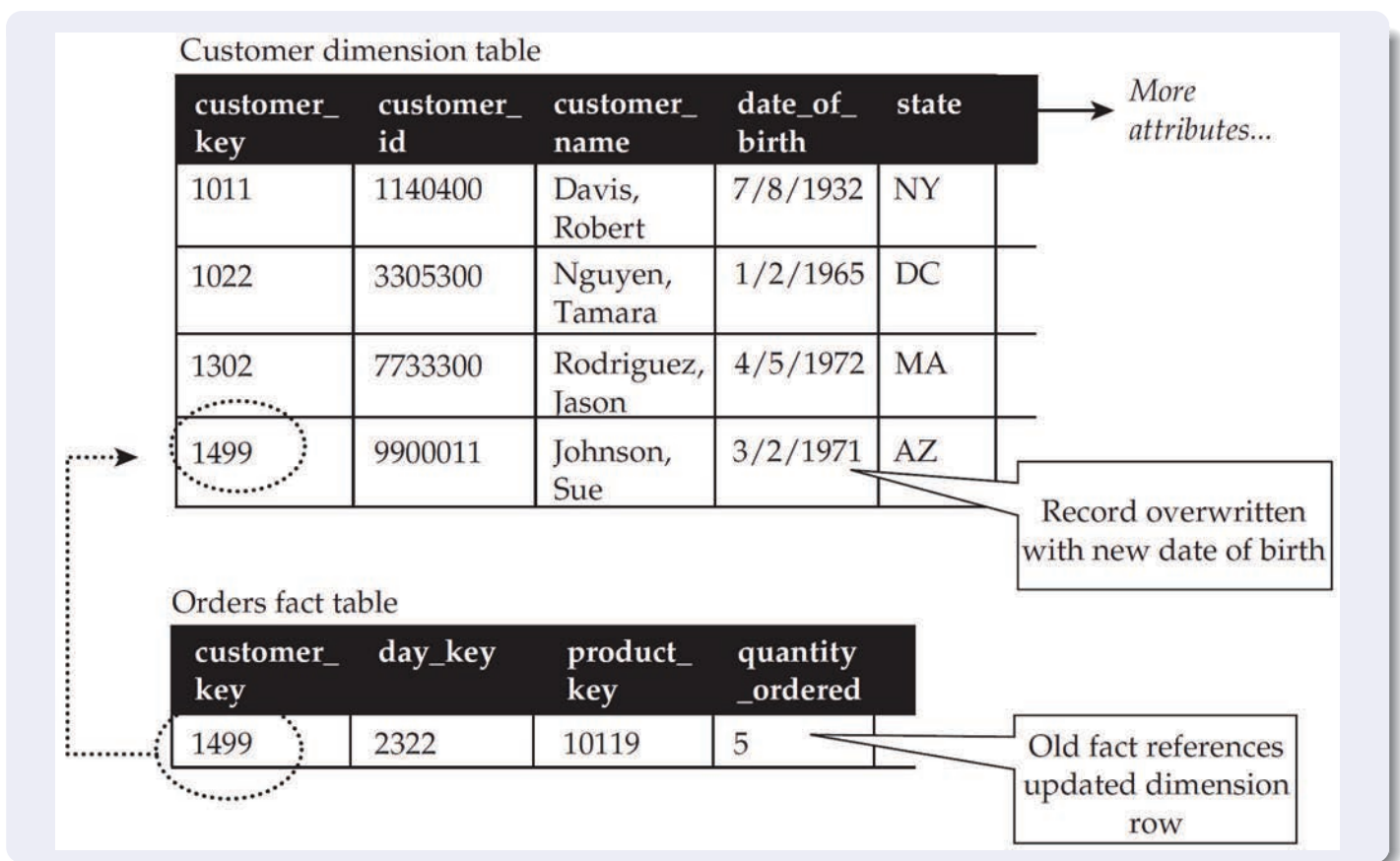


- Llaves generadas.

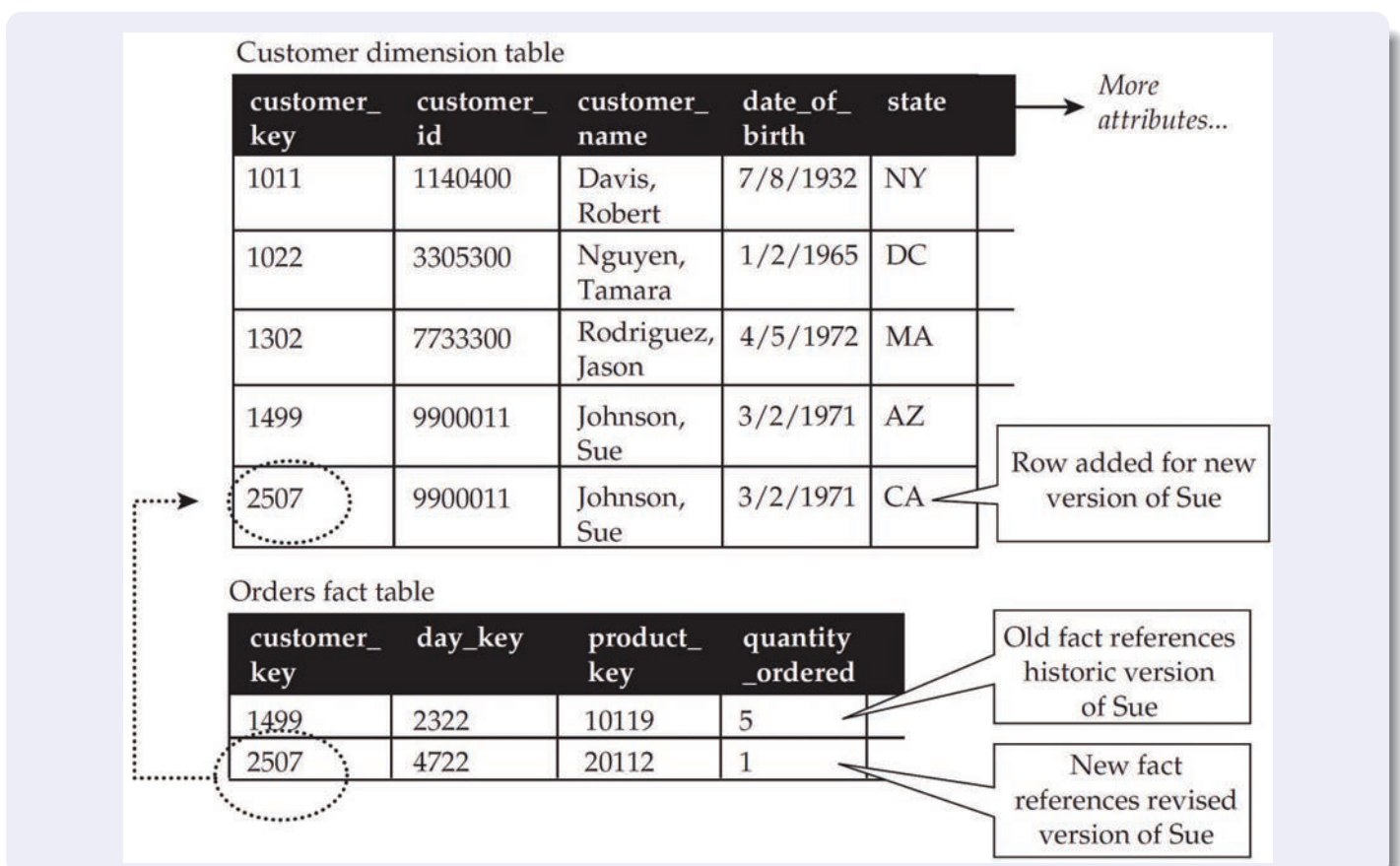
Ejemplo antes



Ejemplo después (i)



Ejemplo después (y ii)



Resumen de la versión simplificada del problema

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

Cómo documentar *SCD*

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

Ejemplo antes

Customer dimension table

customer_key	customer_ID	customer_name	region_current	region_previous
1011	1140400	Davis, Robert	East	East
1022	3305300	Nguyen, Tamara	East	East
1302	7733300	Rodriguez, Jason	West	West
1499	9900011	Johnson, Sue	West	West

Orders fact table

customer_key	day_key	order_dollars
1011	2322	2000
1011	3422	1000

Ejemplo después

Customer dimension table

customer_key	customer_ID	customer_name	region_current	region_previous
1011	1140400	Davis, Robert	Northeast	East
1022	3305300	Nguyen, Tamara	Southeast	East
1302	7733300	Rodriguez, Jason	West	West
1499	9900011	Johnson, Sue	West	West

Orders fact table

customer_key	day_key	order_dollars
1011	2322	2000
1011	3422	1000
1011	6599	1200
1011	8211	2000

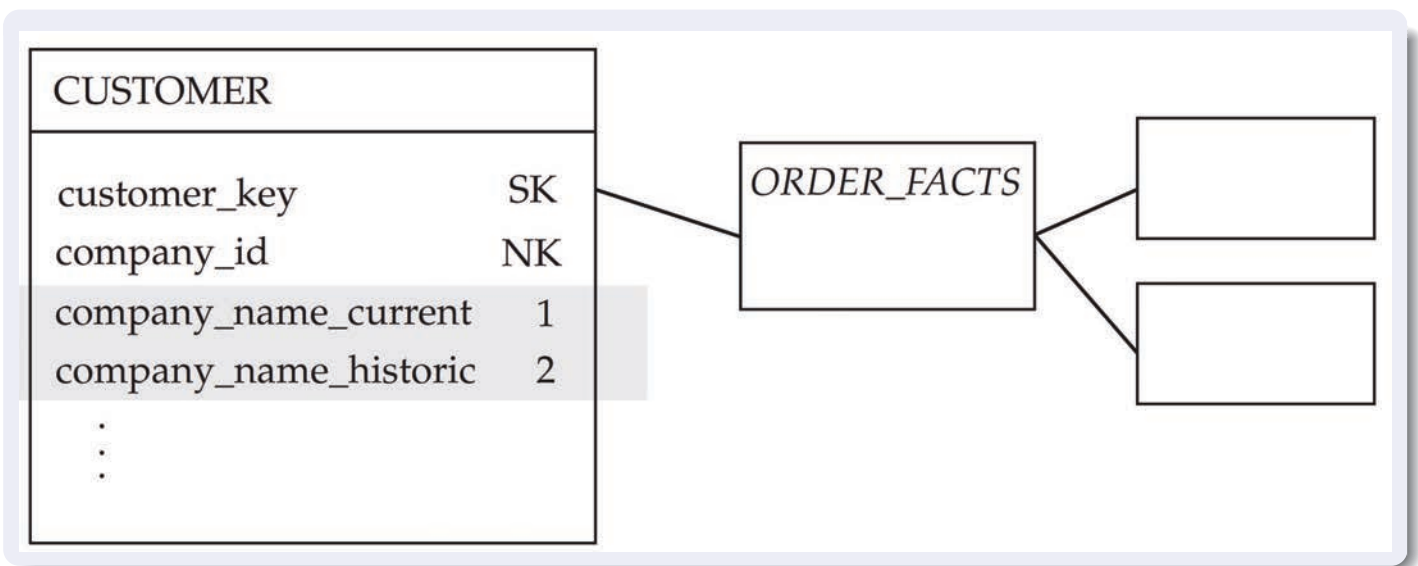
↑
Analysis
with
new values

↑
Analysis
with
old values

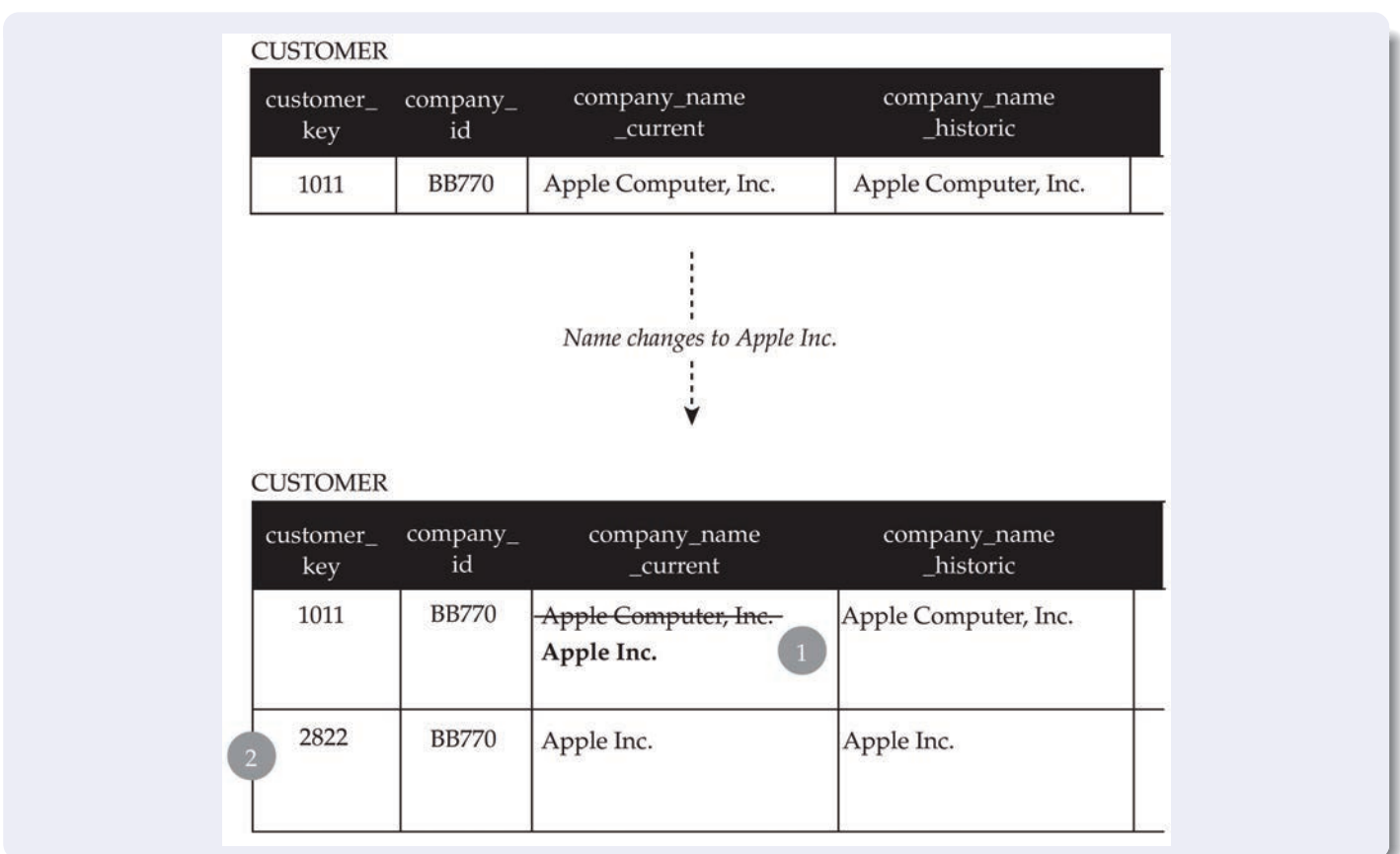
} facts in place before the change

} facts added after the change

¿Documentación del Tipo 3?



Tipo 3 + 2 + 1: Tipo 6 (i)



Tipo 3 + 2 + 1: Tipo 6 (y ii)

CUSTOMER

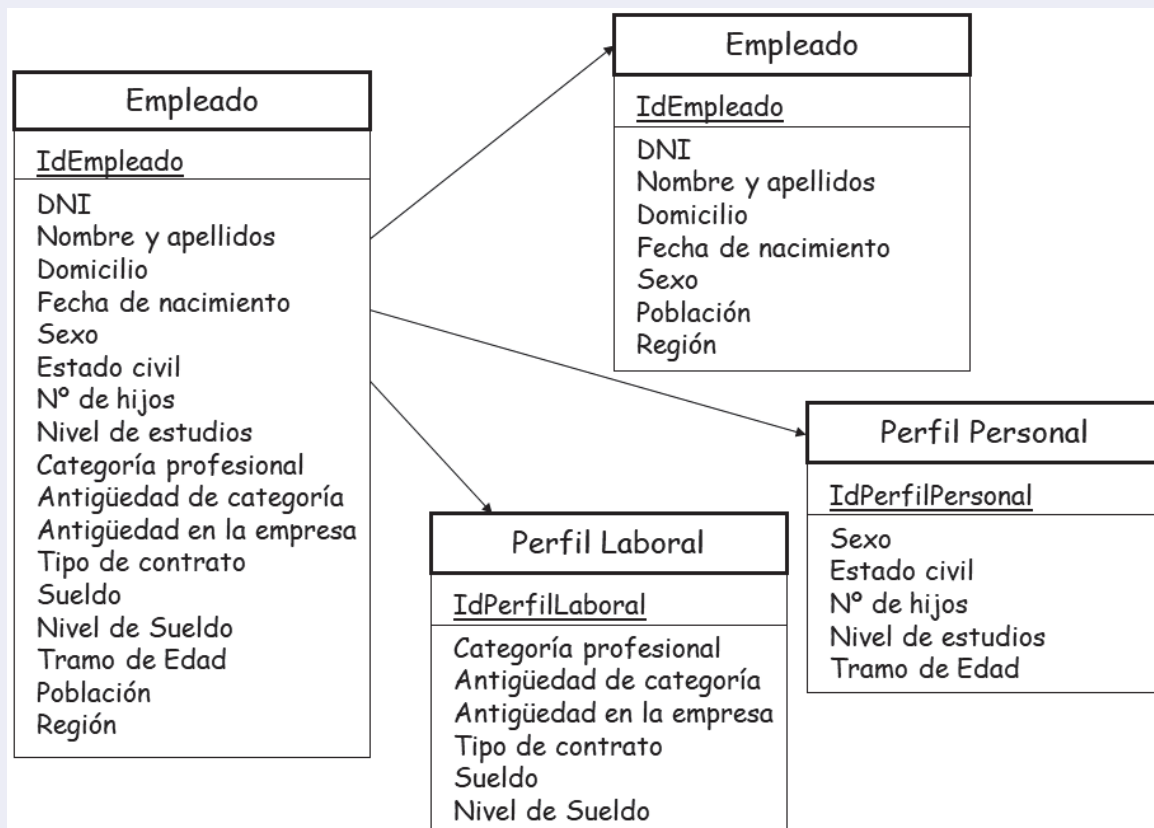
customer_ key	company_ id	company_name _current	company_name _historic	
1011	BB770	Apple Computer, Inc. Apple Inc. iApple Inc.	Apple Computer, Inc.	
2822	BB770	Apple Inc. iApple Inc.	Apple Inc.	
3100	BB770	iApple Inc.	iApple Inc.	

Resumen de *SCD* según Kimball

SCD Type	Dimension Table Action	Impact on Fact Analysis
Type 0	No change to attribute value.	Facts associated with attribute's original value.
Type 1	Overwrite attribute value.	Facts associated with attribute's current value.
Type 2	Add new dimension row for profile with new attribute value.	Facts associated with attribute value in effect when fact occurred.
Type 3	Add new column to preserve attribute's current and prior values.	Facts associated with both current and prior attribute alternative values.
Type 4	Add mini-dimension table containing rapidly changing attributes.	Facts associated with rapidly changing attributes in effect when fact occurred.
Type 5	Add type 4 mini-dimension, along with overwritten type 1 mini-dimension key in base dimension.	Facts associated with rapidly changing attributes in effect when fact occurred, plus current rapidly changing attribute values.
Type 6	Add type 1 overwritten attributes to type 2 dimension row, and overwrite all prior dimension rows.	Facts associated with attribute value in effect when fact occurred, plus current values.
Type 7	Add type 2 dimension row with new attribute value, plus view limited to current rows and/or attribute values.	Facts associated with attribute value in effect when fact occurred, plus current values.

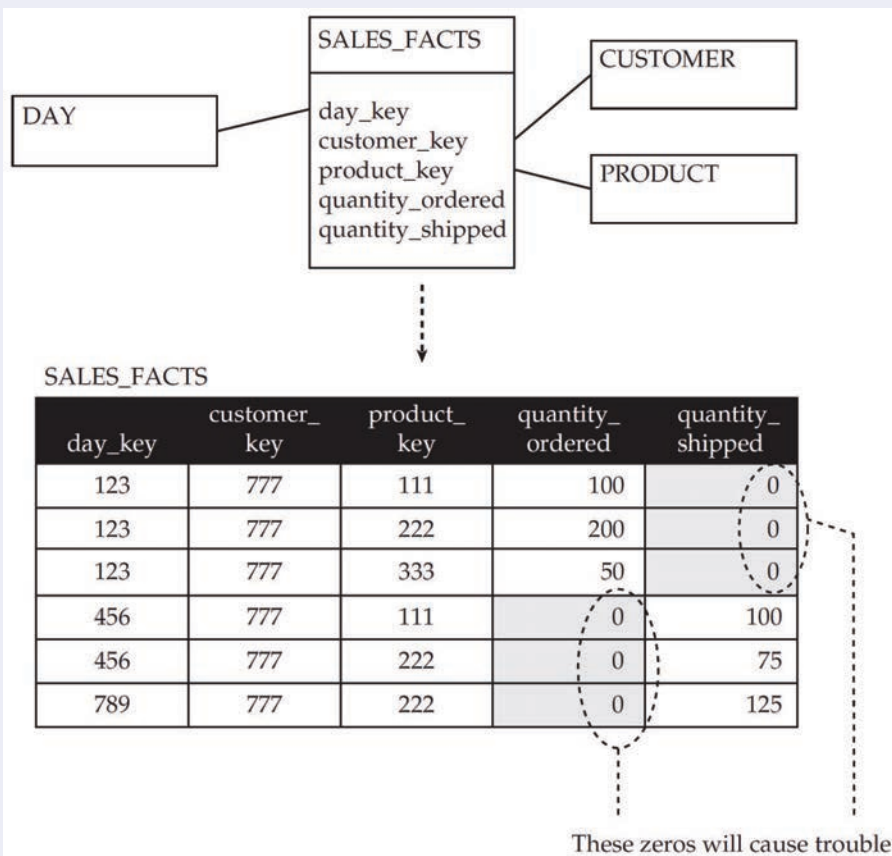
Desdoblamiento de dimensiones

Dimensiones cambiantes, no-lentamente-cambiantes

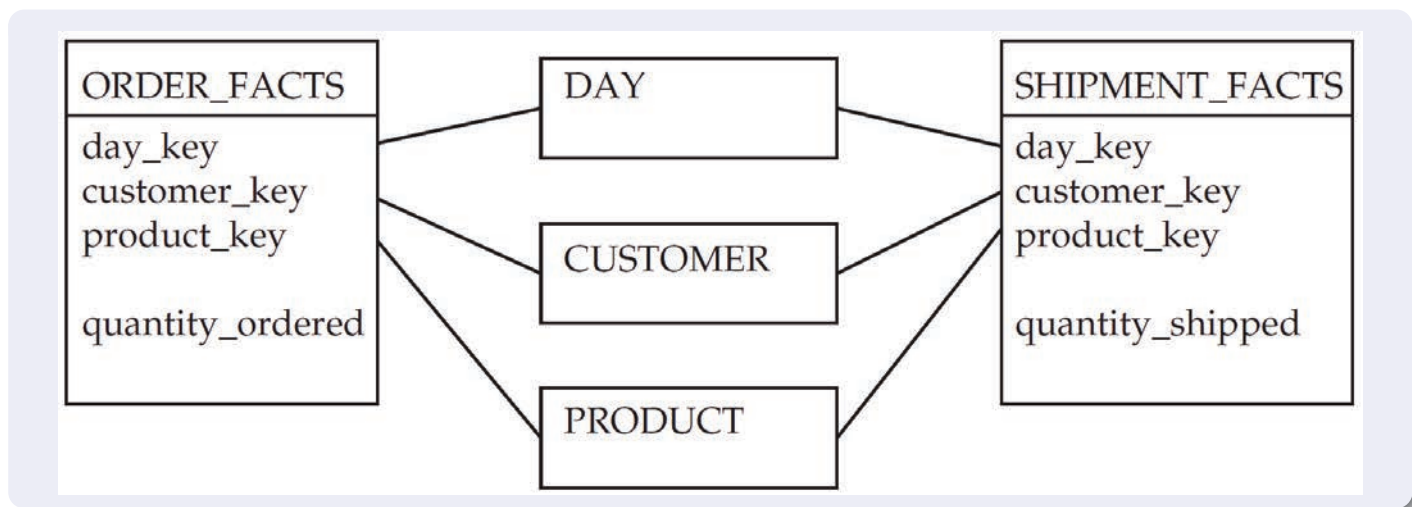


Dimensión de mediciones

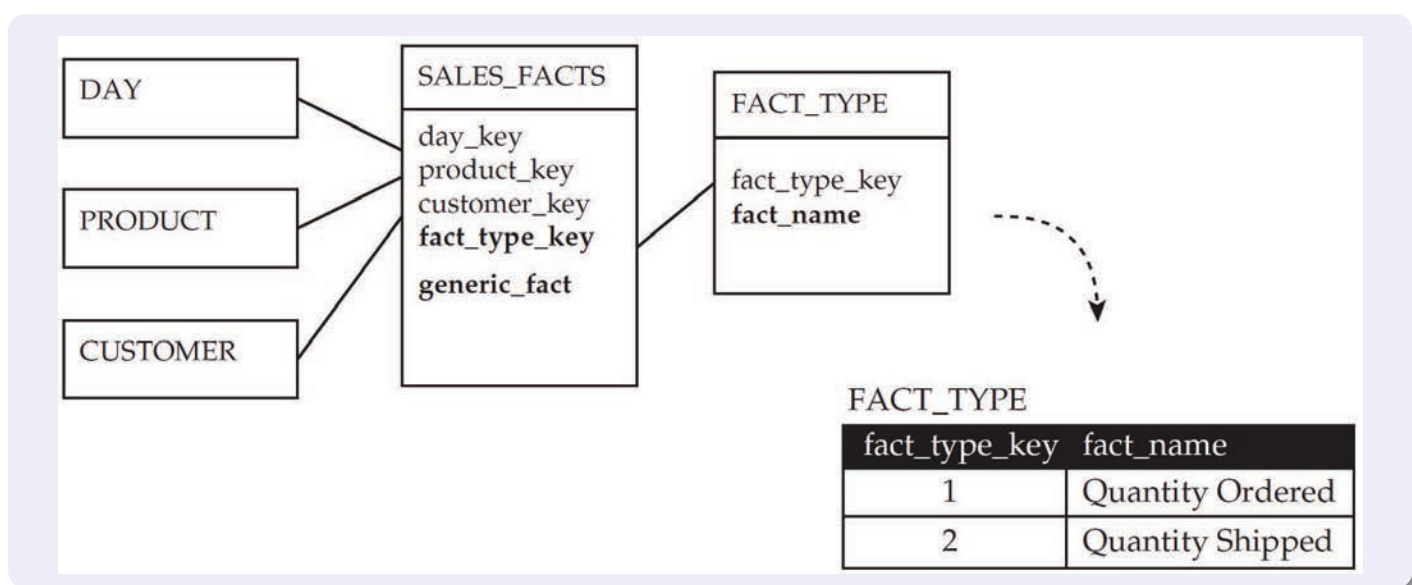
Mediciones no definidas



Dos focos de atención

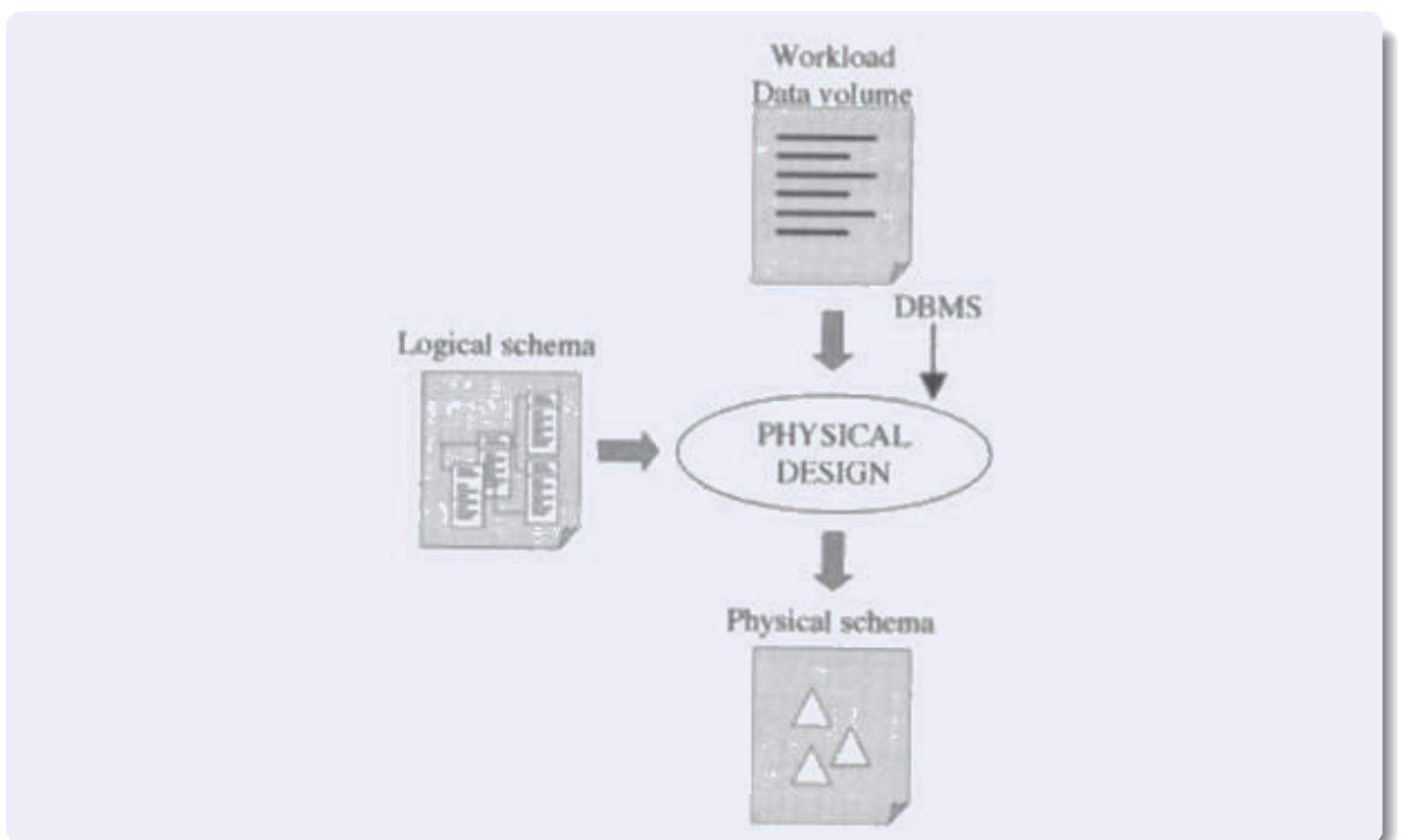


Dimensión de mediciones



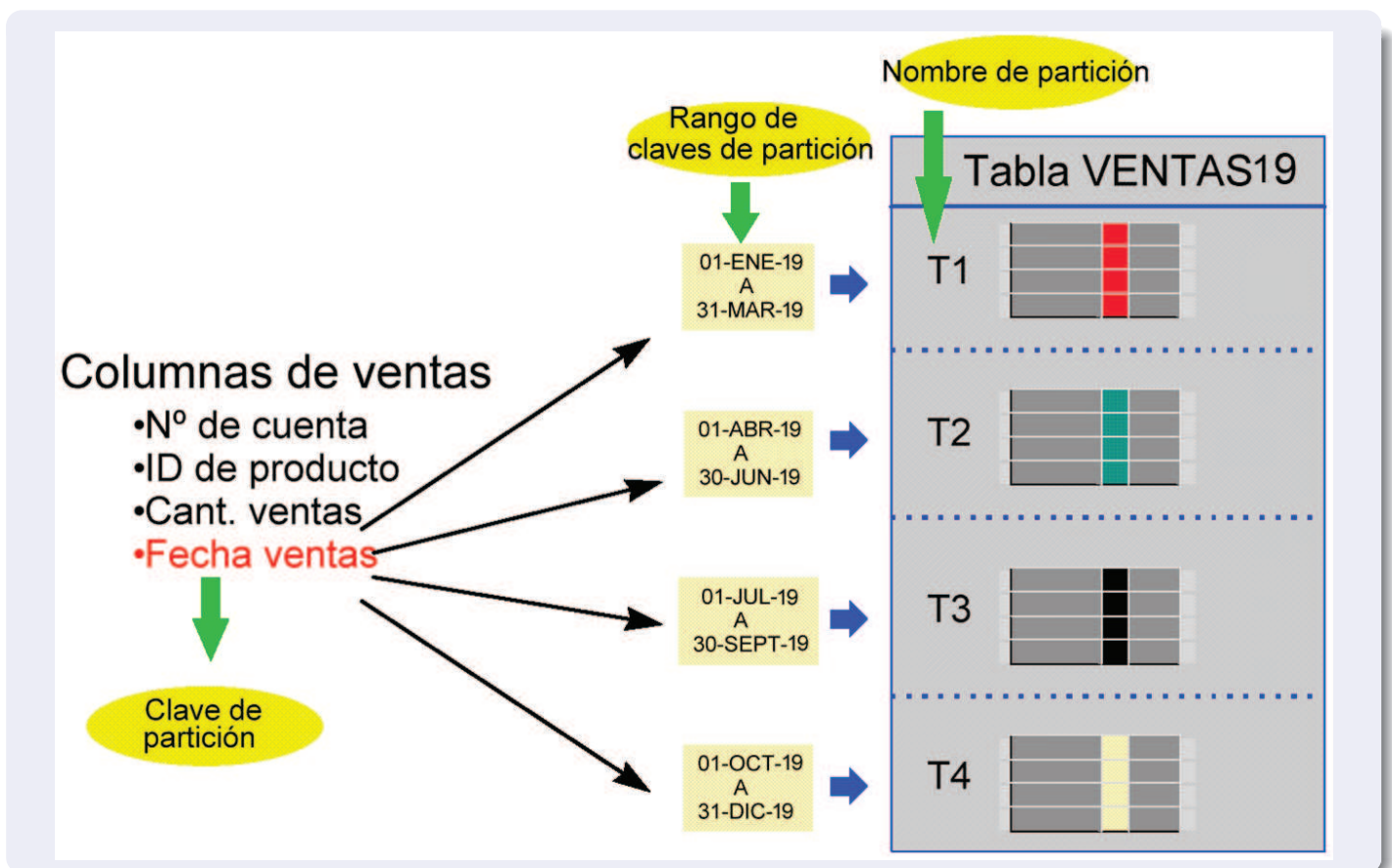
Diseño físico

Diseño físico

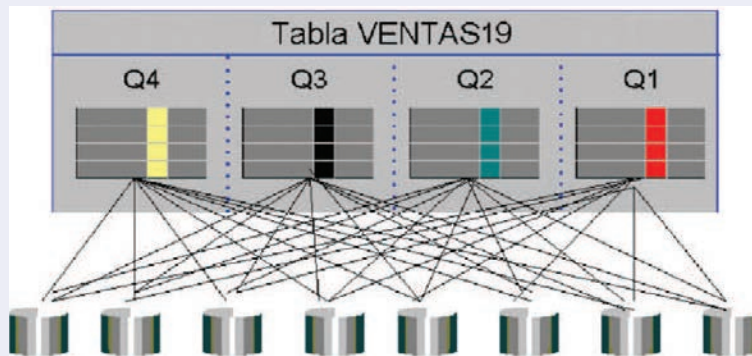
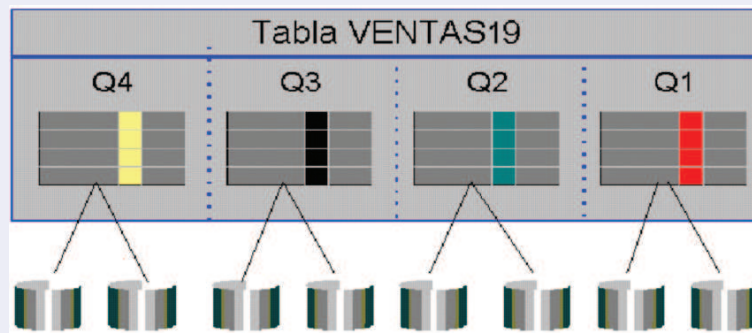


Partición de la tabla de hechos

Partición de la tabla de hechos



Asignación de las particiones a disco



- *RAID* (redundant array of inexpensive disk).

Índices de mapa de bits y de join

Índice de mapa de bits

Nombres					
Juan	María	José	Carmen	Luis	Andrés
1	0	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	1	0	0	0
0	0	0	0	1	0
0	1	0	0	0	0
0	0	0	0	0	1

Edades			
20	21	25	26
1	0	0	0
1	0	0	0
0	1	0	0
1	0	0	0
1	0	0	0
0	0	1	0
0	1	0	0
0	0	0	1

CREATE BITMAP INDEX Nombres
ON Clientes(Nombre)

CREATE BITMAP INDEX Edades
ON Clientes(Edad)

rId	Nombre	Edad
r4	Juan	20
r18	María	20
r19	José	21
r34	Carmen	20
r35	José	20
r36	Luis	25
r5	María	21
r41	Andrés	26

Mapa de bits vs. árbol B

• <https://www.oracle.com/technetwork/es/articles/sql/indices-mapa-de-bits-y-arbol-b-2393445-es.html>

Índice de join

RowID	BookID	Title	Genre	...
1	7493	Tropical Food	Cooking	...
2	9436	Winnie the Pooh	Childrens' books	...
3	9948	Gone With the Wind	Fiction	...
4	9967	Italian Food	Cooking	...

Book (dimension table)

RowID	BookID	ShopID	SalesID	DayID	Count	Price
1	9436	854	1021	2475	2	30
2	7493	854	1021	2475	1	20
3	9948	876	2098	3456	1	20
4	7493	876	2231	3456	2	40
5	7493	876	3049	2475	1	20
6	9436	854	3362	3569	2	30
7	9967	731	3460	3569	1	35
8	7493	731	3460	3569	1	15
9	9948	731	3460	3569	1	15

Sales (fact table)

Book_RowID	Sales_RowID
1	(2, 4, 5, 8)
2	(1, 6)
3	(3, 9)
4	(7)

Join index for Book and Sales

Índice de join mediante mapa de bits

<i>RowID</i>	<i>BookID</i>	<i>Title</i>	<i>Genre</i>	...
1	7493	Tropical Food	Cooking	...
2	9436	Winnie the Pooh	Childrens' books	...
3	9948	Gone With the Wind	Fiction	...
4	9967	Italian Food	Cooking	...

Book (dimension table)

<i>RowID</i>	<i>BookID</i>	<i>ShopID</i>	<i>SalesID</i>	<i>DayID</i>	<i>Count</i>	<i>Price</i>
1	9436	854	1021	2475	2	30
2	7493	854	1021	2475	1	20
3	9948	876	2098	3456	1	20
4	7493	876	2231	3456	2	40
5	7493	876	3049	2475	1	20
6	9436	854	3362	3569	2	30
7	9967	731	3460	3569	1	35
8	7493	731	3460	3569	1	15
9	9948	731	3460	3569	1	15

Sales (fact table)

<i>Book_RowID</i>	<i>Sales_positions</i>
1	010110010
2	100001000
3	001000001
4	000000100

Join index for Book and Sales

Bibliografía

Bibliografía

- Ada10 C. Adamson: *Star Schema: The Complete Reference*. McGraw-Hill, 2010.
- GR09 M. Golfarelli, S. Rizzi: *Data Warehouse Design: Modern Principles and Methodologies*. McGraw-Hill, 2009.
- 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.