Multidimensional Design

Knowledge Objectives

1. Justify the usefulness of multidimensional analysis against operational databases and spreadsheets
2. Define OLAP (On-Line Analytical Processing)
3. Describe a data cube
4. Describe the most typical multidimensional operations over the cube
5. Distinguish the main kinds of multidimensional tools (ROLAP, MOLAP and HOLAP)

Understanding Objectives

1. Explain the meaning of a star (or its variants snowflake and galaxy) shape multidimensional schema
2. Translate a multidimensional UML diagram into a relational star-join schema
3. Elaborate on the advantages and disadvantages of ROLAP in front of MOLAP

MOTIVATION AND DEFINITION

Spreadsheets

- Absence of metadata
  - Rows and columns without associated meaning
  - Difficult query and interpretations
- Limited amount of data
  - M$Excel (65,000*256=16,000,000 cells)
- The position limits operations
- Aggregation hierarchies are not managed

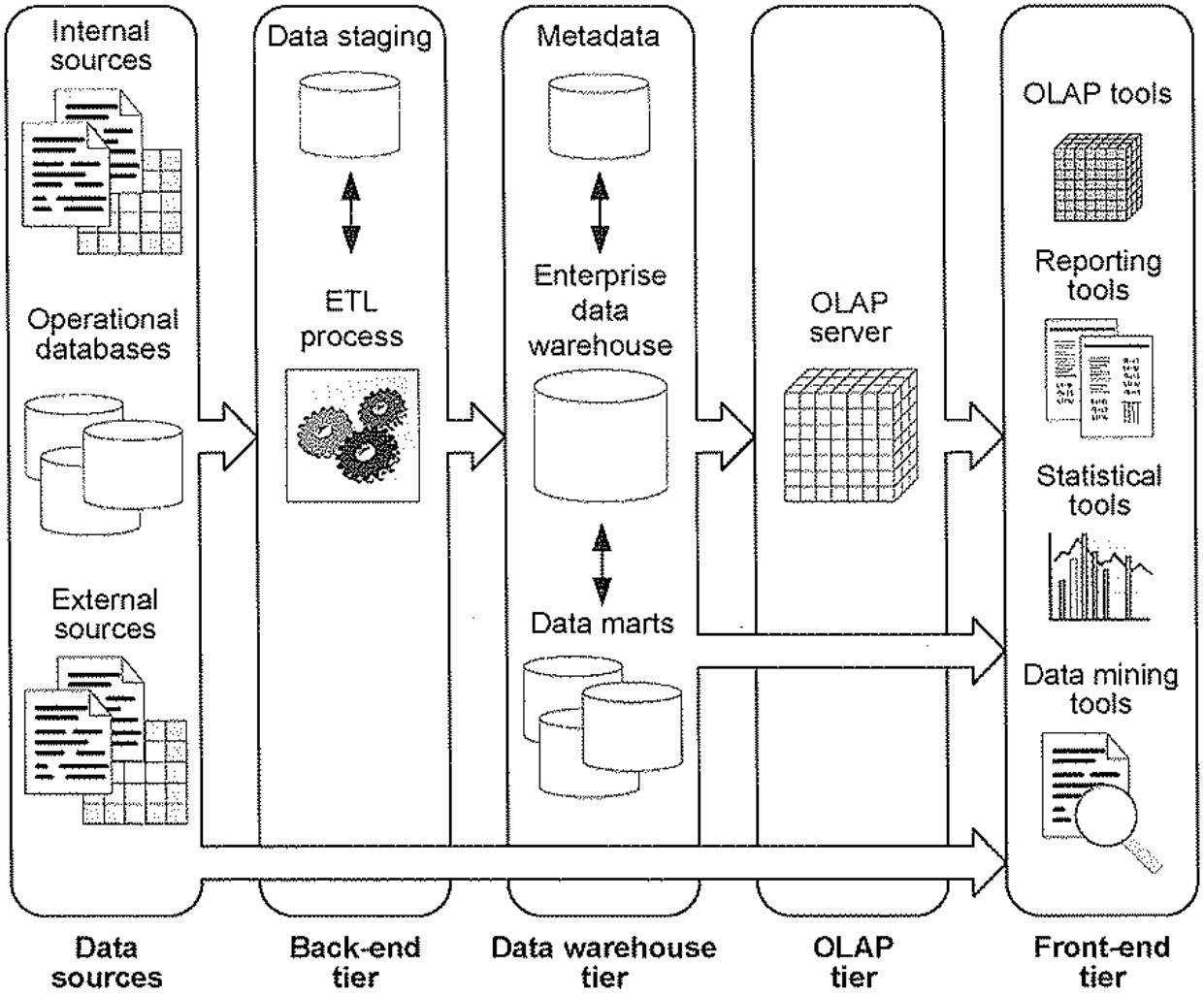
FASMI test

- Fast
- Analysis
- of Shared
- Multidimensional
- Information

Nigel Pendse, 1995

Spreadsheet -> FAMSI
Operational DB -> SIFAM

Reference architecture



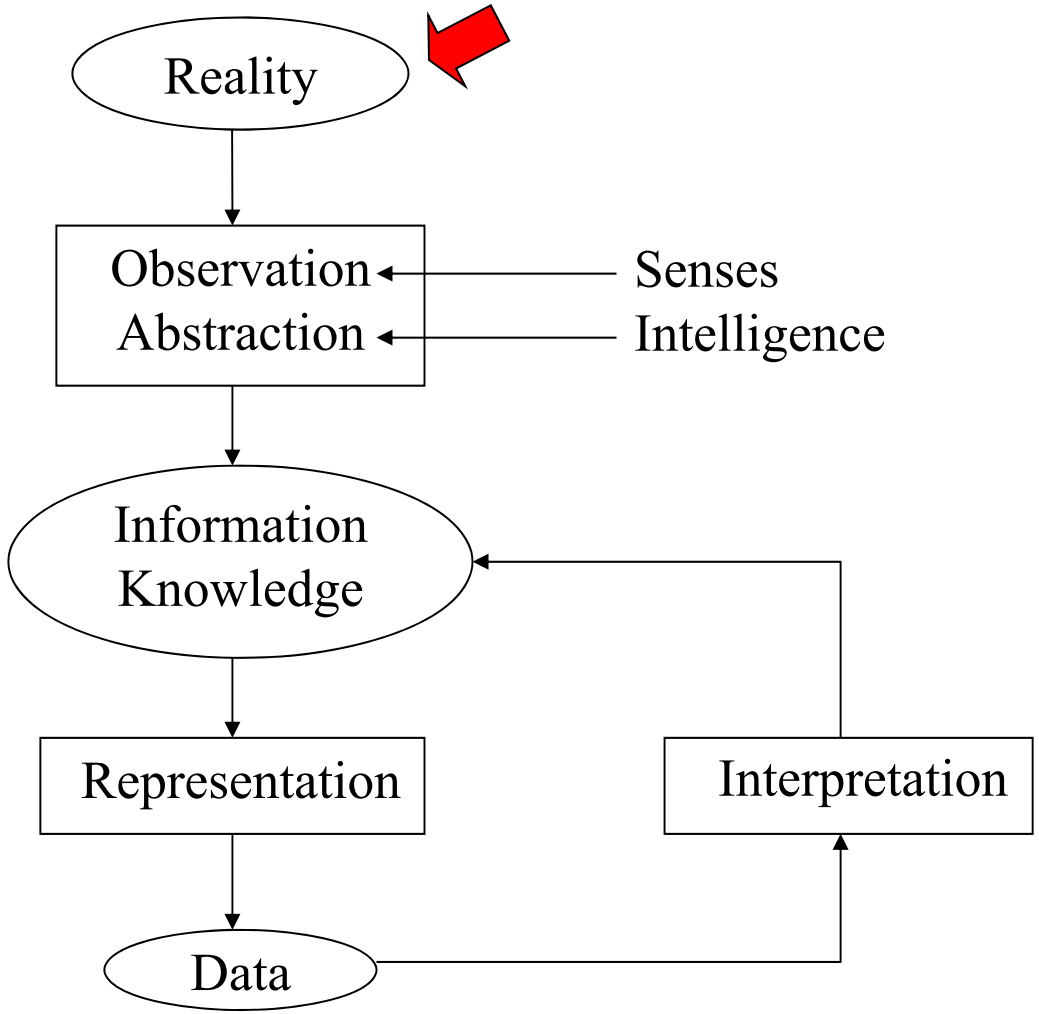
REAL WORLD VIEW

Three worlds

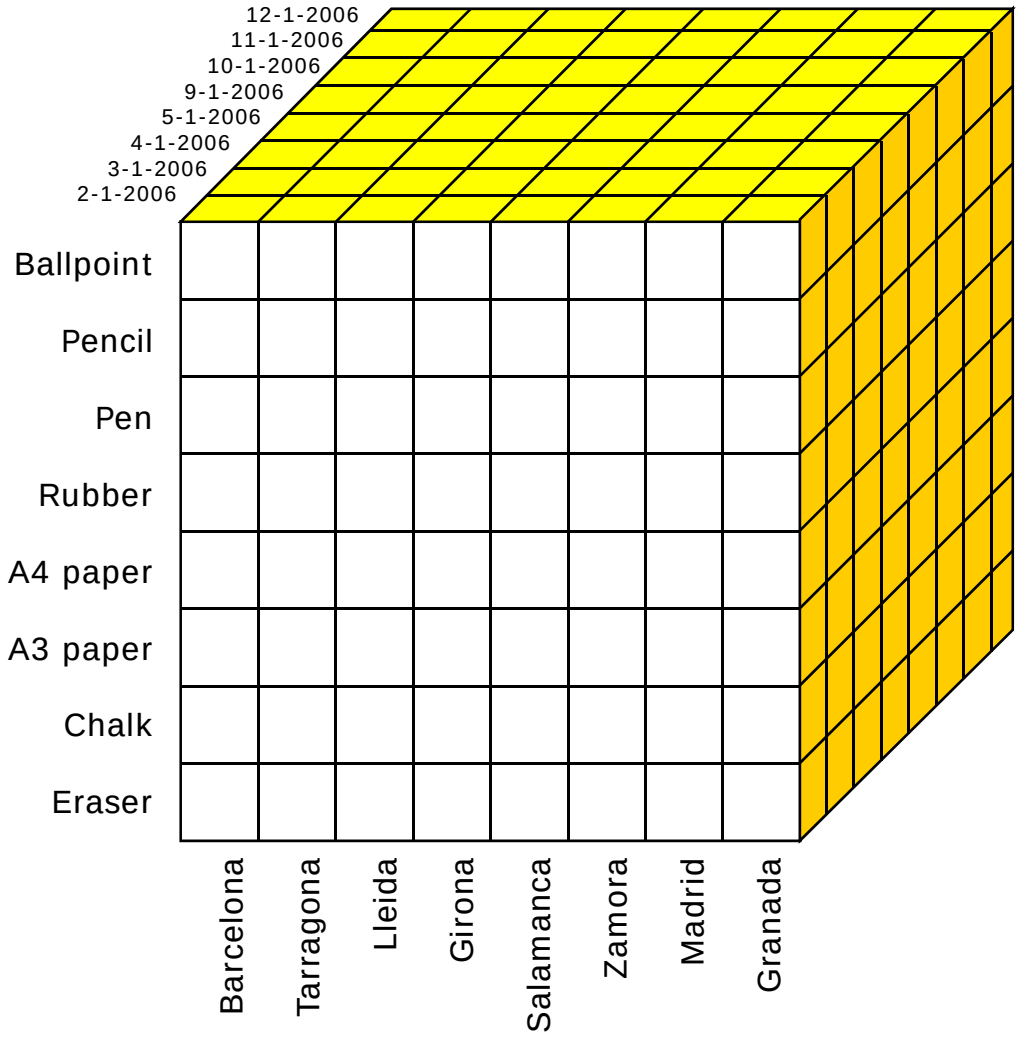
Real world
(unique)

Conceptual world
(multiple)

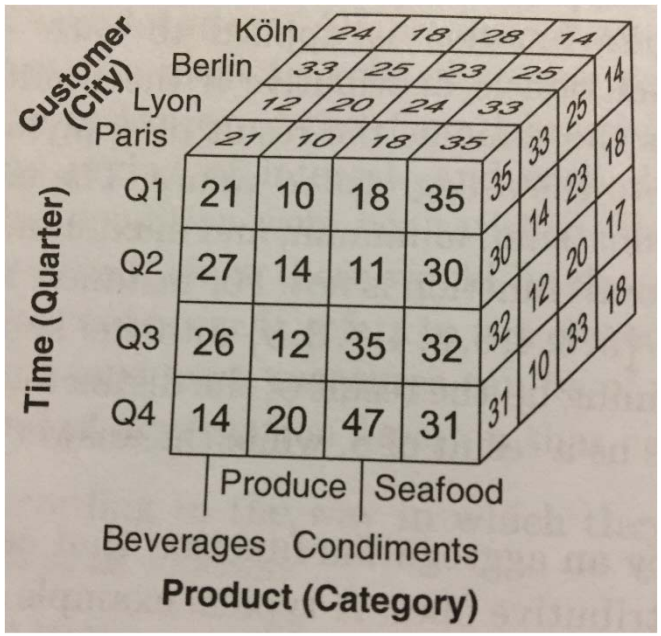
Representation world
(multiple)



Cube (Hyperprism)

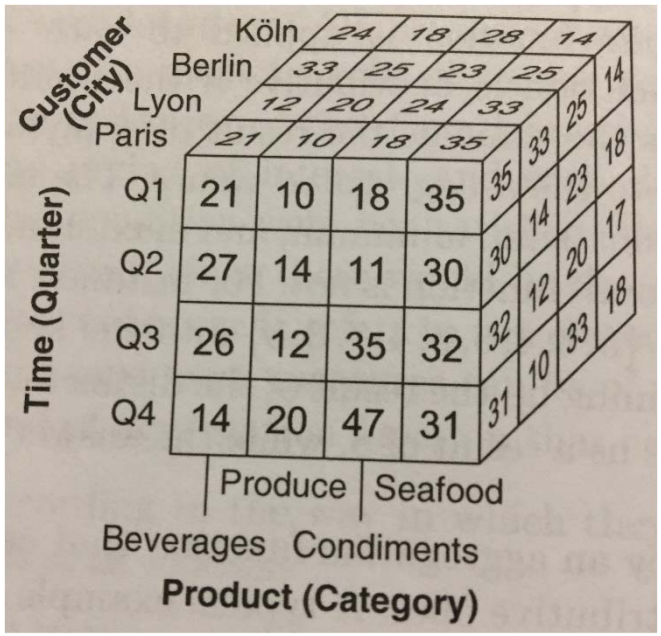


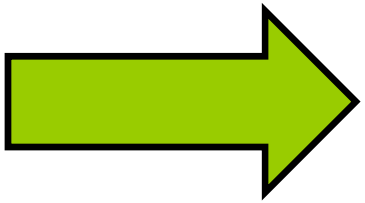
Cross-tab view

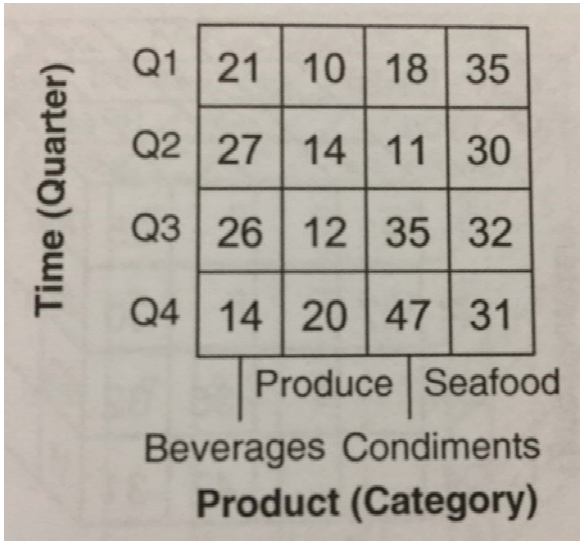


		Q1	Q2	Q3	Q4
Paris	Beverages	21	10	18	35
	Produce	27	14	11	30
	Condiments	26	12	35	32
	Seafood	14	20	47	31
Lyon	Beverages	12	20	24	33
	Produce				
	Condiments				
	Seafood				
Berlin	Beverages	33	25	23	25
	Produce				
	Condiments				
	Seafood				
Köln	Beverages	24	18	28	14
	Produce				
	Condiments				
	Seafood				

Slice

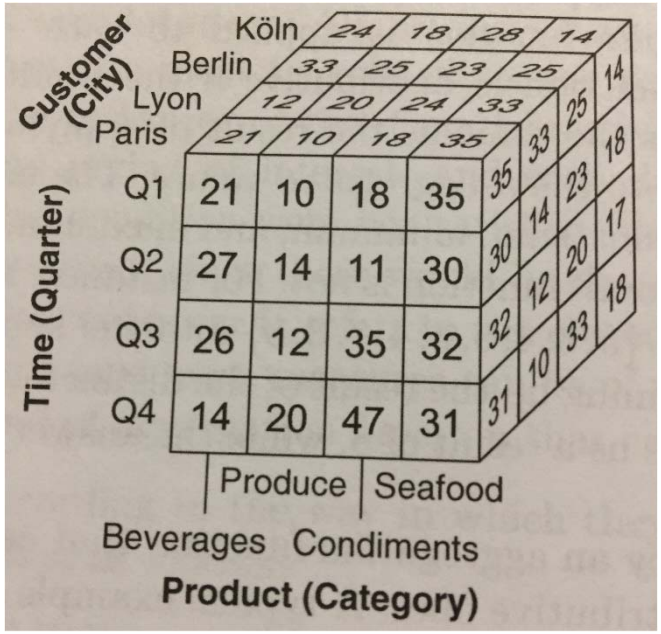


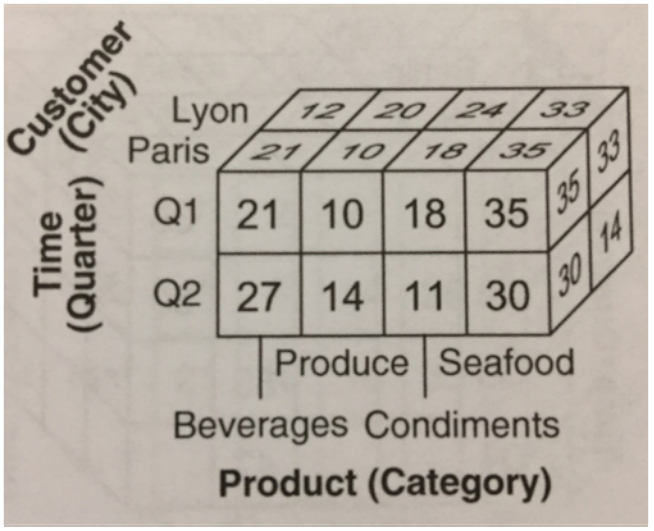




A. Vaisman & E.Zimanyi

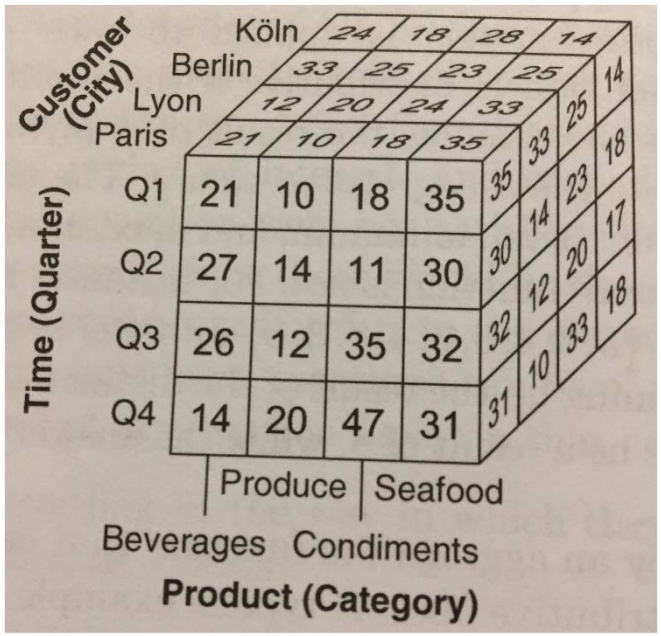
Dice

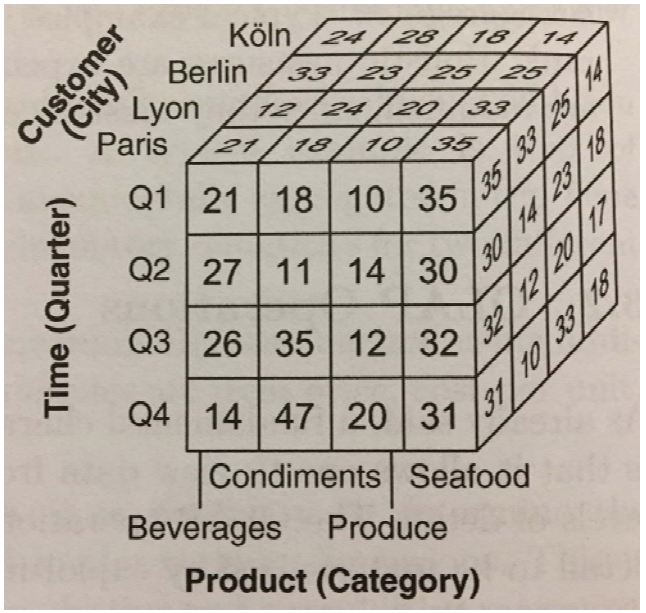




A. Vaisman & E.Zimanyi

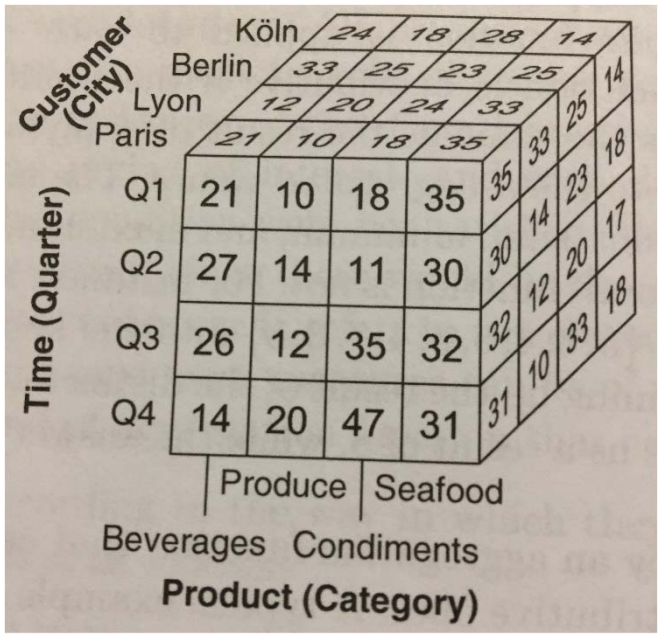
Sort

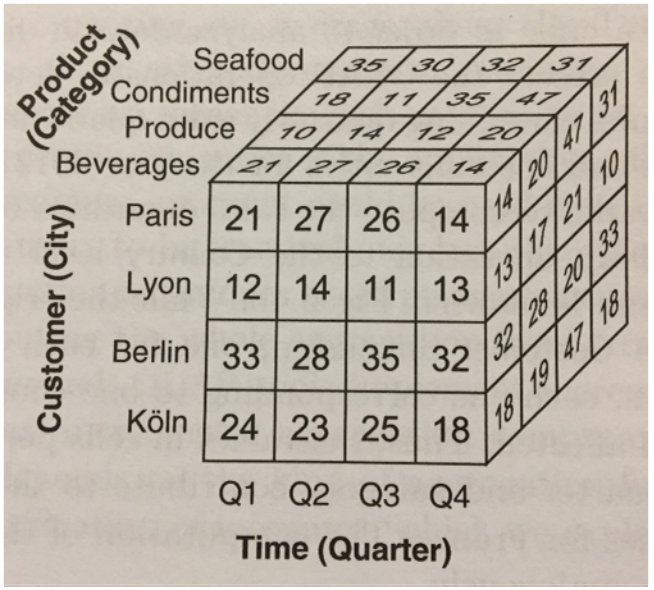




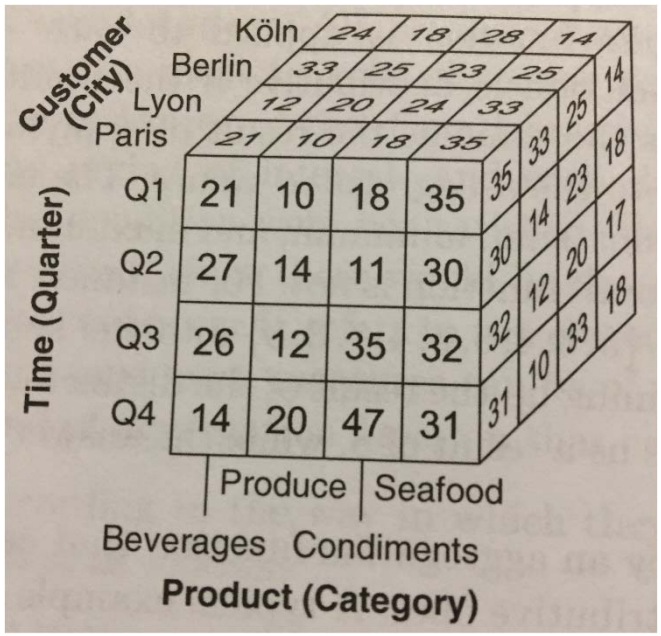
A. Vaisman & E.Zimanyi

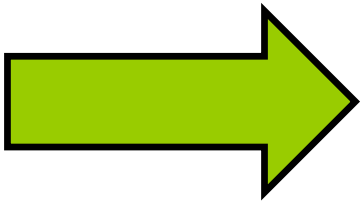
Pivot

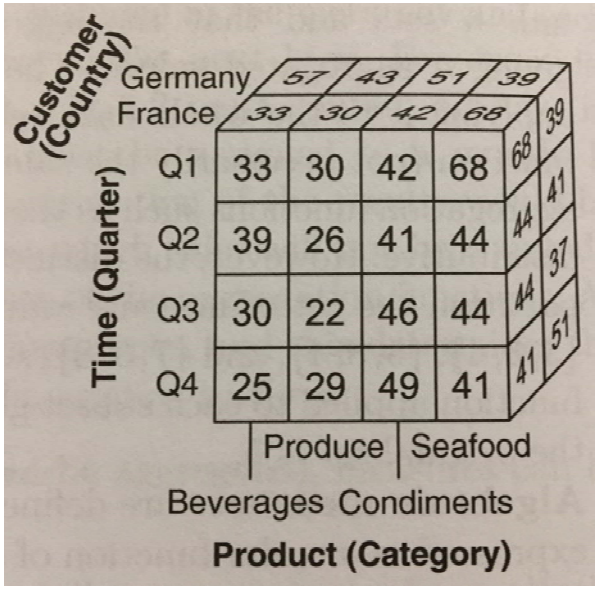

Customer (City)
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1
Q2
Q3
Q4
Produce
Seafood
Beverages
Condiments
Product (Category)


Product (Category)
Seafood
Condiments
Produce
Beverages
Customer (City)
Paris
Lyon
Berlin
Köln
Q1
Q2
Q3
Q4
Time (Quarter)

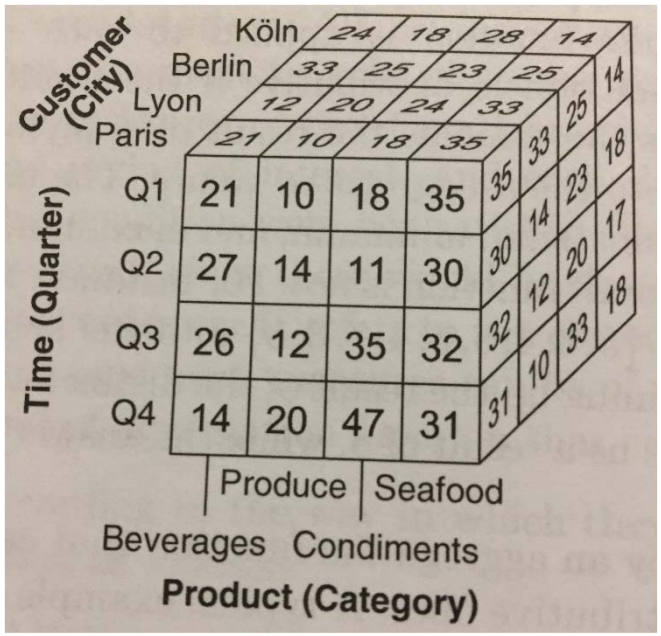
Roll-up (I)

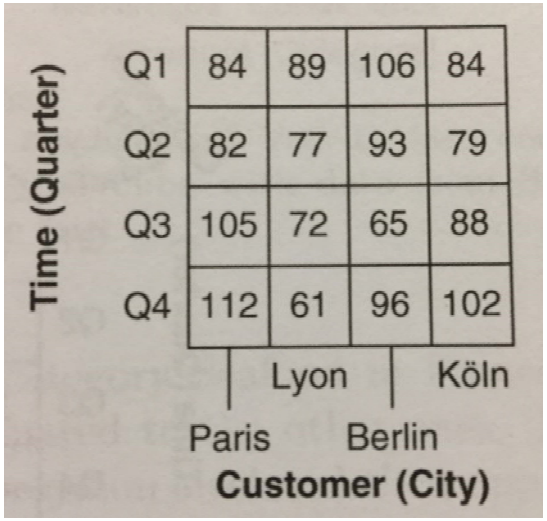

Customer (City)
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1
Q2
Q3
Q4
Produce
Seafood
Beverages
Condiments
Product (Category)



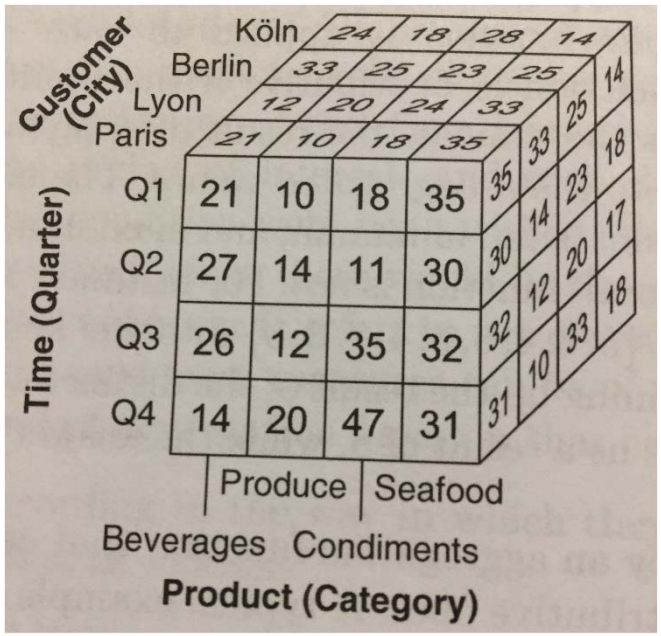

Customer (Country)
Germany
France
Time (Quarter)
Q1
Q2
Q3
Q4
Produce
Seafood
Beverages
Condiments
Product (Category)

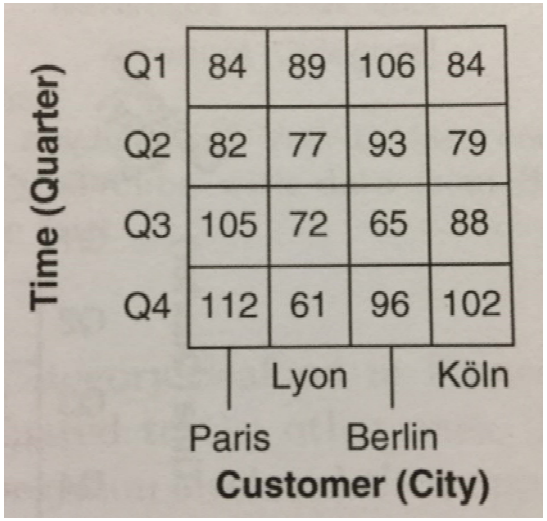
Roll-up (II)


Customer (City)
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1
Q2
Q3
Q4
Produce
Seafood
Beverages
Condiments
Product (Category)

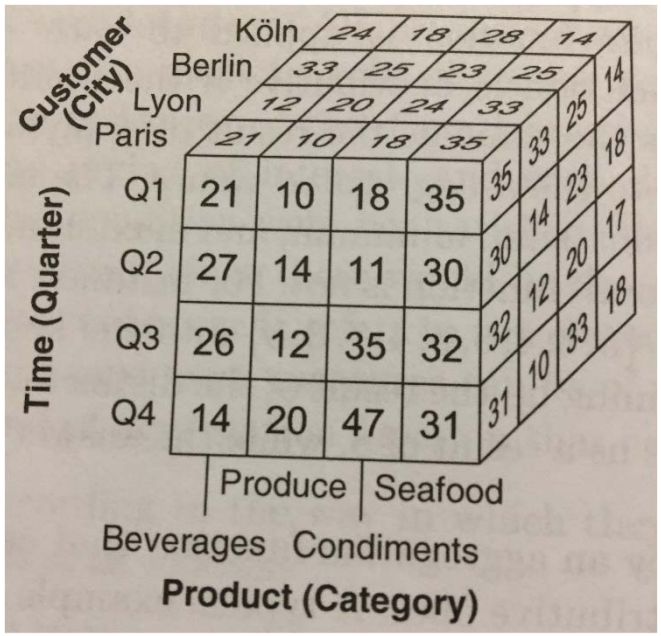

Time (Quarter)
Q1 84 89 106 84
Q2 82 77 93 79
Q3 105 72 65 88
Q4 112 61 96 102
Lyon
Köln
Paris
Berlin
Customer (City)

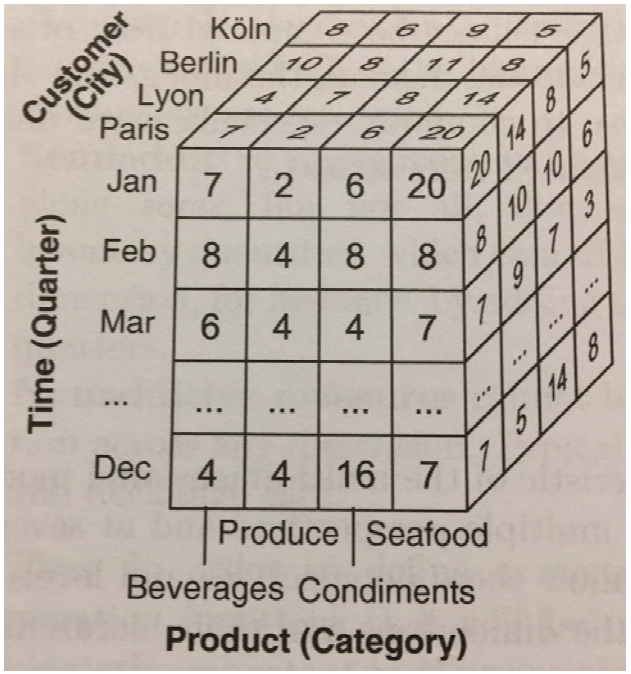
Roll-up (III)


Customer (City)
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1
Q2
Q3
Q4
Produce
Seafood
Beverages
Condiments
Product (Category)


Time (Quarter)
Q1 84 89 106 84
Q2 82 77 93 79
Q3 105 72 65 88
Q4 112 61 96 102
Lyon
Köln
Paris
Berlin
Customer (City)

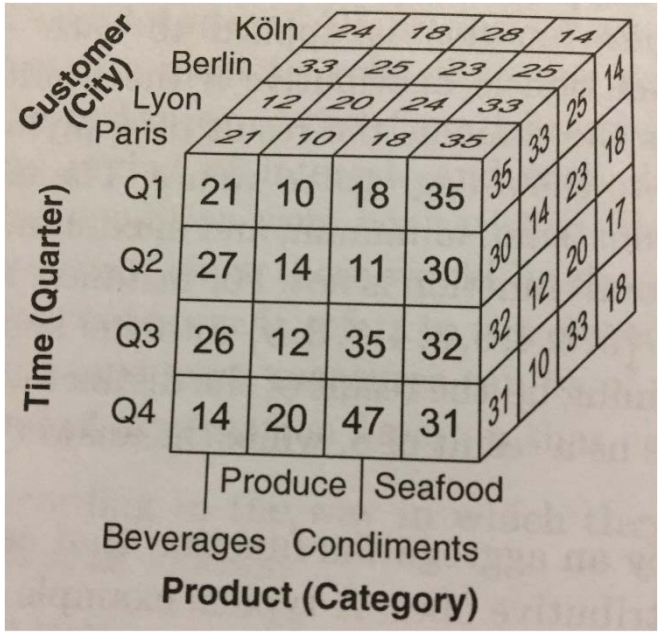
Drill-down

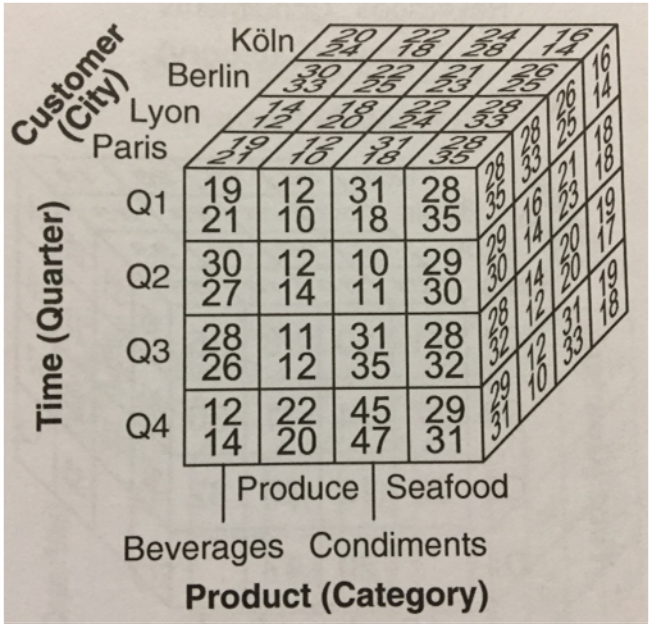




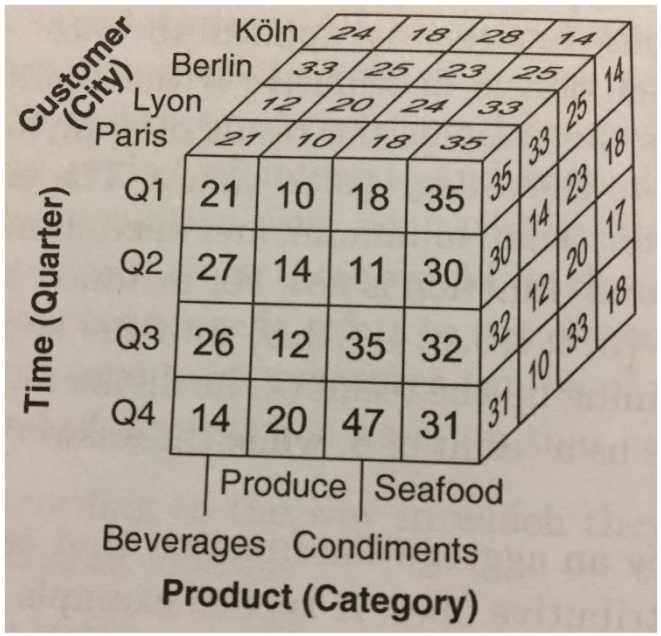
A. Vaisman & E.Zimanyi

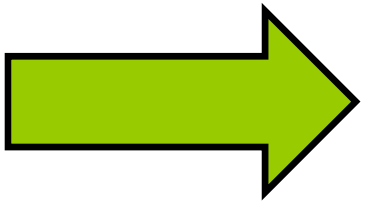
Drill-across

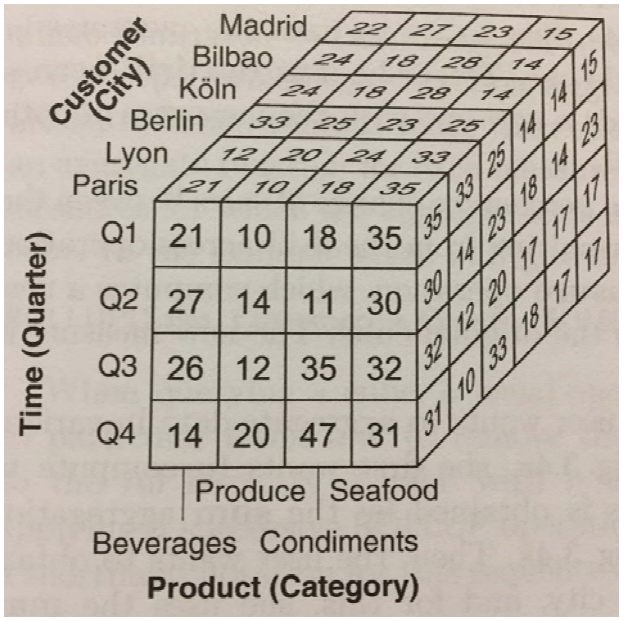




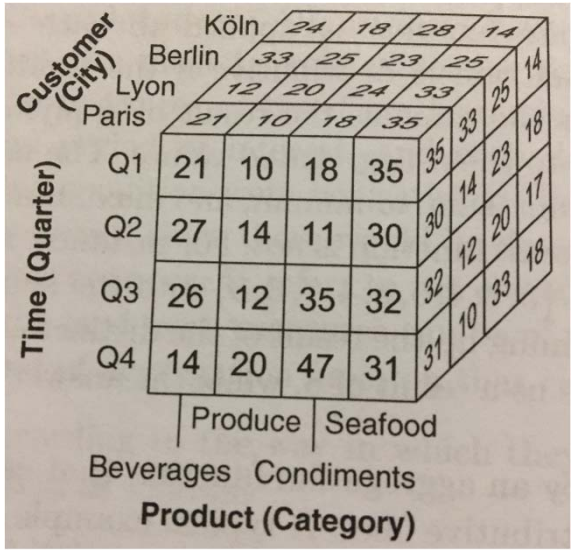
Union

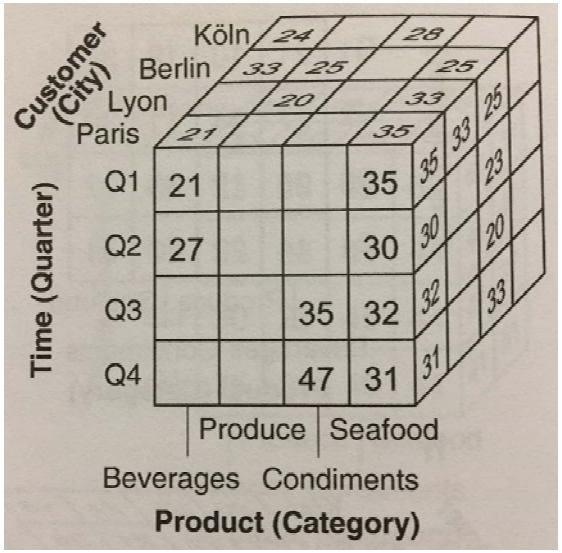

Customer (City)
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1
Q2
Q3
Q4
21 10 18 35
27 14 11 30
26 12 35 32
14 20 47 31
Produce
Seafood
Beverages
Condiments
Product (Category)

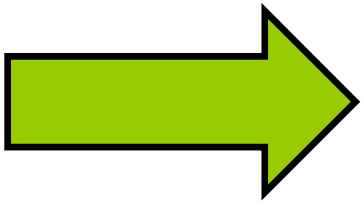


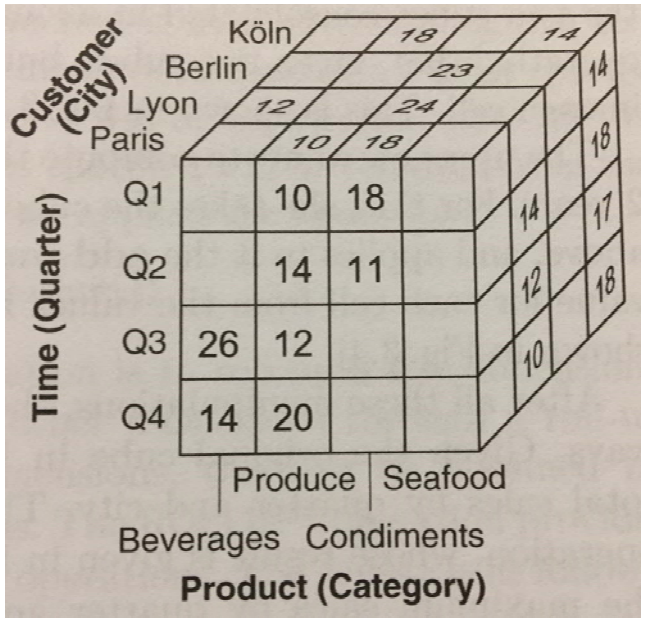

Customer (City)
Madrid
Bilbao
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1
Q2
Q3
Q4
21 10 18 35
27 14 11 30
26 12 35 32
14 20 47 31
Produce
Seafood
Beverages
Condiments
Product (Category)

Difference









A. Vaisman & E.Zimanyi

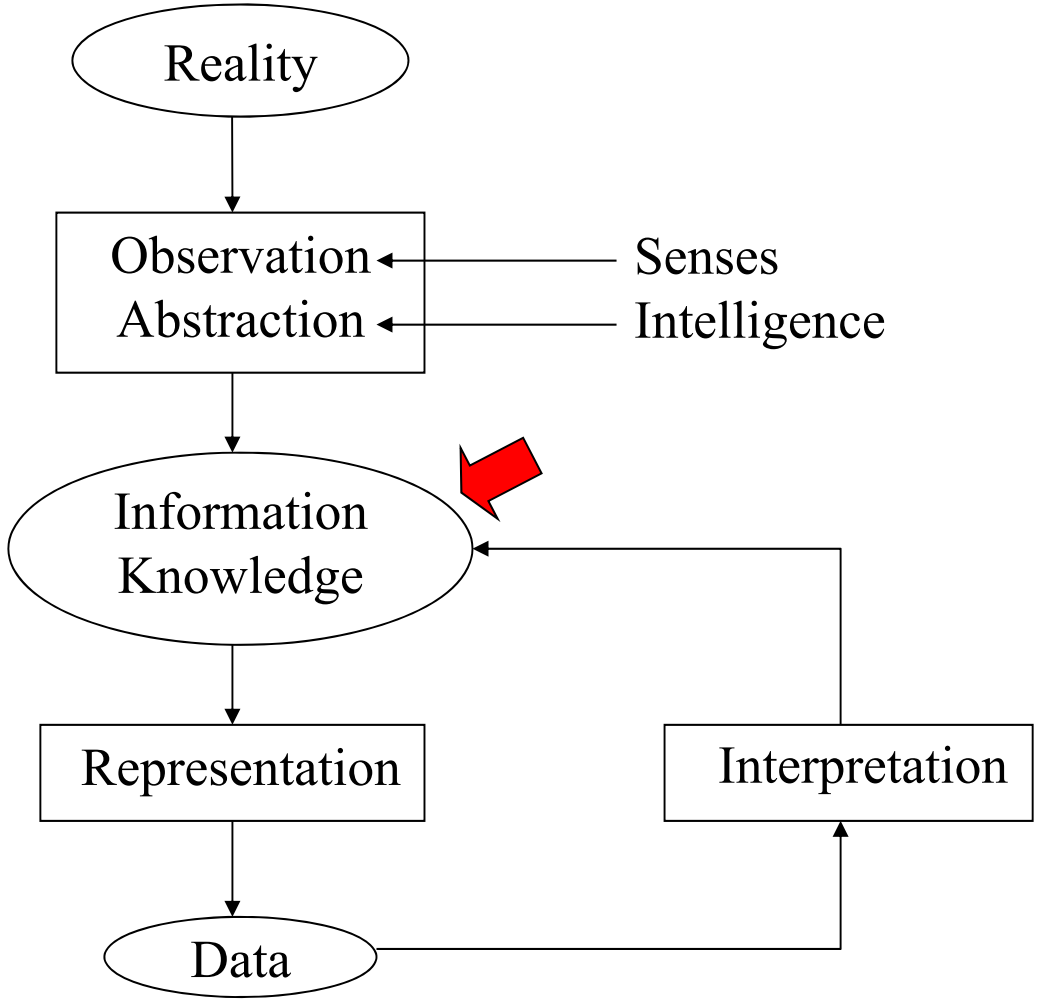
CONCEPTUAL WORLD VIEW

Three worlds

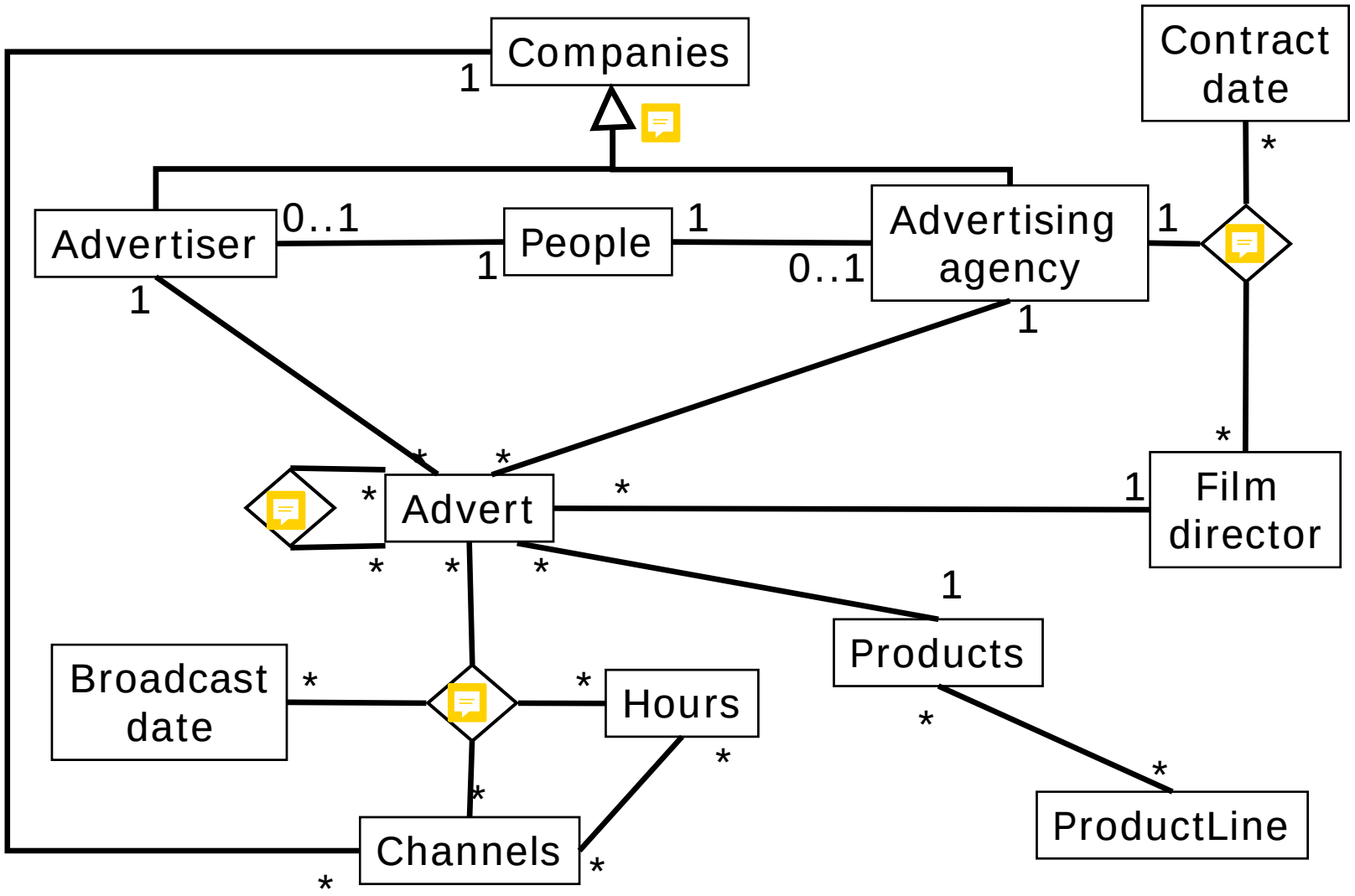
Real world
(unique)

Conceptual world
(multiple)

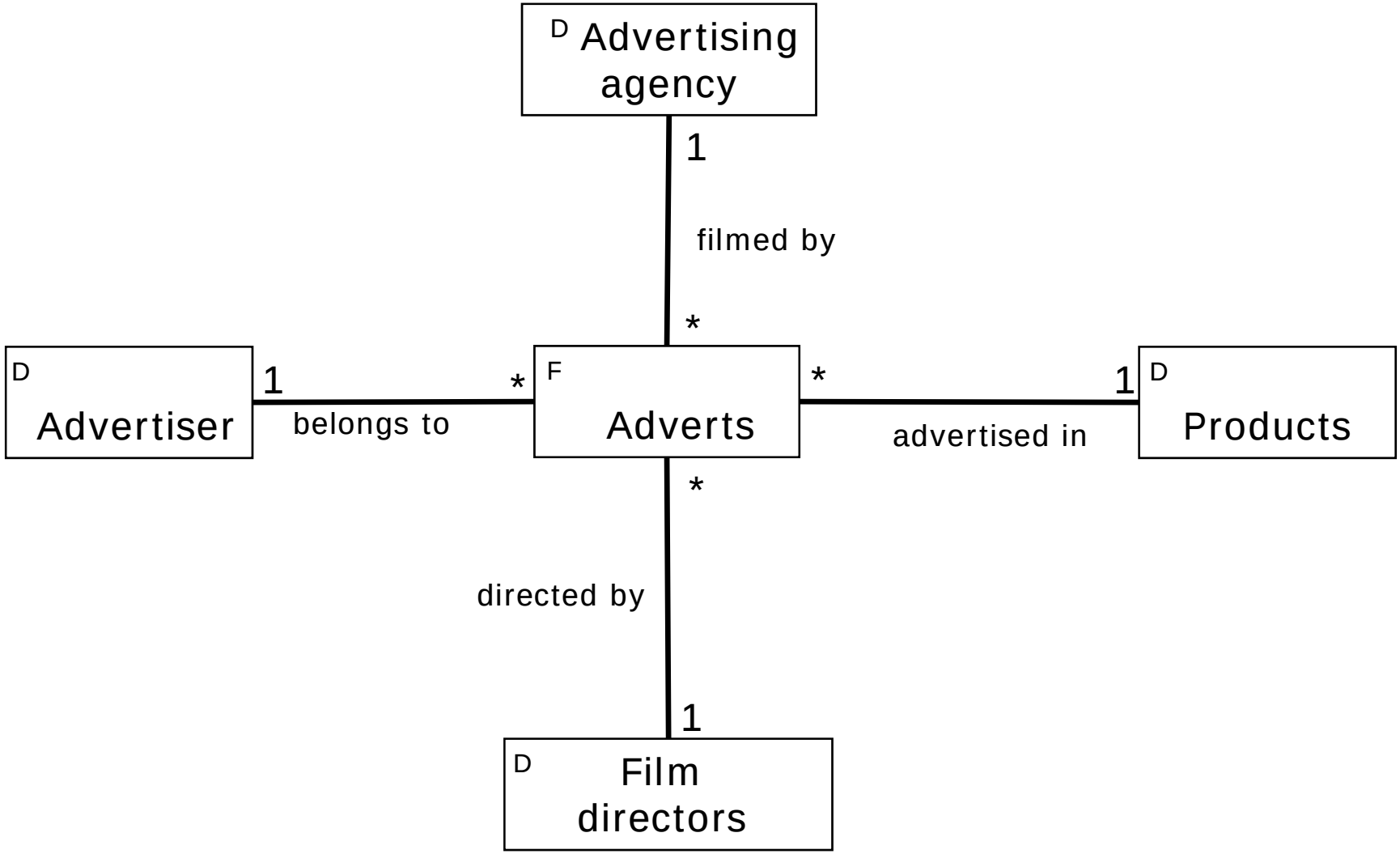
Representation world
(multiple)



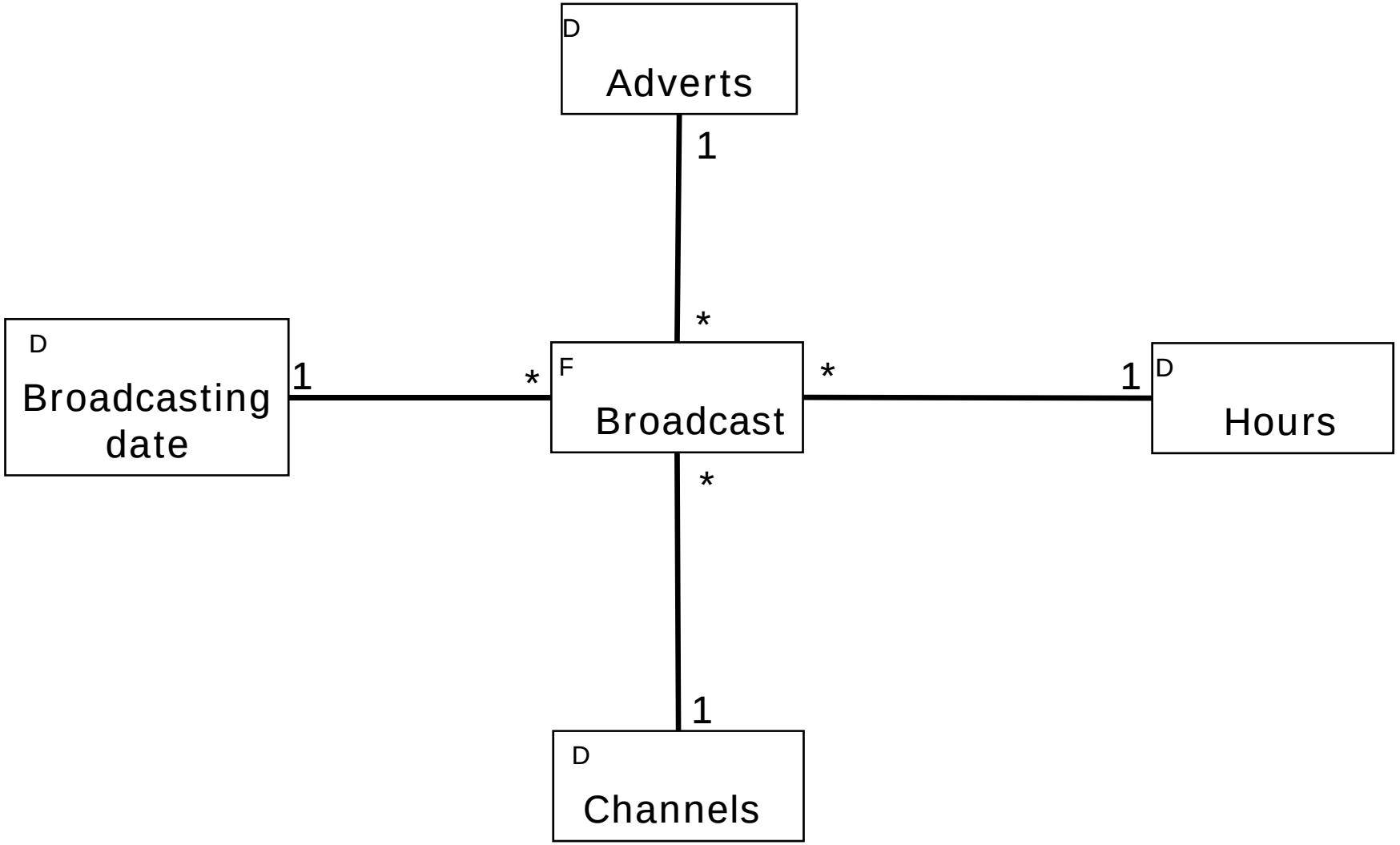
Example of transactional modeling



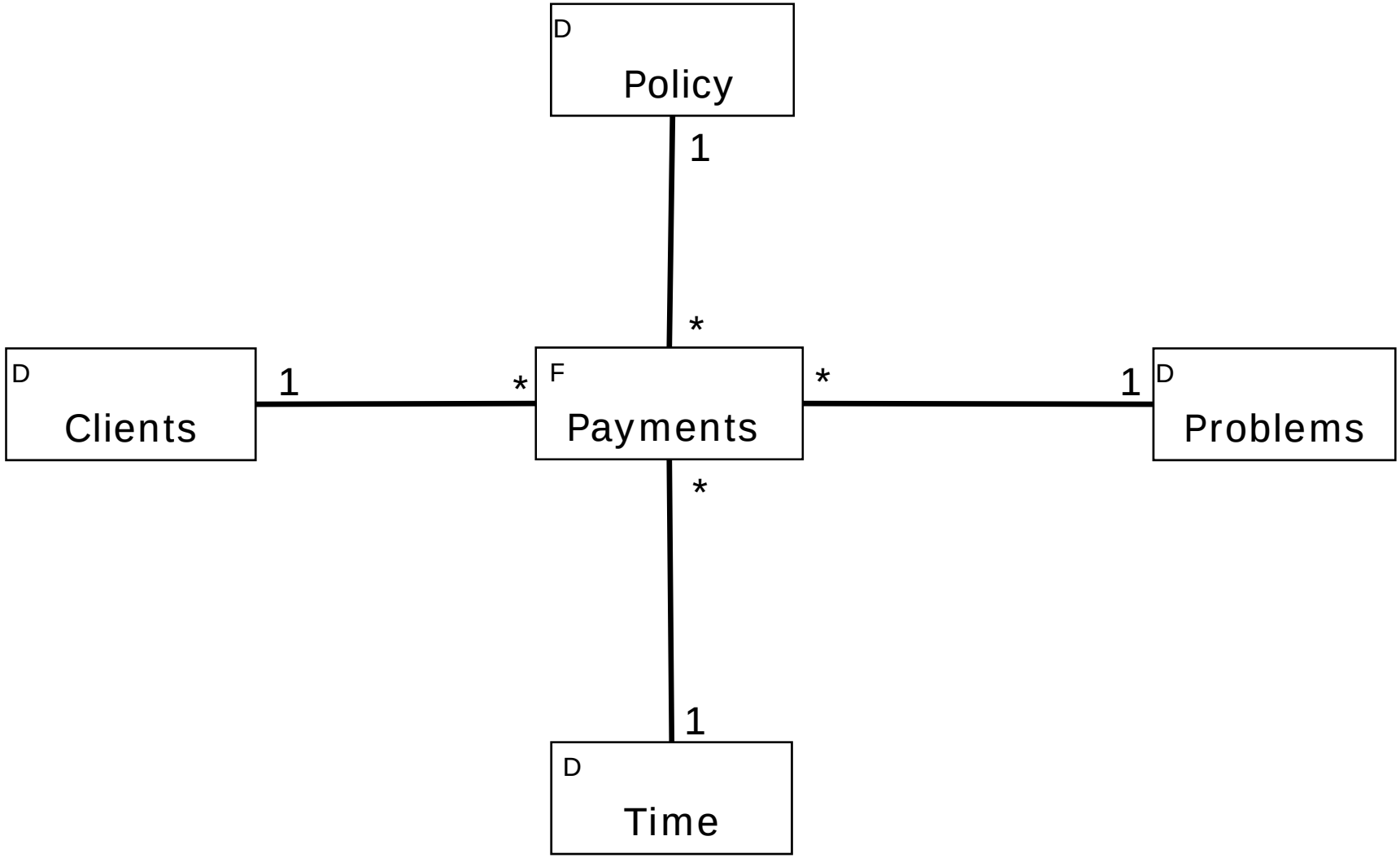
Star schema (I)



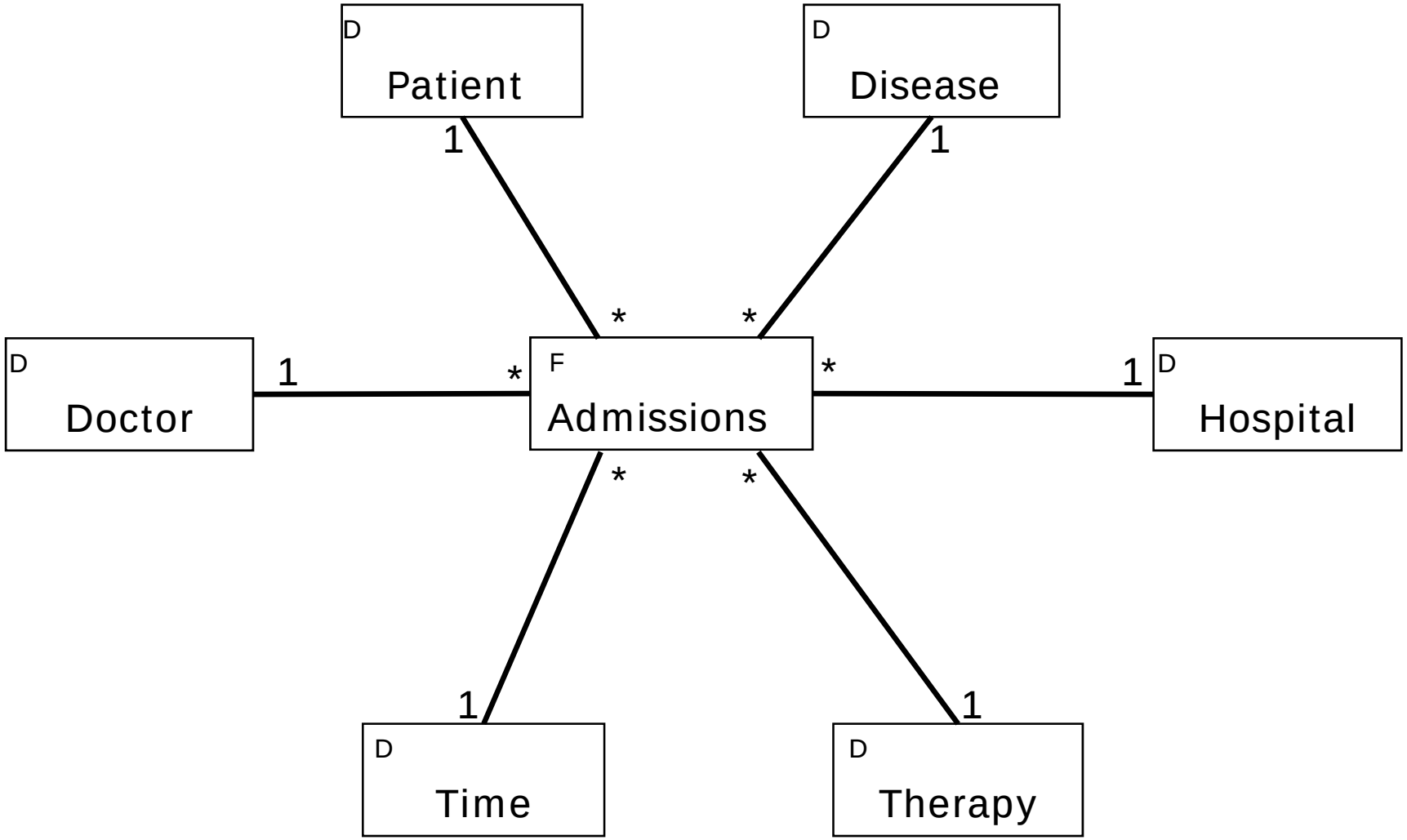
Star schema (II)



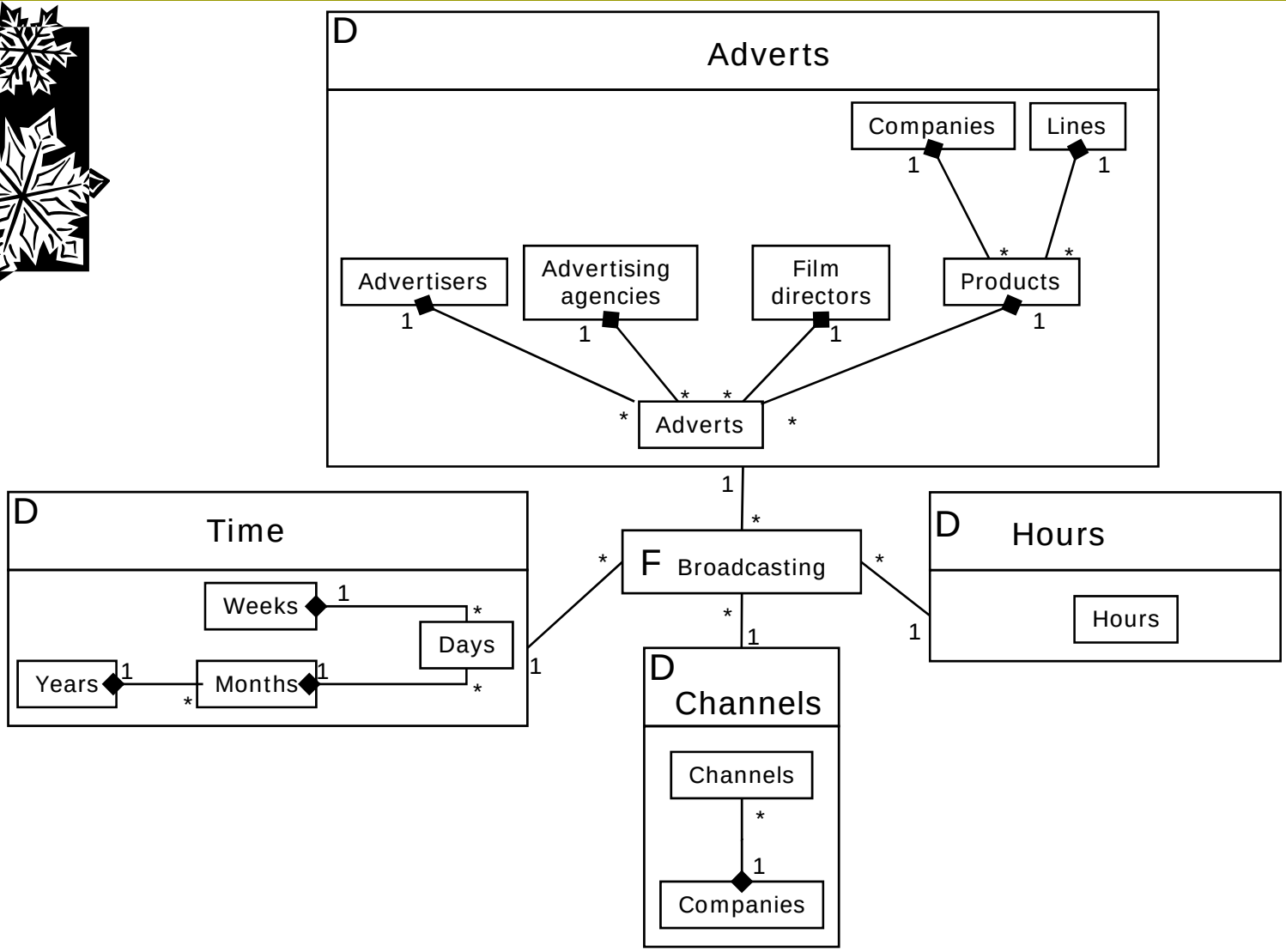
Star schema (III)



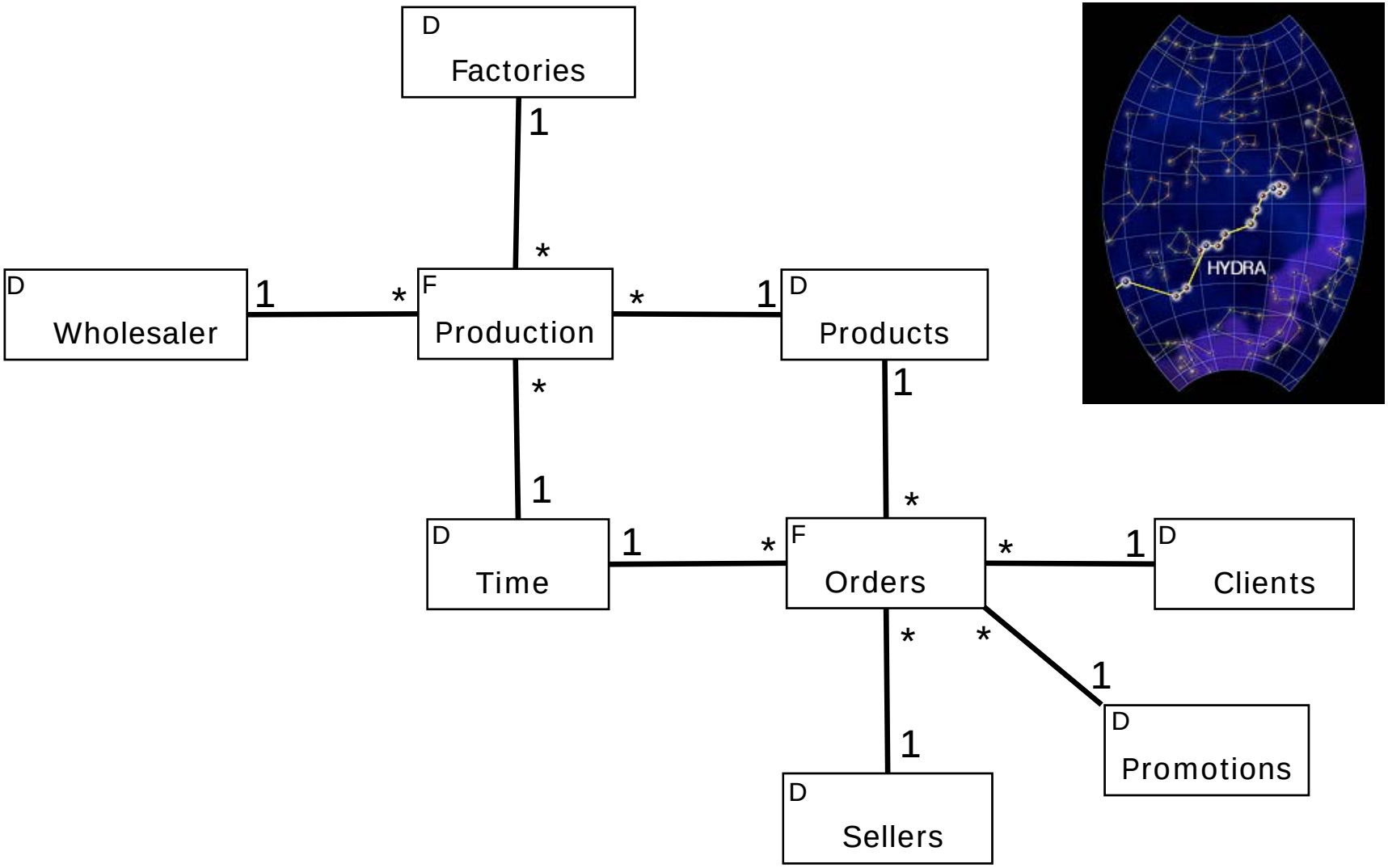
Star schema (IV)



Snowflake


D
Adverts
Companies
Lines
1
1
Advertisers
Advertising agencies
Film directors
Products
1
1
1
1
*
*
*
*
*
Adverts
*
1
*
D
Time
Weeks
Days
Years
Months
1
1
1
1
*
*
*
*
F
Broadcasting
*
*
*
1
1
D
Hours
Hours
1
D
Channels
Channels
*
1
Companies

Galaxy or Constellation



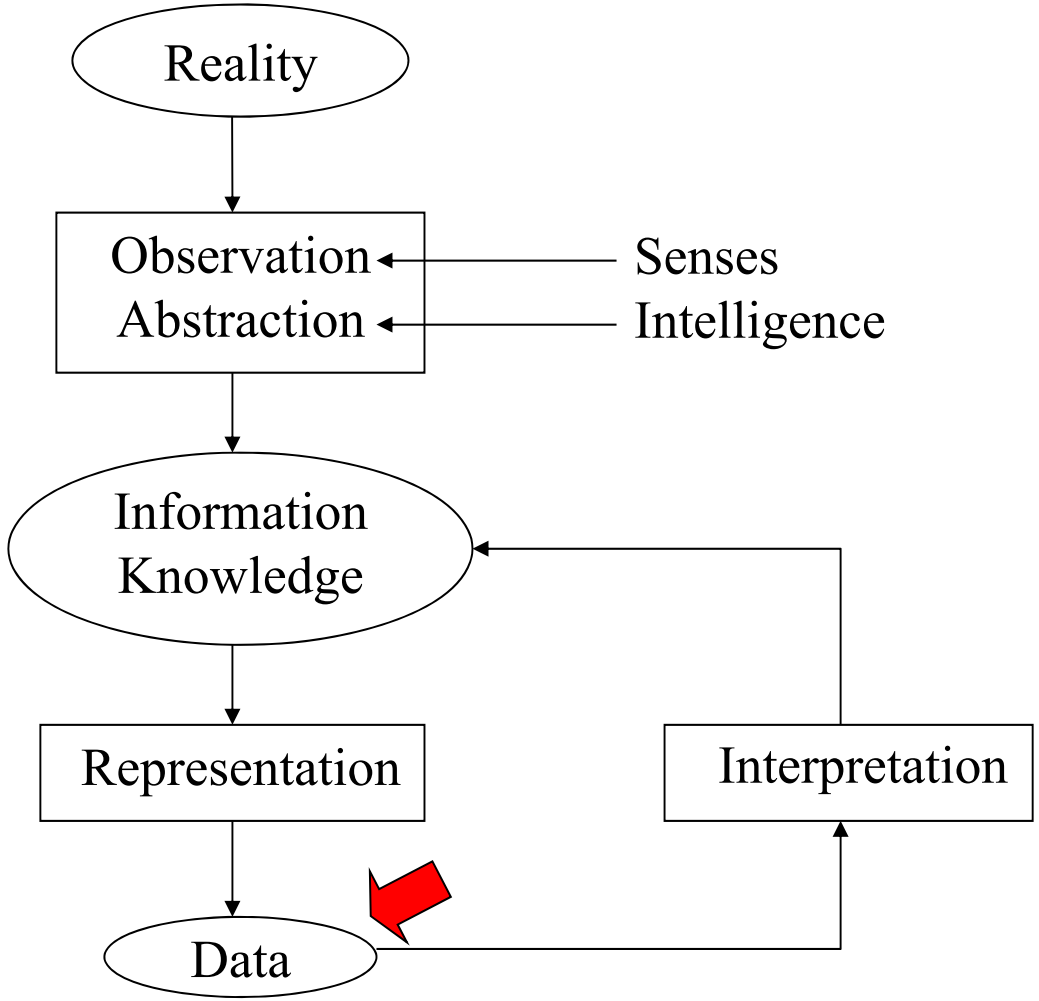
REPRESENTATION WORLD VIEW

Three worlds

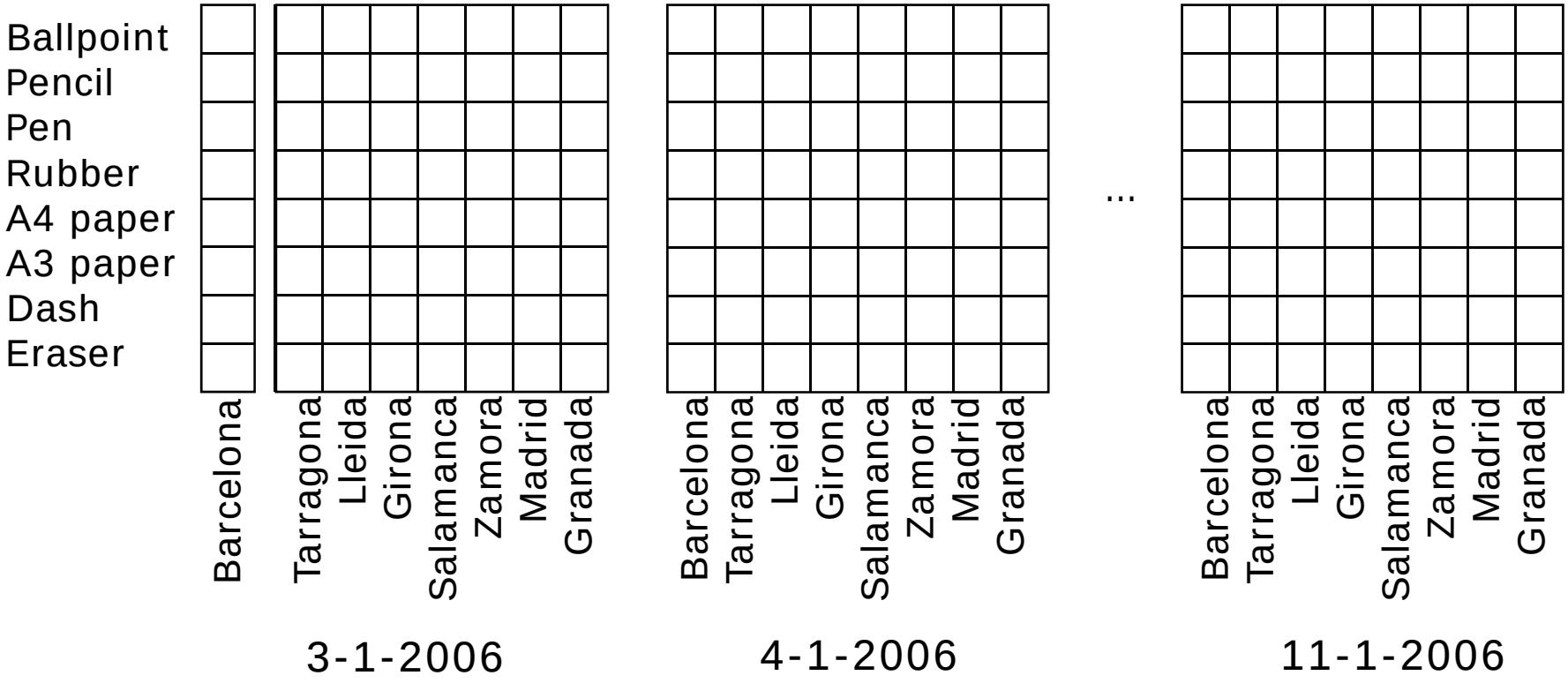
Real world
(unique)

Conceptual world
(multiple)

Representation world
(multiple)



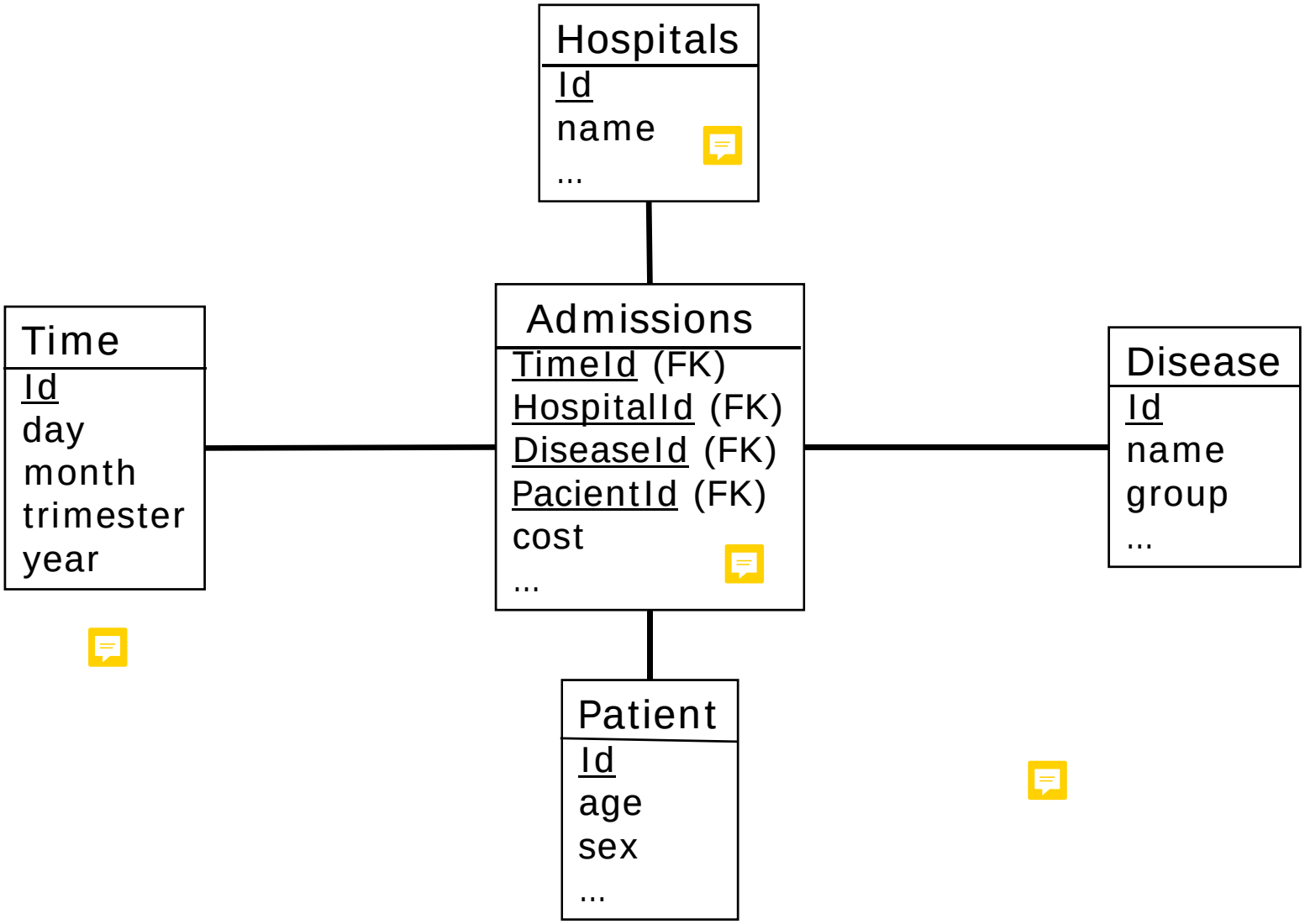
MOLAP: Multidimensional matrix



ROLAP: Characteristics

- Relational DBMS with multidimensional views
  - Two levels: Storage and Translation
- Use standard SQL
  - Easy to obtain
  - Independent of the DBMS
- Performance problems
  - Relational DBMS conceived for OLTP
  - OLAP operations are missing
  - Generate too many joins
- Used in huge Data Marts

ROLAP: Star-join schema



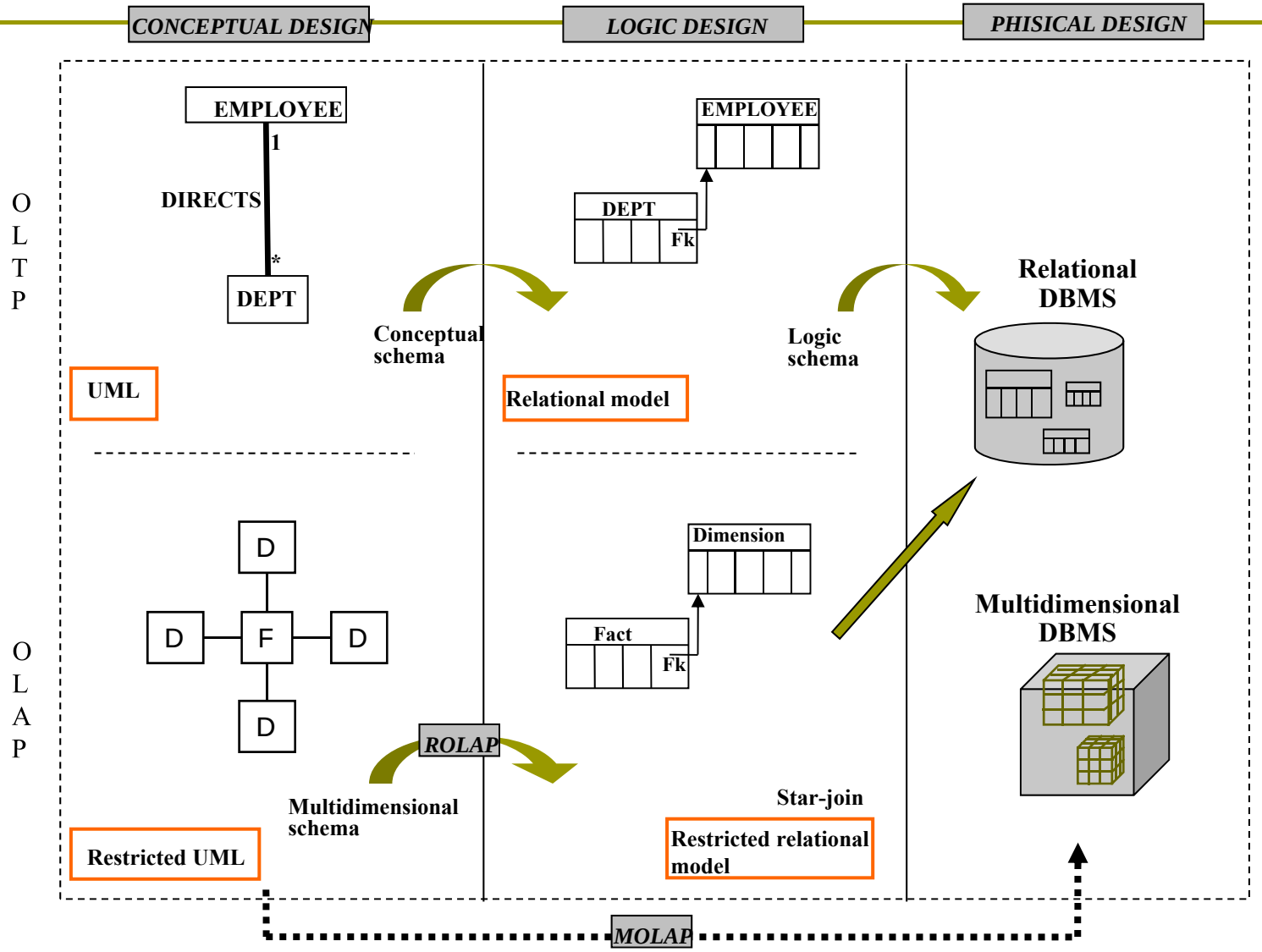
ROLAP: Cube-Query

SELECT $d_1.attr$, ..., $d_n.attr$, $F(f.Measure_1)$, ...
FROM Fact f, $Dimension_1\ d_1$, ..., $Dimension_n\ d_n$
WHERE $f.key_1 = d_1.ID$ *AND ... AND* $f.key_n = d_n.ID$
GROUP BY $d_1.attr$, ..., $d_n.attr$
ORDER BY $d_1.attr$, ..., $d_n.attr$

ROLAP: Results table

Hospital	Month	Average Cost
Duran i Reinals	January'06	3300
Duran i Reinals	February'06	4500
Duran i Reinals	...	...
Duran i Reinals	All	4300
Bellvitge	January'06	180
Bellvitge	February'06	300
Bellvitge	...	...
Bellvitge	All	200

Comparison of design steps



Reasons for ROLAP

- Integrates existing technology
- Does not show scalability problems
- Query tools are independent from the DBMS
- Improves efficiency by codifying and compressing
- Allows using parallelism
- MOLAP does not allow ad-hoc queries
- MOLAP makes data actualization difficult

Reasons for MOLAP

- Matrixes are really efficient
- Relational tables are unnatural
- Multidimensionality and SQL do not fit each other
- ROLAP gains efficiency with MOLAP techniques

HOLAP

a) Store dense chunks in MOLAP and sparse chunks in ROLAP

b) Store atomic data in ROLAP and aggregates in MOLAP

c) Store frequenly accessed data in MOLAP and the remaining data in ROLAP

CLOSING

Summary

- OLAP definition
- Cube
- Multidimensional schemas
  - Star
  - Snowflake
  - Galaxy or Constellation
- Kinds of multidimensional tools
  - ROLAP
  - MOLAP
  - HOLAP
- Star-join schema
- Cube-Query

Bibliography

- E. F. Codd, S. B. Codd and C. T. Salley. *Providing OLAP to user-analysts: An IT mandate.* Technical report, E. F. Codd & Associates, 1993
- R. Kimball, L. Reeves, M. Ross and W. Thornthwaite. *The Data Warehouse lifecycle toolkit*. John Wiley & Sons, 1998
- E. Malinowski and E. Zimányi. Advanced Data Warehouse Design. Springer, 2008
- A. Abelló and O. Romero. *On-Line Analytical Processing*. In Encyclopedia of Database Systems. Springer, 2009
- M. Golfarelli and S. Rizzi. *Data Warehouse Design*. McGrau-Hill, 2009
- BI-verdict. www.bi-verdict.com, June 2010
- A. Vaisman and E. Zimányi. Data Warehouse Systems. Springer, 2014