

Integración de Sistemas

José Samos Jiménez

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

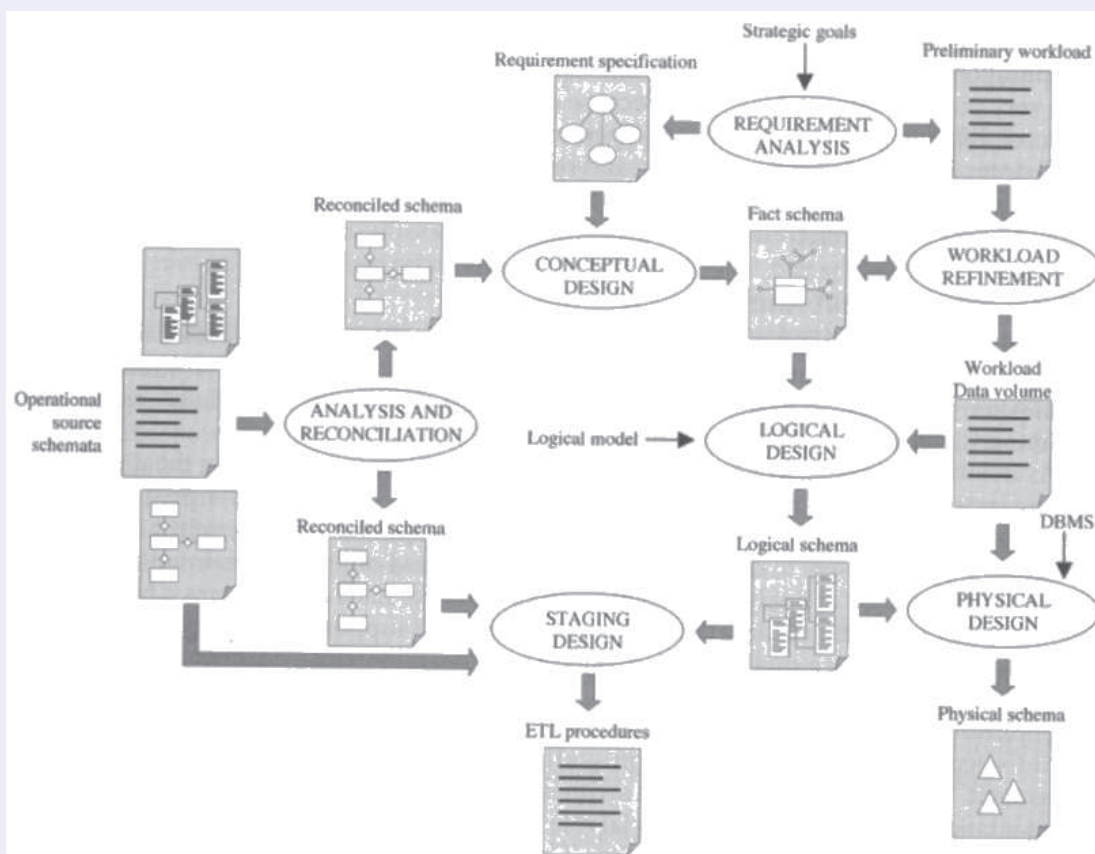
Curso 2019-20

Contenido

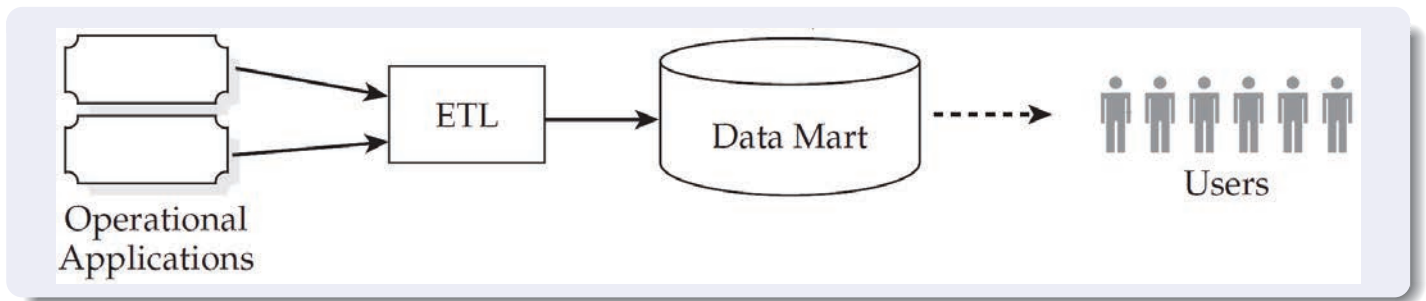
- 1 Desarrollo de sistemas multidimensionales
- 2 Arquitecturas de integración de sistemas
- 3 Metadatos
- 4 Implementación de la arquitectura
- 5 Errores comunes
- 6 Bibliografía

Desarrollo de sistemas multidimensionales

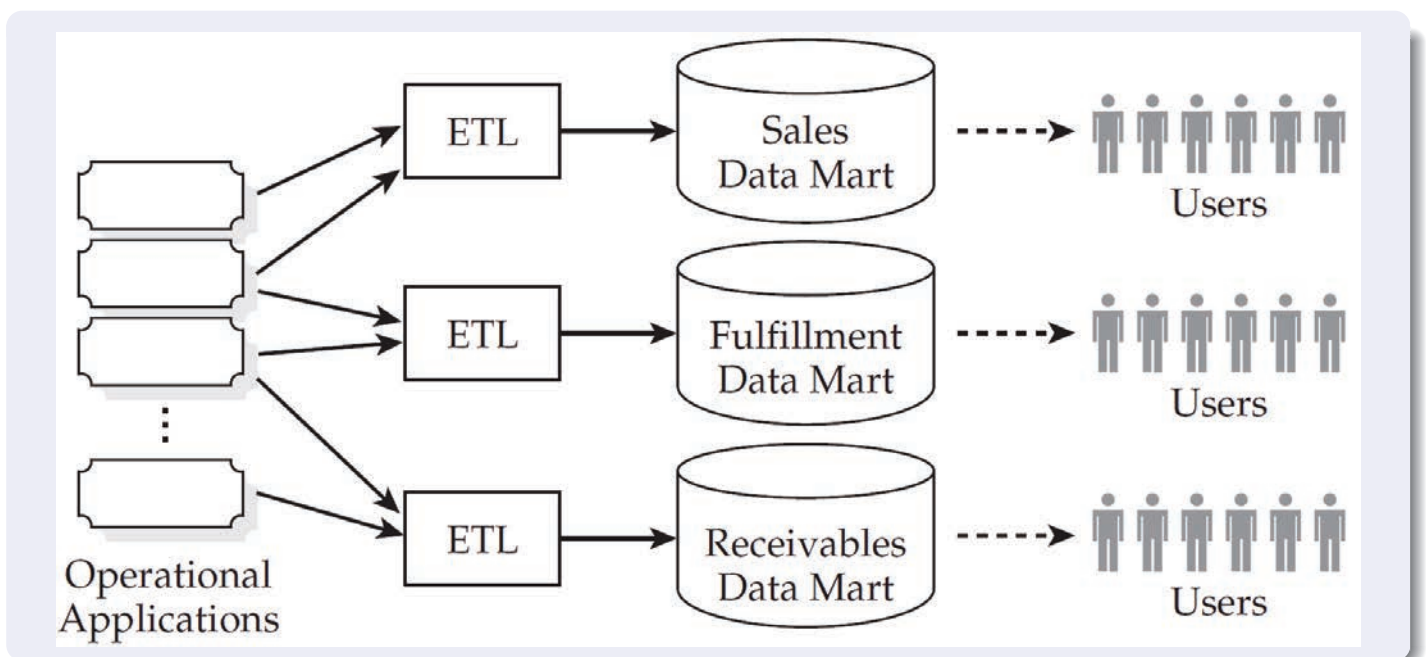
Enfoque mixto [GR09]



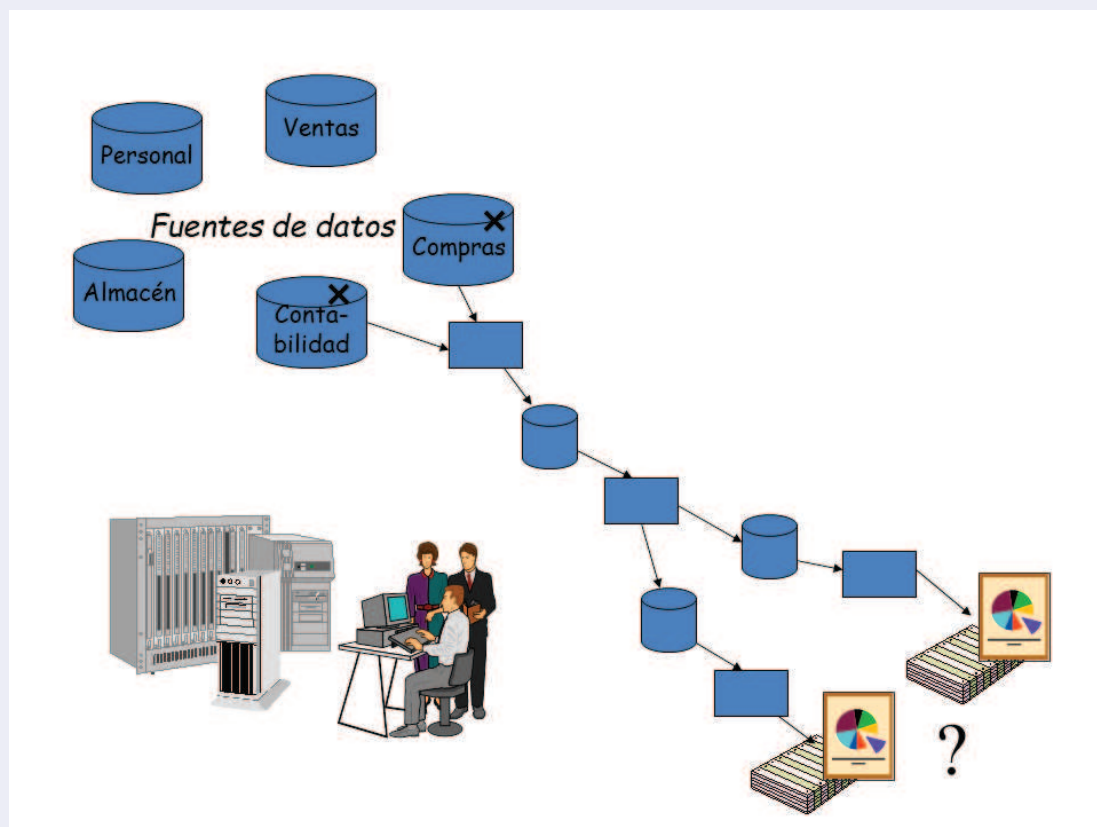
Desarrollo individual



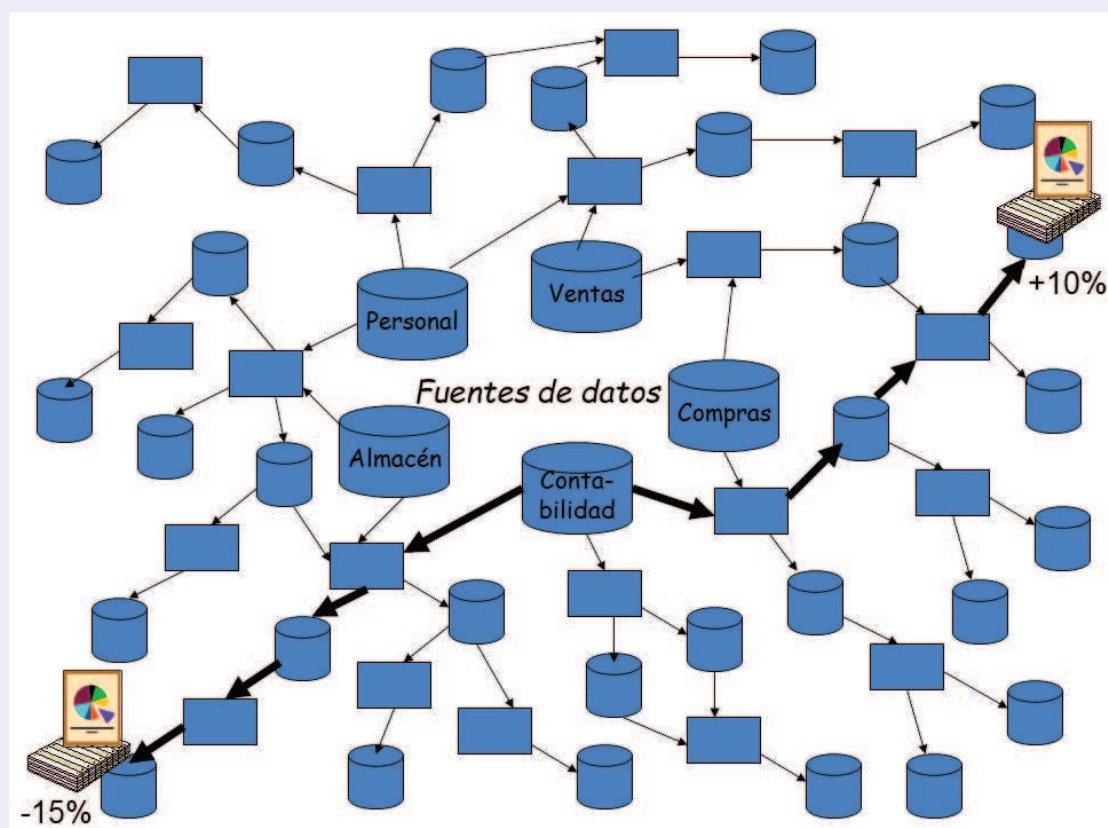
Múltiples desarrollos individuales



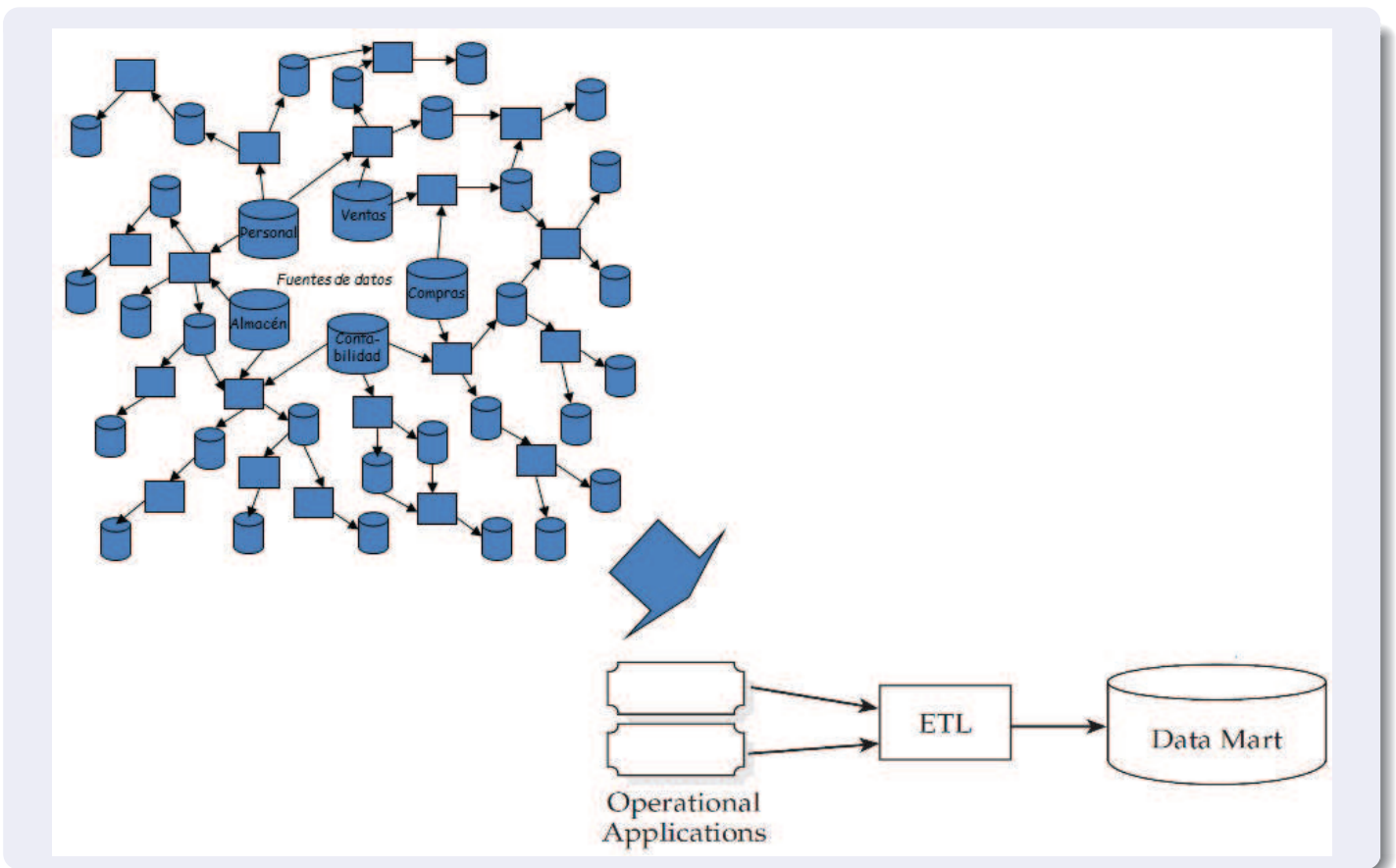
Desarrollo de un informe



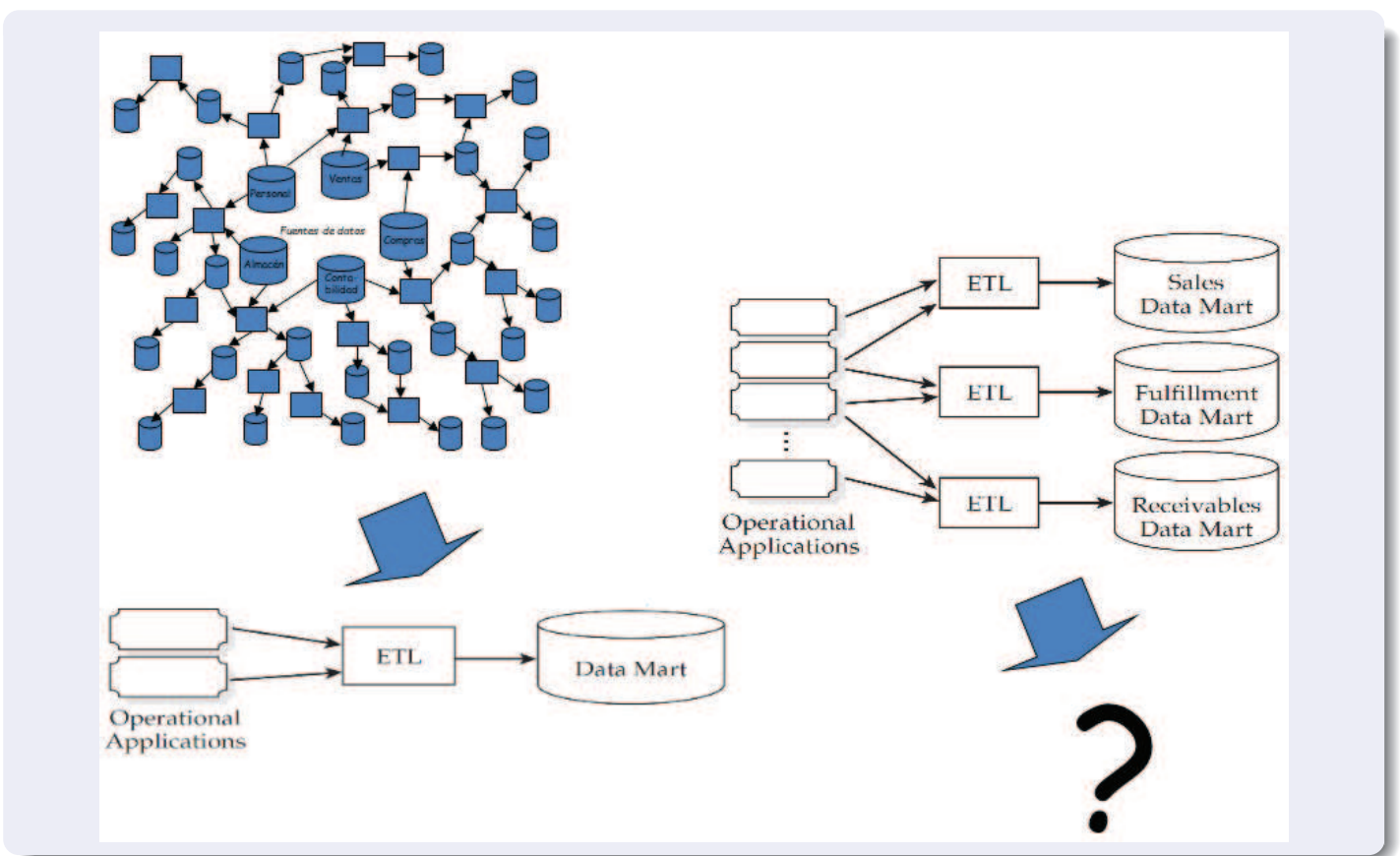
Tela de araña operacional (Inmon)



Sistema multidimensional como solución

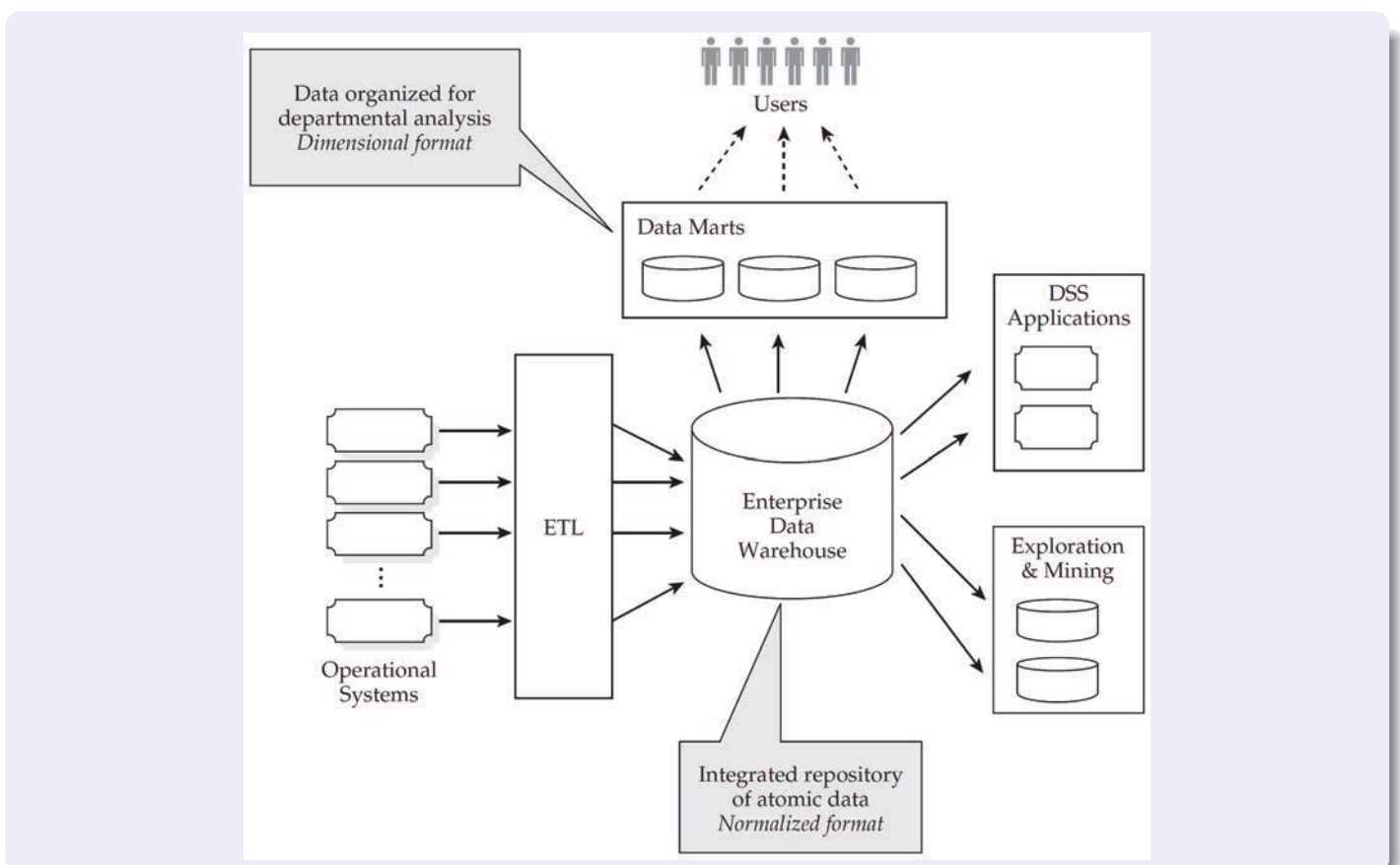


Proliferación de sistemas multidimensionales: ¿solución?

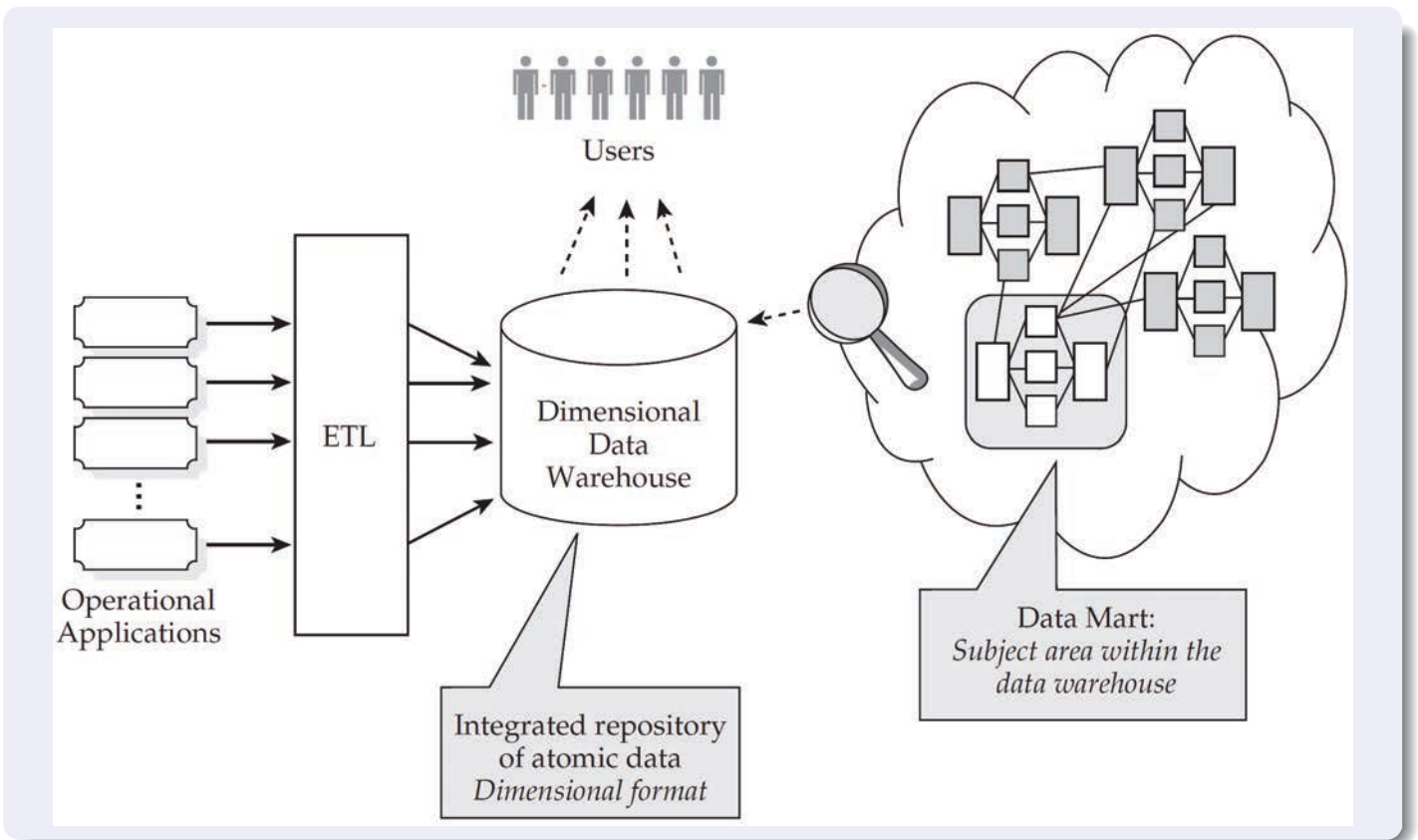


Arquitecturas de integración de sistemas

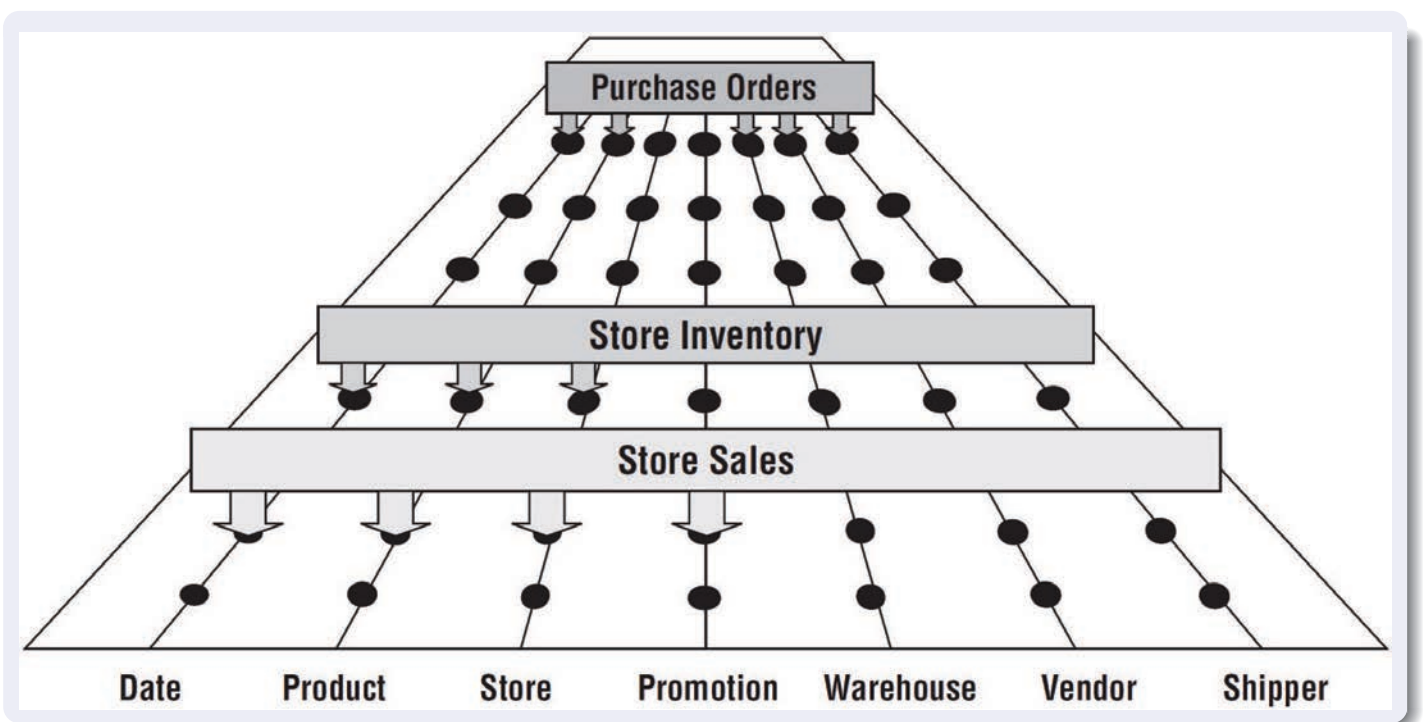
Arquitectura de Inmon



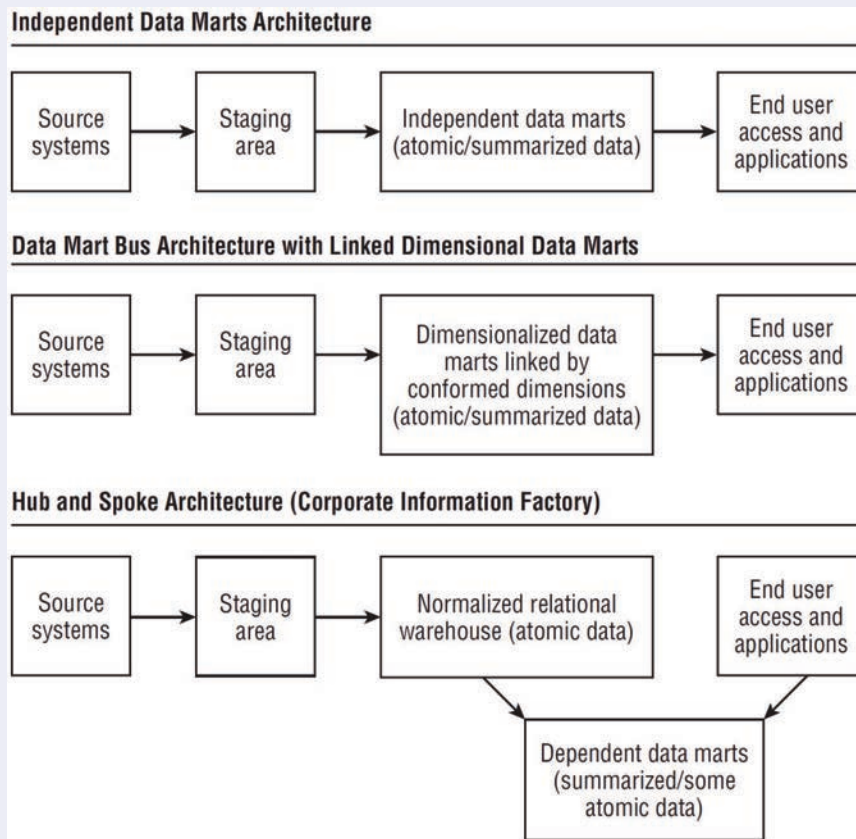
Arquitectura de Kimball



The Data Warehouse Bus Architecture



Comparación de arquitecturas (i)



Comparación de arquitecturas (ii)

Architecture	Advocate	Also Known As	Description	Role of Dimensional Design
Corporate Information Factory	Bill Inmon	<ul style="list-style-type: none"> Atomic data warehouse Enterprise data warehouse 	<ul style="list-style-type: none"> Enterprise data warehouse component is an integrated repository of atomic data It is <i>not</i> accessed directly Data marts reorganize data for departmental use/analysis 	Dimensional design used for data marts only
Dimensional Data Warehouse	Ralph Kimball	<ul style="list-style-type: none"> Enterprise data warehouse Bus architecture Architected data marts Virtual data marts 	<ul style="list-style-type: none"> Dimensional data warehouse is an integrated repository of atomic data It may be accessed directly Subject areas within the dimensional data warehouse sometimes called <i>data marts</i> Data marts not required to be separate databases 	All data is organized dimensionally
Stand-Alone Data Marts	No takers, yet common	<ul style="list-style-type: none"> Data mart Silo Stovepipe Island 	<ul style="list-style-type: none"> Subject area implementation without an enterprise context 	May employ dimensional design

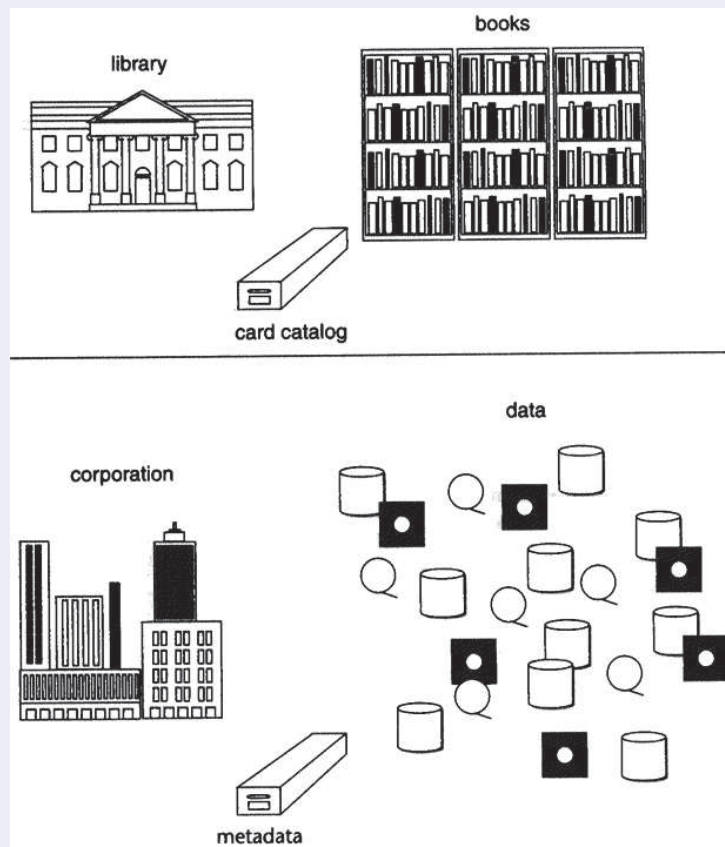
Comparación de arquitecturas (y iii)

	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	✗	n/a	n/a	Physical	Dimensional*	Yes

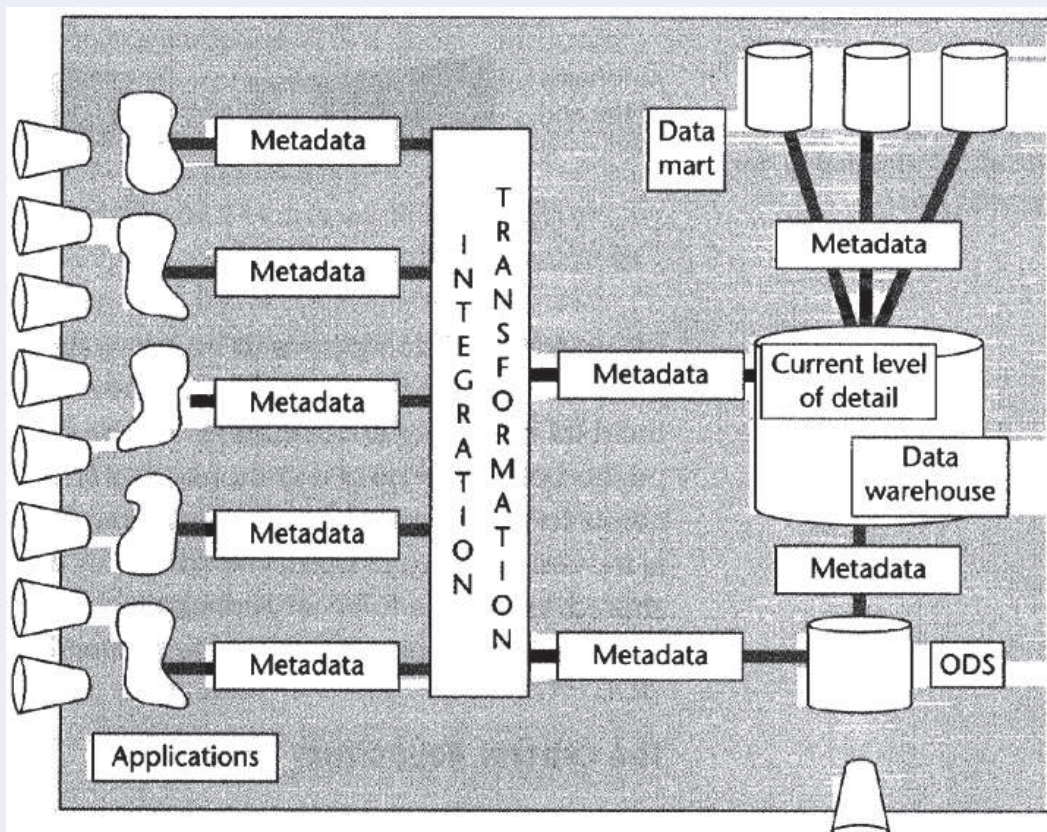
* Optional

Metadatos

Datos y metadatos



Metadatos en la arquitectura



Tipos y ejemplos de metadatos

- Metadatos de **desarrollo**: se crean y usan en el proceso de desarrollo de las aplicaciones y las BD.
- Metadatos de **control**: destinados a gestionar el funcionamiento de los almacenes de datos.
- Metadatos de **uso**: orientados a los usuarios, para ayudarles a entender los datos de negocio.

Data layout. Describing the relative position and format of data in a storage media—customer record.

- ❑ Cust-id char (15)
- ❑ Cust-name varchar (45)
- ❑ Cust-address varchar (45)
- ❑ Cust-balance dec fixed (15, 2)

Content. There are 150,000 occurrences of transaction X in table PLK.

Indexes. Table XYZ has indexes on the following columns:

- ❑ Column HYT
- ❑ Column BFD
- ❑ Column MJI

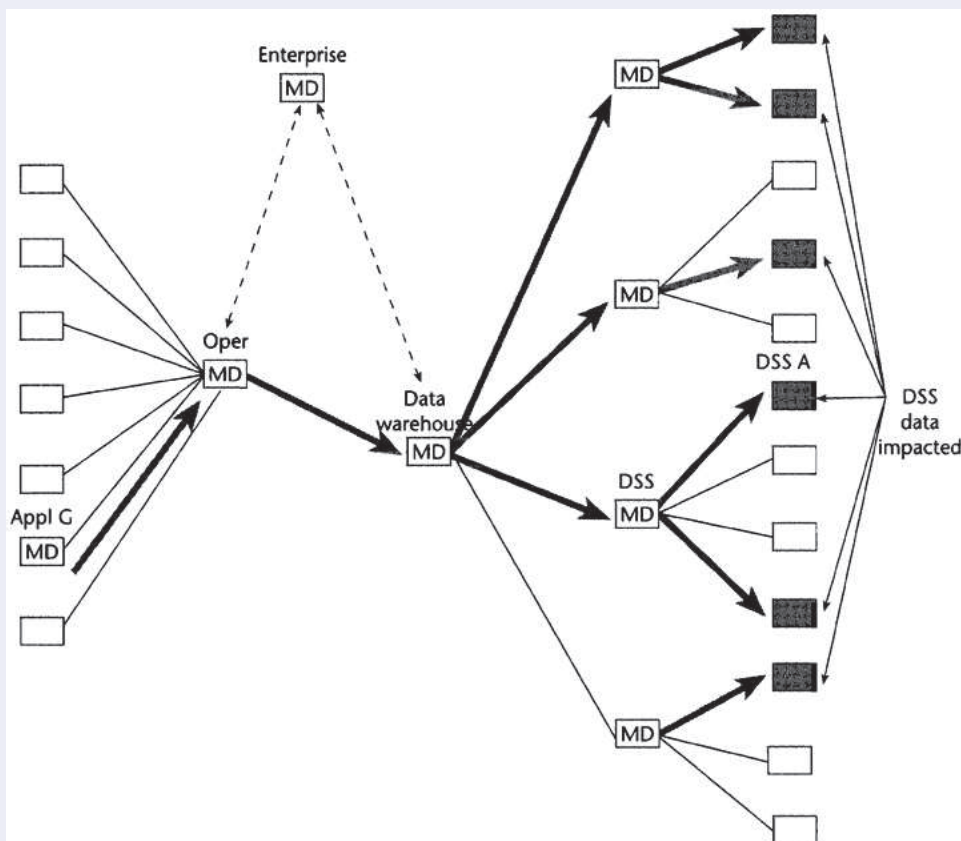
Refreshment scheduling. Table ABC is refreshed every Tuesday at 2:00 P.M.

Usage. Only 2 of the 10 columns in table ABC have been used over the past six months.

Referential integrity. Table XYZ is related to table ABC by means of the key QWE.

General documentation. "Table ABC was designed in 1975 as part of the new accounts payable system. Table ABC contains accounts overdue data as calculated by . . ."

Ejemplo 1: análisis de impacto



Ejemplo 2: resultados del proceso ETL

ETL Statistics Report										January 17, 2000	
Target Table	ETL Process	Target Table Type	Source System(s)	Processing Time	Load Cycle	Number of Records Inserted	Number of Records Updated	Records In Error	Total Number of Records	Load Date	Load Time
Customer	cs0001	Dimension	ERP	0:43:12	3	223	196	2	1,194	March 2, 2000	5:18:10
Customer	cs0002	Dimension	Order Processing	0:27:39	3	22	5	0	1,194	March 2, 2000	5:18:10
Customer	cs0003	Dimension	Sales Force	0:18:45	3	51	17	0	1,194	March 2, 2000	5:18:10
Product	pd001	Dimension	Order Processing	0:29:28	3	41	28	1	112	March 2, 2000	5:18:10
Employee	ee001	Dimension	ERP	0:37:55	3	23	15	0	276	March 2, 2000	5:18:10
Time	tm001	Dimension	External Source	0:05:41	3	1	0	0	3	March 2, 2000	5:18:10
Sales	sl001	Fact	Order Processing	2:35:30	3	36,908,928	0	17,172	58,095,382	March 2, 2000	5:18:10
Customer	cs0001	Dimension	ERP	0:41:45	2	287	201	0	898	February 1, 2000	3:36:38
Customer	cs0002	Dimension	Order Processing	0:25:32	2	15	1	0	898	February 1, 2000	3:36:38
Customer	cs0003	Dimension	Sales Force	0:16:48	2	78	24	0	898	February 1, 2000	3:36:38
Product	pd001	Dimension	Order Processing	0:28:34	2	31	24	0	71	February 1, 2000	3:36:38
Employee	ee001	Dimension	ERP	0:34:59	2	9	2	0	253	February 1, 2000	3:36:38
Time	tm001	Dimension	External Source	0:04:13	2	1	0	0	2	February 1, 2000	3:36:38
Sales	sl001	Fact	Order Processing	1:04:47	2	16,130,774	0	230	21,186,454	February 1, 2000	3:36:38
Customer	cs0001	Dimension	ERP	0:49:42	1	334	0	0	518	January 3, 2000	3:47:35
Customer	cs0002	Dimension	Order Processing	0:28:21	1	37	0	0	518	January 3, 2000	3:47:35
Customer	cs0003	Dimension	Sales Force	0:31:09	1	147	0	22	518	January 3, 2000	3:47:35
Product	pd001	Dimension	Order Processing	0:45:33	1	40	0	0	40	January 3, 2000	3:47:35
Employee	ee001	Dimension	ERP	0:50:18	1	244	0	0	244	January 3, 2000	3:47:35
Time	tm001	Dimension	External Source	0:02:21	1	1	0	0	1	January 3, 2000	3:47:35
Sales	sl001	Fact	Order Processing	0:20:11	1	5,055,680	0	0	5,055,680	January 3, 2000	3:47:35

Ejemplo 3: información adicional en los informes (i)

1999 Monthly Global Sales Report				January 7, 2000	
Month	Product Category	Sales \$ (in thousands) U.S.	Sales \$ (in thousands) International	Sales \$ (in thousands) Total	
December	TV	22,101	10,200	32,301	
	VCR	11,190	4,300	15,490	
	Cellular Phone	12,190	7,193	19,383	
	Digital	4,002	1,301	5,303	
	Miscellaneous	1,209	870	2,079	
November	TV	42,000	22,200	64,200	
	VCR	21,190	9,878	31,068	
	Cellular Phone	28,193	12,193	40,386	
	Digital	8,901	2,901	11,802	
	Miscellaneous	2,730	1,530	2,730	
October	TV	70,100	32,950	103,050	
	VCR	31,900	14,878	46,778	
	Cellular Phone	41,700	17,550	59,250	
	Digital	20,000	4,100	24,100	
	Miscellaneous	4,850	2,850	7,700	

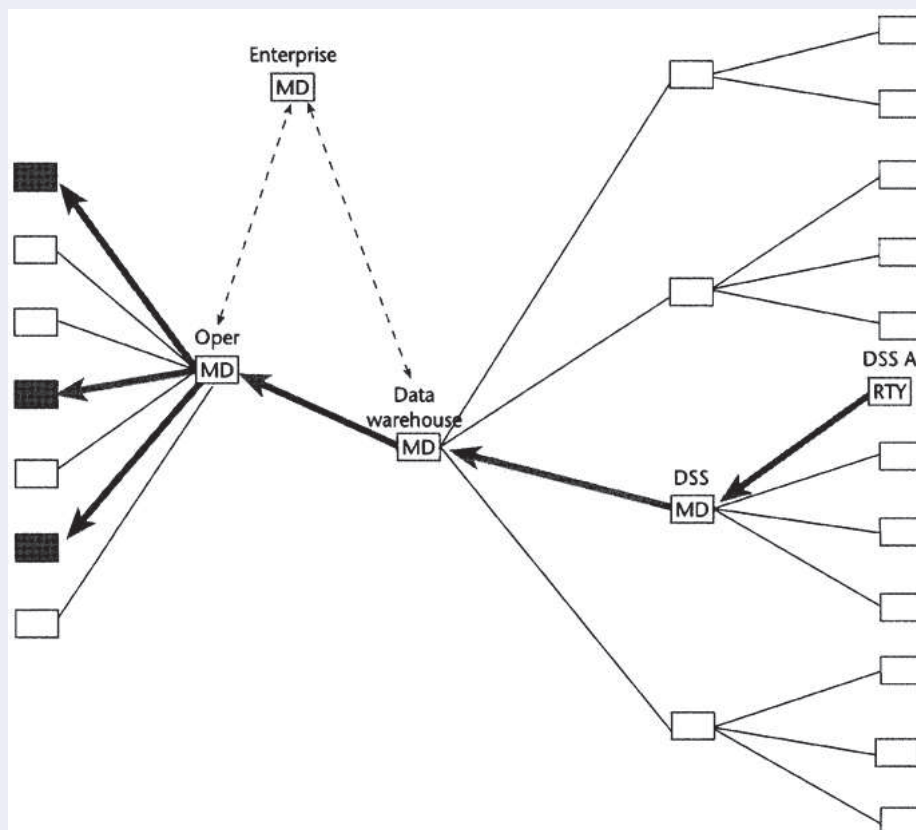
U.S. sales includes the United States, Canada, and Mexico, but does not subtract sales dollars from returned orders

Ejemplo 4: información adicional en los informes (y ii)

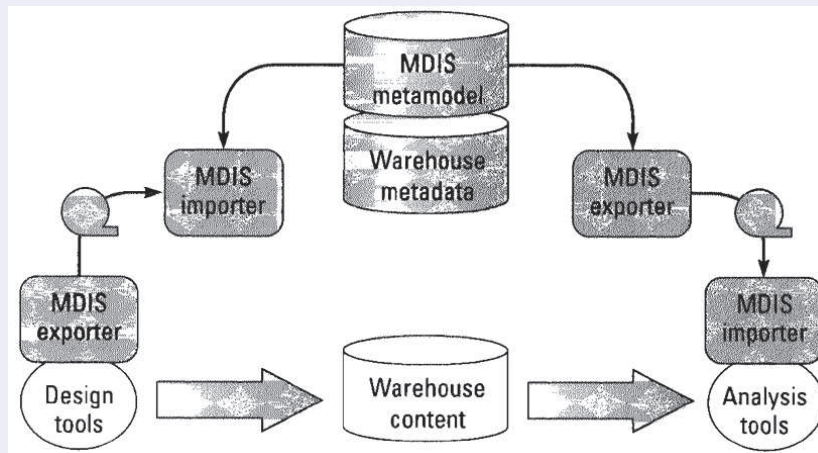
1999 Monthly Global Sales Report				January 7, 2000
Month	Product Category	Sales \$ (in thousands) U.S.	Sales \$ (in thousands) International	Sales \$ (in thousands) Total
December	TV	22,101	10,200	32,301
	VCR	11,190	4,300	15,490
	Cellular Phone	12,190	7,193	19,383
	Digital	4,002	1,301	5,303
	Miscellaneous	1,209	870	2,079
November	TV	42,000	22,200	64,200
	VCR	21,190	9,878	31,068
	Cellular Phone	28,193	12,193	40,386
	Digital	8,901	2,901	11,802
	Miscellaneous	2,730	1,530	2,730
October	TV	70,100	32,950	103,050
	VCR	31,900	14,878	46,778
	Cellular Phone	41,700	17,550	59,250
	Digital	20,000	4,100	24,100
	Miscellaneous	4,850	2,850	7,700

Data Quality Tracking Statistics
 8.4% of the dollar value were not loaded
 1.7% of the records were not loaded

Ejemplo 5: traza de los datos

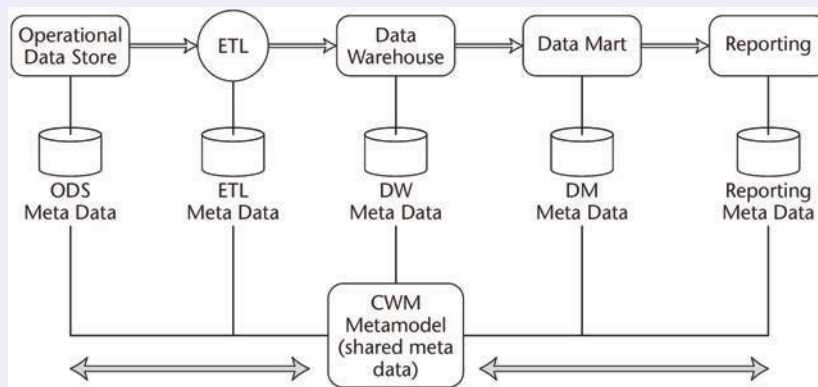


Estándares de metadatos



- MDIS (*Metadata Interchange Specifications*).
- CWM (*Common Warehouse Metamodel*), de OMG (*Object Management Group*), basado en:
 - ▶ MOF (Meta Object Facility),
 - ▶ UML (Unified Modeling Language),
 - ▶ XMI (XML Metadata Interchange).

CWM



Capas del CWM

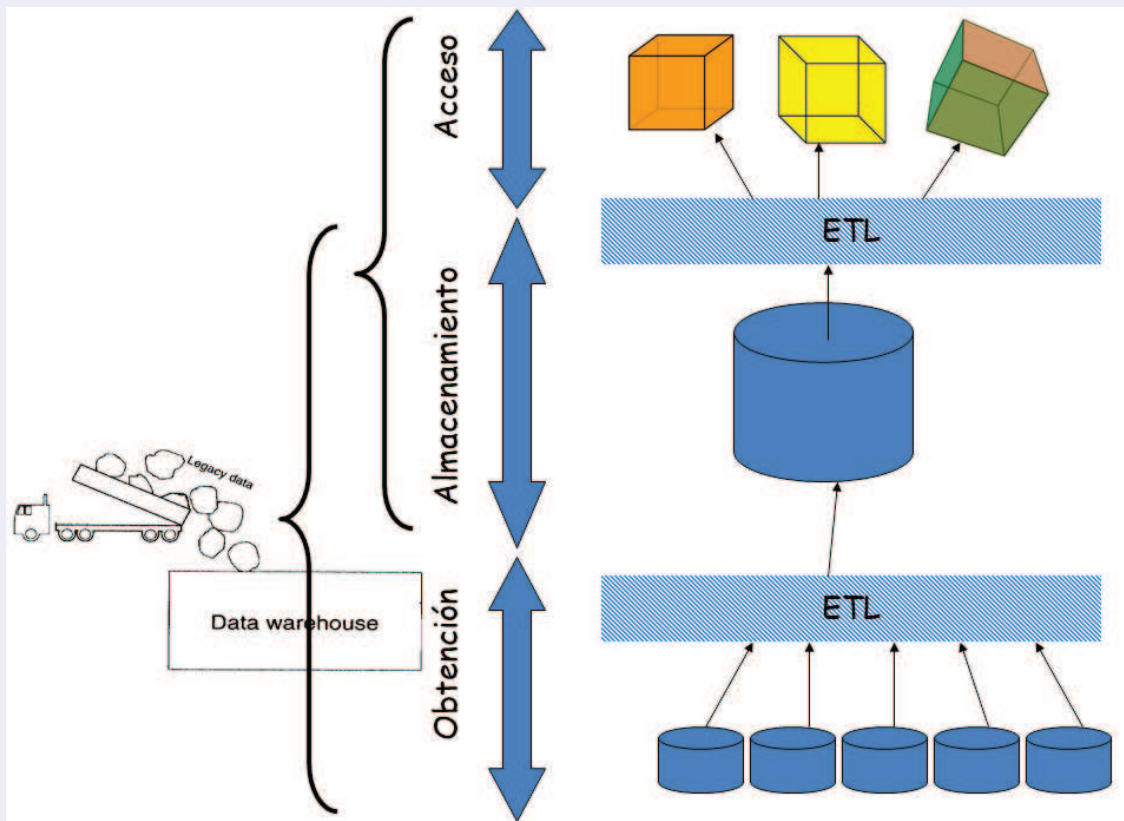
Management	Warehouse Process			Warehouse Operation	
Analysis	Transformation	OLAP	Data Mining	Information Visualization	Business Nomenclature
Resource	Object	Relational	Record	Multi-dimensional	XML
Foundation	Business Information	Data Types	Expressions	Keys and Indexes	Software Deployment
Object Model	Core	Behavioral	Relationships	Instance	

Implementación de la arquitectura

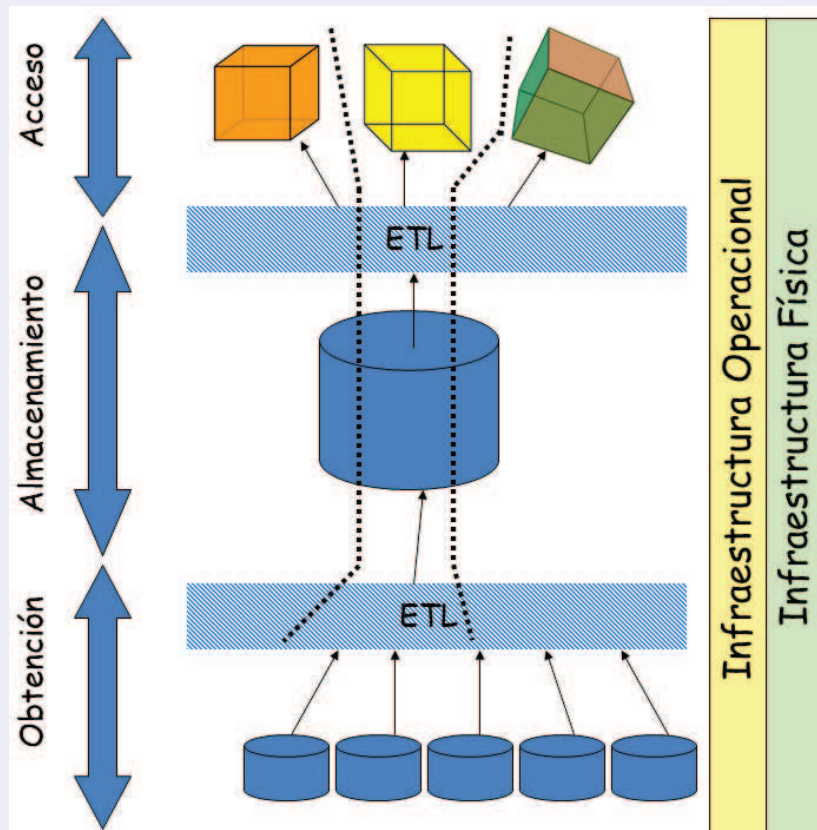
Propuesta de pasos para el desarrollo de un proyecto

1. Analysis (four weeks)—Recruit users into JAD sessions to develop data requirements and screen layouts. (JAD: Joint Application Development)
2. Design (two weeks)—Normalize the data, generate DDL, and build tables.
3. Identify sources (one week)—Identify where the data should come from.
4. Data extraction (three months)—Find some programmers to read the data out of the old system for loading into the new.
5. Load (one week)—load the extracted into the tables.
6. Install the data-mining product (one month)—Install the data-mining product and point it towards the right tables.
7. User training (one week)—Send the users to a one-week training class in how to use the tool.
8. Project done.

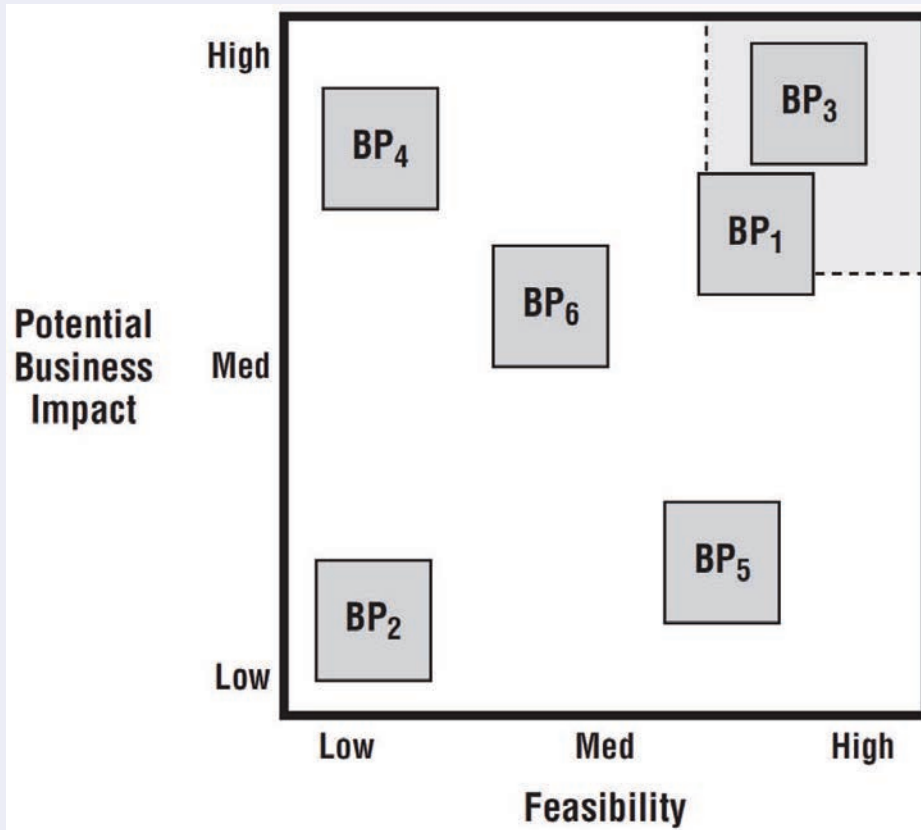
Proyectos parciales (*Back Room* vs. *Front Room*)



División en proyectos



Priorización proyectos



Errores comunes

Common Pitfalls to Avoid (Kimball)

- **Pitfall 10:** Become overly enamored with technology and data rather than focusing on the business's requirements and goals.
- **Pitfall 9:** Fail to embrace or recruit an influential, accessible, and reasonable senior management visionary as the business sponsor of the DW/BI effort.
- **Pitfall 8:** Tackle a galactic multiyear project rather than pursuing more manageable, although still compelling, iterative development efforts.
- **Pitfall 7:** Allocate energy to construct a normalized data structure, yet run out of budget before building a viable presentation area based on dimensional models.
- **Pitfall 6:** Pay more attention to back room operational performance and ease-of-development than to front room query performance and ease of use.
- **Pitfall 5:** Make the supposedly queryable data in the presentation area overly complex. Database designers who prefer a more complex presentation should spend a year supporting business users; they'd develop a much better appreciation for the need to seek simpler solutions.
- **Pitfall 4:** Populate dimensional models on a standalone basis without regard to a data architecture that ties them together using shared, conformed dimensions.
- **Pitfall 3:** Load only summarized data into the presentation area's dimensional structures.
- **Pitfall 2:** Presume the business, its requirements and analytics, and the underlying data and the supporting technology are static.
- **Pitfall 1:** Neglect to acknowledge that DW/BI success is tied directly to business acceptance. If the users haven't accepted the DW/BI system as a foundation for improved decision making, your efforts have been exercises in futility.

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.
- KR13 R. Kimball, M. Ross: *The Data Warehouse Toolkit (3rd Edition)*. Wiley, 2013.
- Poo03 J. Poole et al.: *Common Warehouse Metamodel Developer's Guide*. Wiley, 2003.