TDWI European Conference Amsterdam
13 November 2006
SAP Business Information Warehouse

Dr. Michael Hahne

Main Agenda

- Multidimensional Data Structures
- Conceptual Design
- Data Model of SAP Business Information Warehouse
- Modeling hierarchies in SAP BW
- Model Management
- Enterprise Data Warehousing with BW

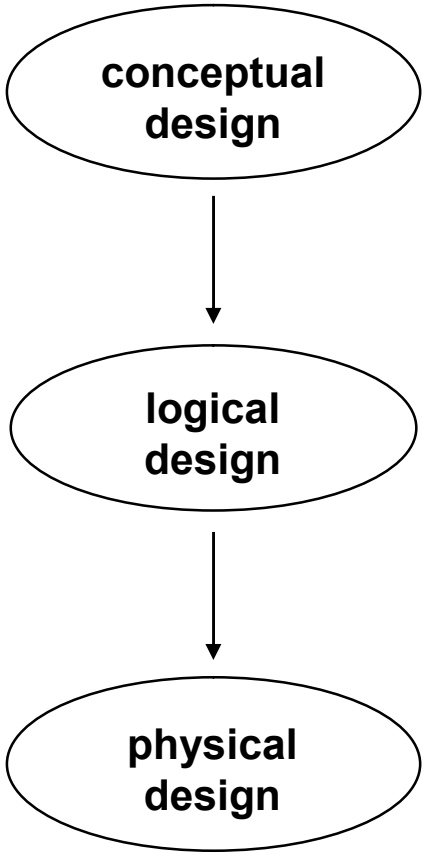
Agenda: Multidimensional Data Structures

- Steps of data modeling
- Elements of multidimensional data structures
- Hierarchical dimension structures
- Temporal aspects
- Guidelines for modeling

Agenda: Multidimensional Data Structures

- Steps of data modeling
- Elements of multidimensional data structures
- Hierarchical dimension structures
- Temporal aspects
- Guidelines for modeling

Steps of data modeling

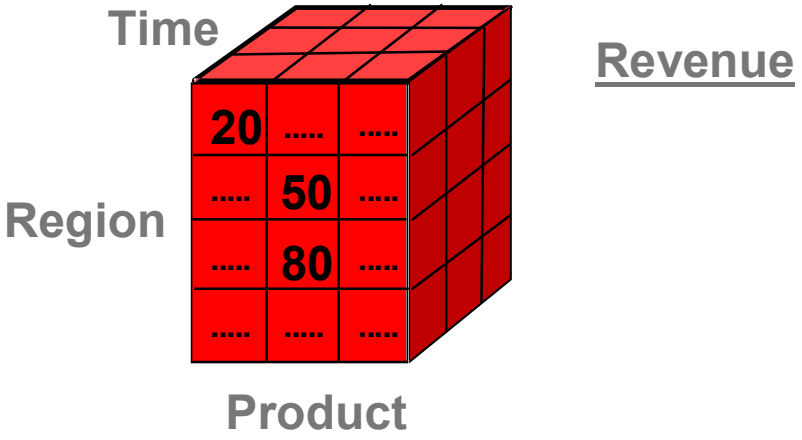


Agenda: Multidimensional Data Structures

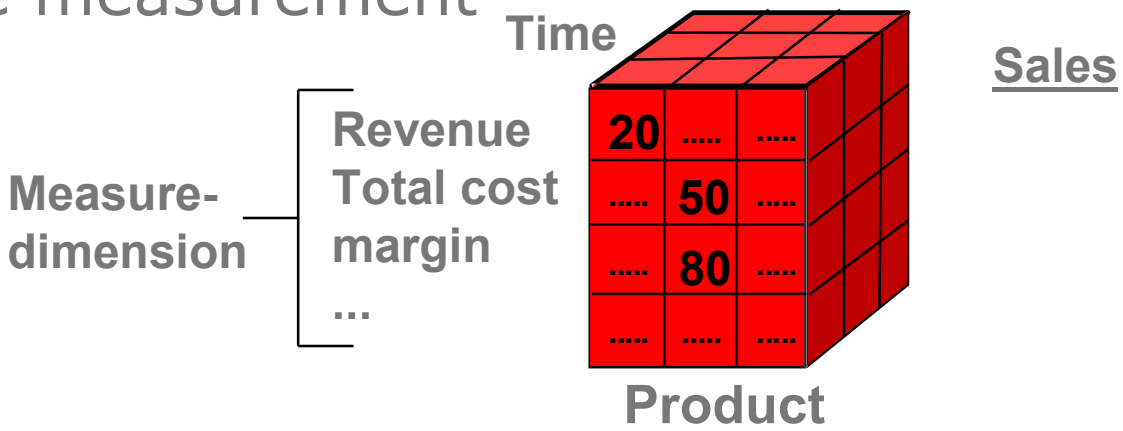
- Steps of data modeling
- Elements of multidimensional data structures
- Hierarchical dimension structures
- Temporal aspects
- Guidelines for modeling

Key performance indicators in cubes

- One measure per cube



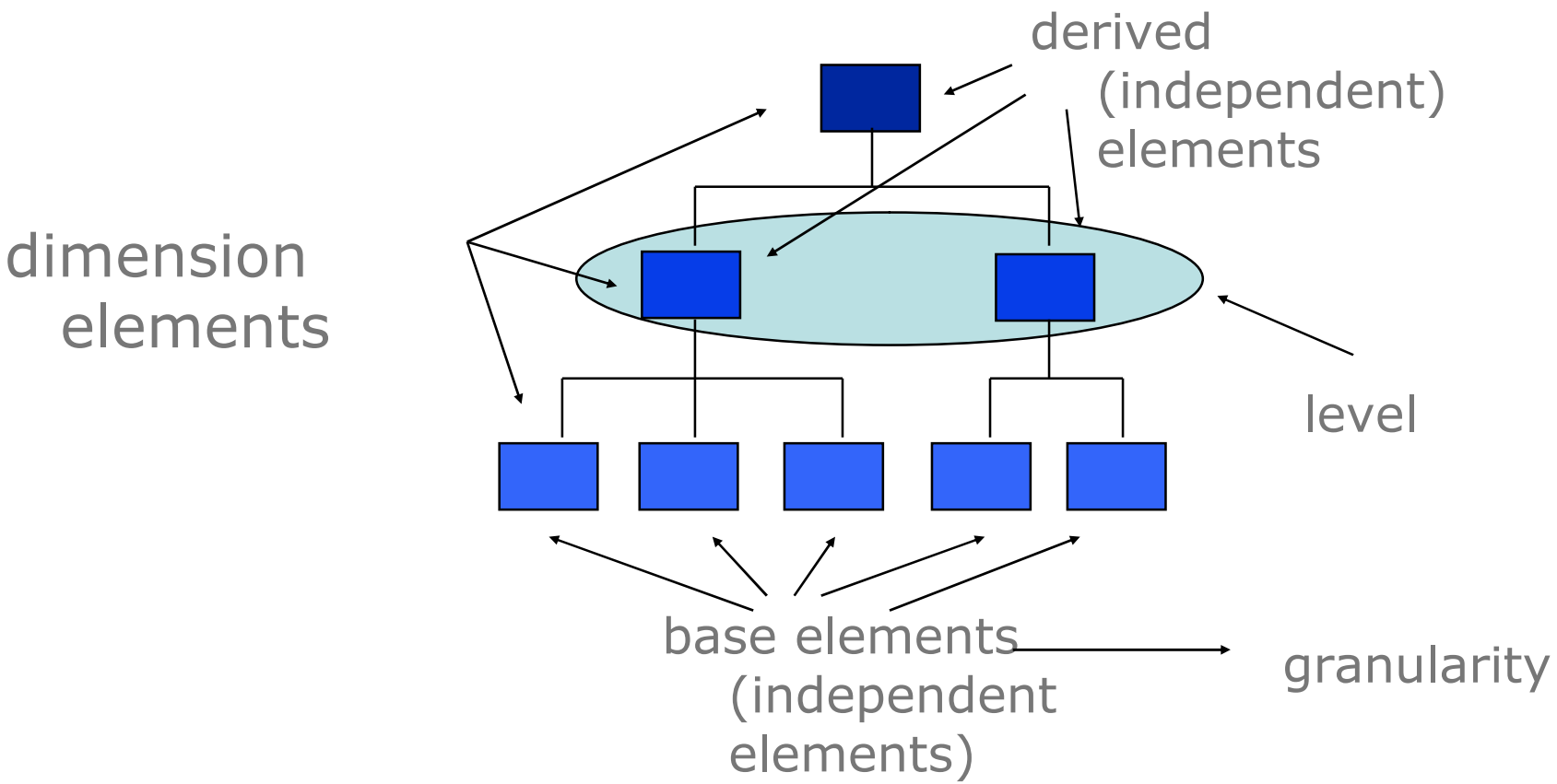
- Performance measurement system



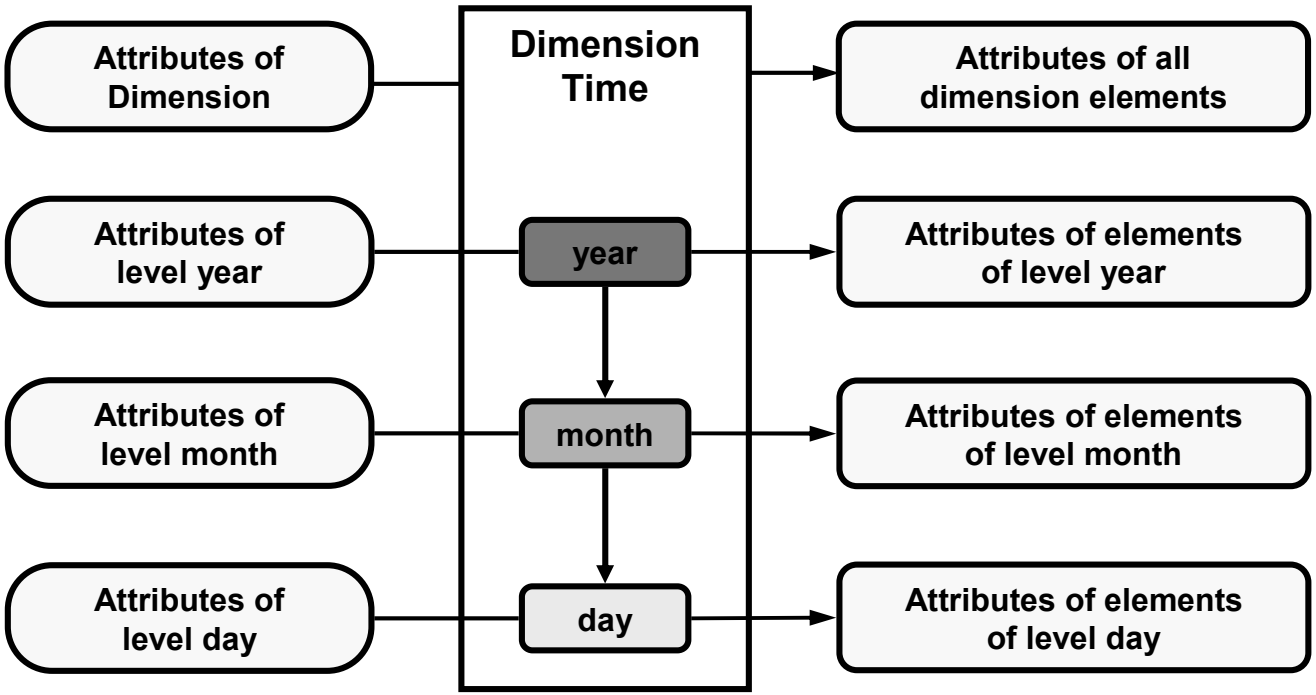
Dimensions

- Role:
  Concretizing and qualifying of measures
- Content:
  Similar objects relating to the same aspect of business
- Variants:
  - Untyped dimension (determined by elements)
    Few dimension elements (e.g. scenario or measure dimension)
  - Typed dimension (determined by levels)
    Huge dimensions (e.g. customer or material dimension)
    Characterization by set of objects

Constituent components of dimensions



Attributes


Attributes of Dimension
Dimension Time
Attributes of all dimension elements
Attributes of level year
year
Attributes of elements of level year
Attributes of level month
month
Attributes of elements of level month
Attributes of level day
day
Attributes of elements of level day

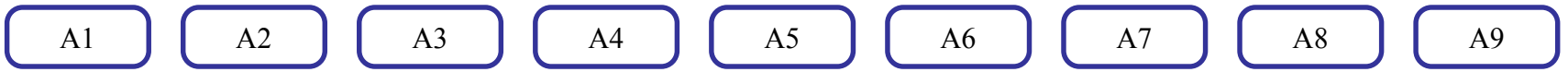
Types of dimensions

- “Standard” (e.g. Material, Customer, Region)
- Measure Dimension
- Scenario dimension (Actual, Budget, Forecast, ...)
- Time dimension

Agenda: Multidimensional Data Structures

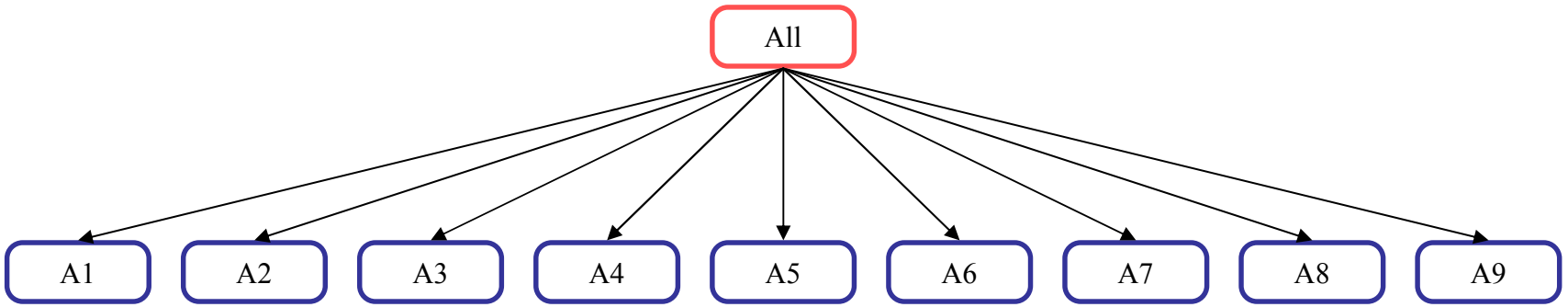
- Steps of data modeling
- Elements of multidimensional data structures
- Hierarchical dimension structures
- Temporal aspects
- Guidelines for modeling

Flat structure

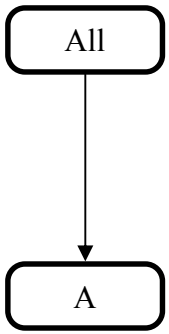


Typed flat structure

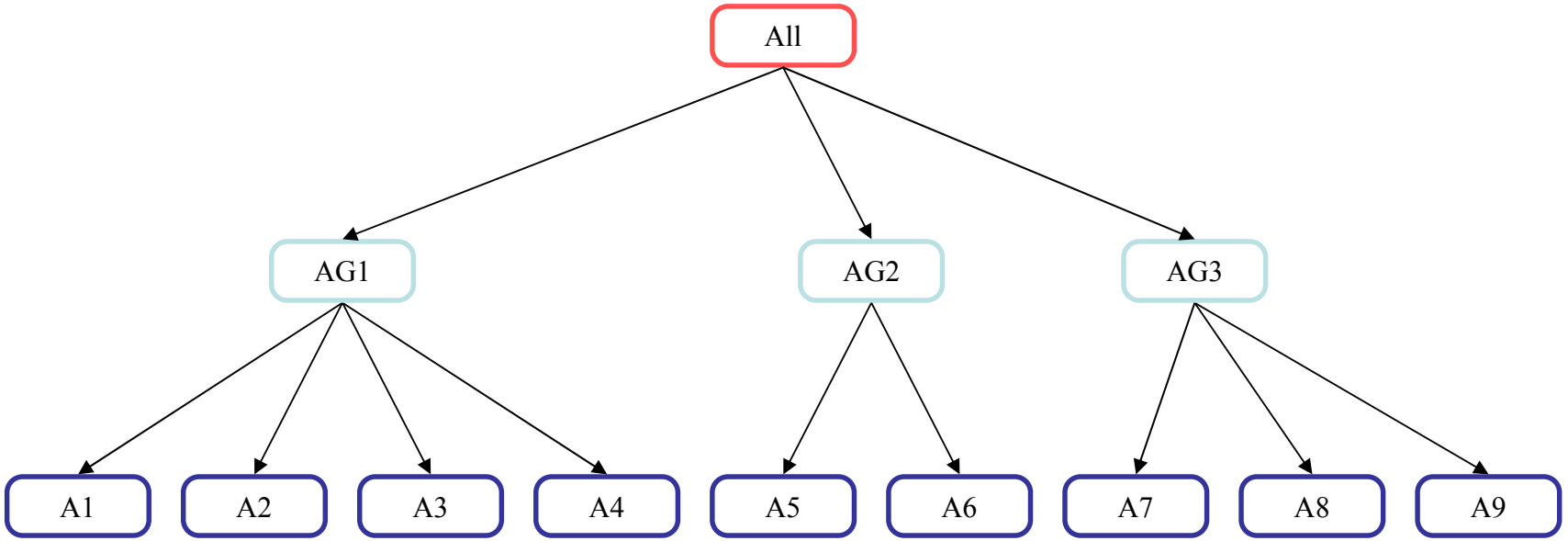
Flat structure with “total”



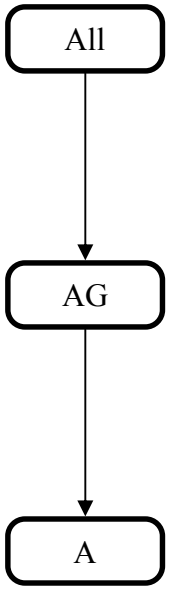
Typed flat structure with “total”



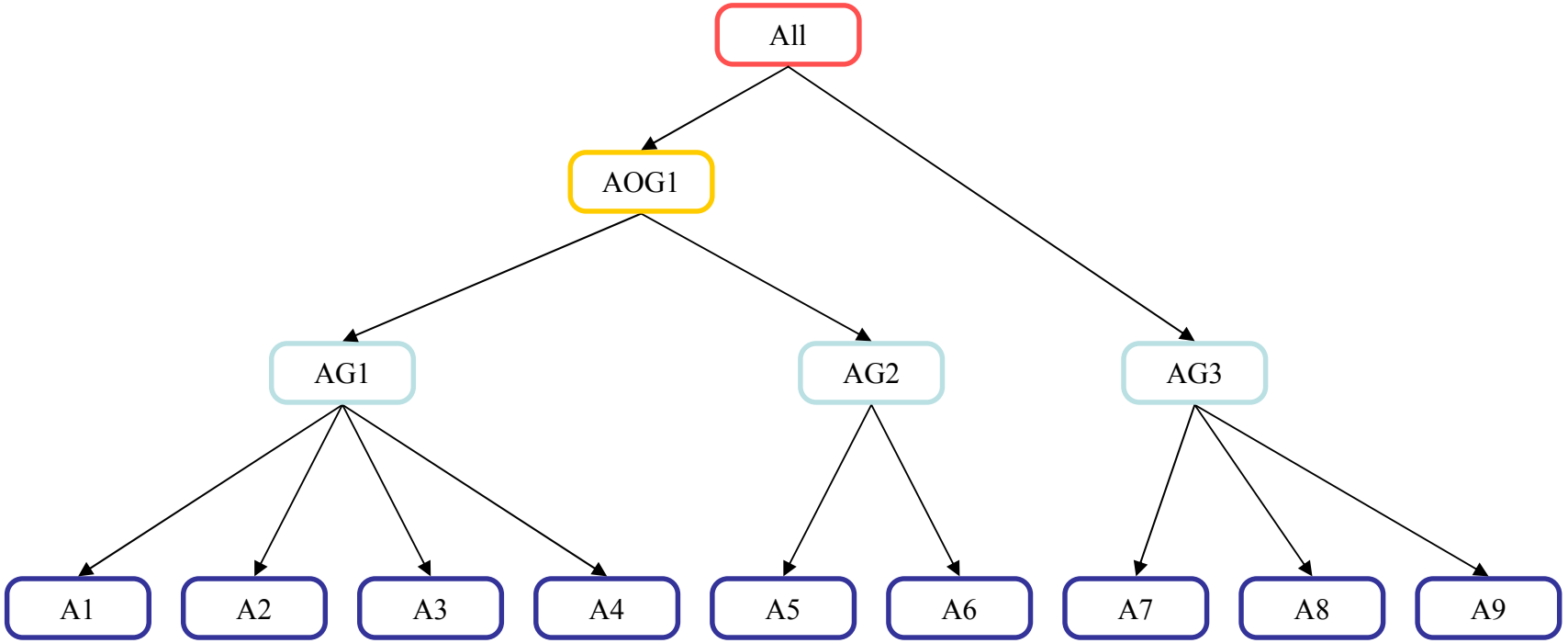
Tree structure



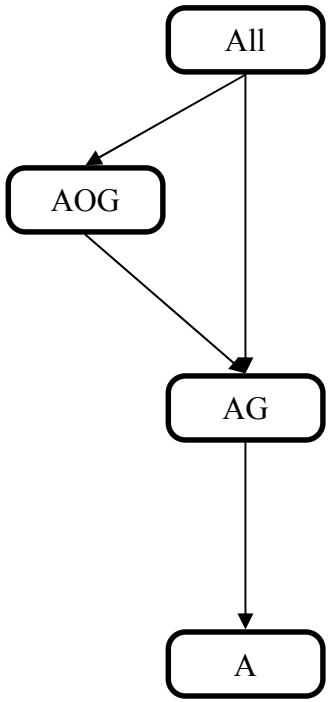
Typed tree structure



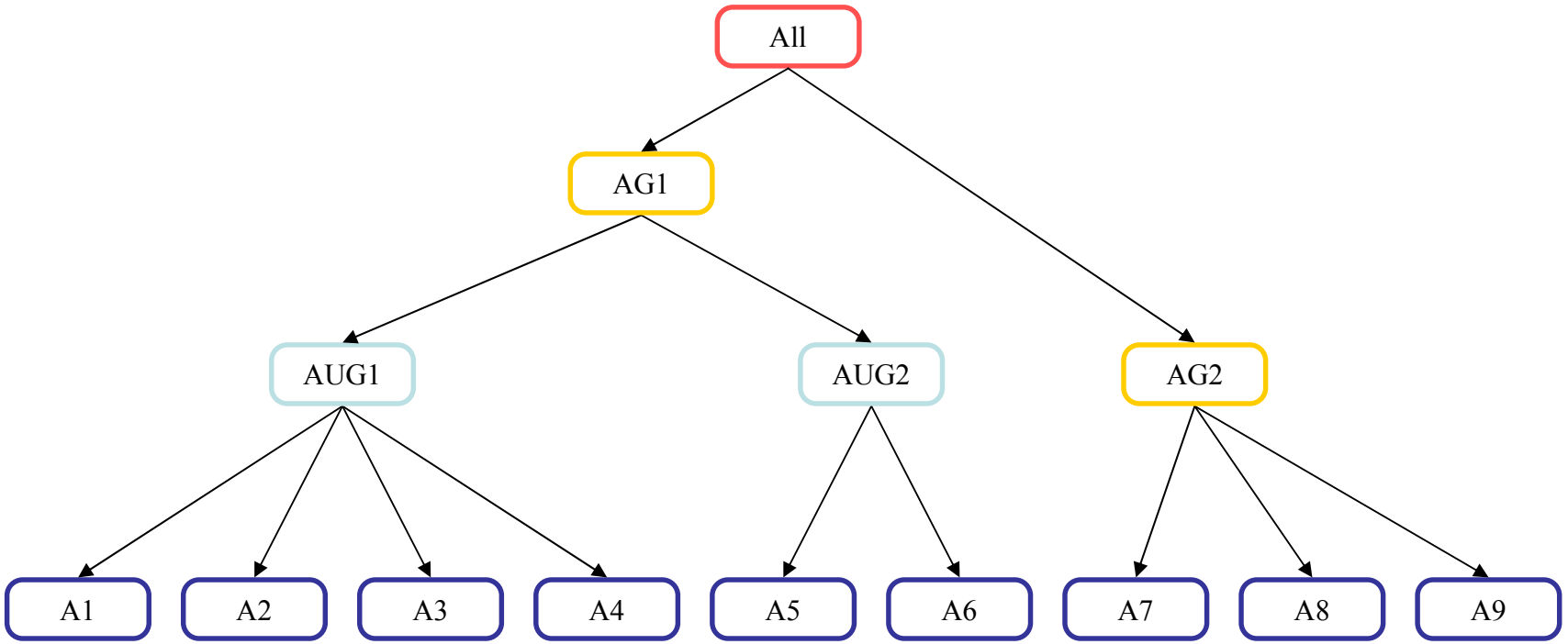
Unbalanced tree structure - variant I



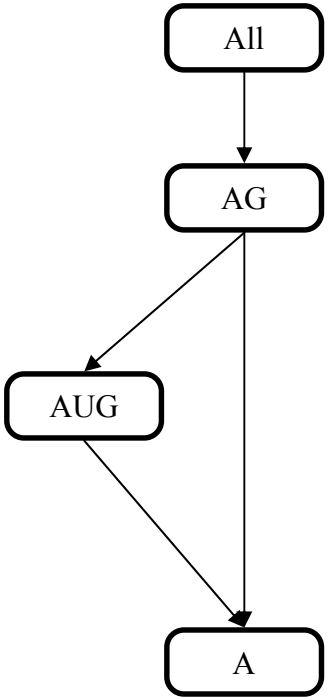
Typed unbalanced tree structure - variant I



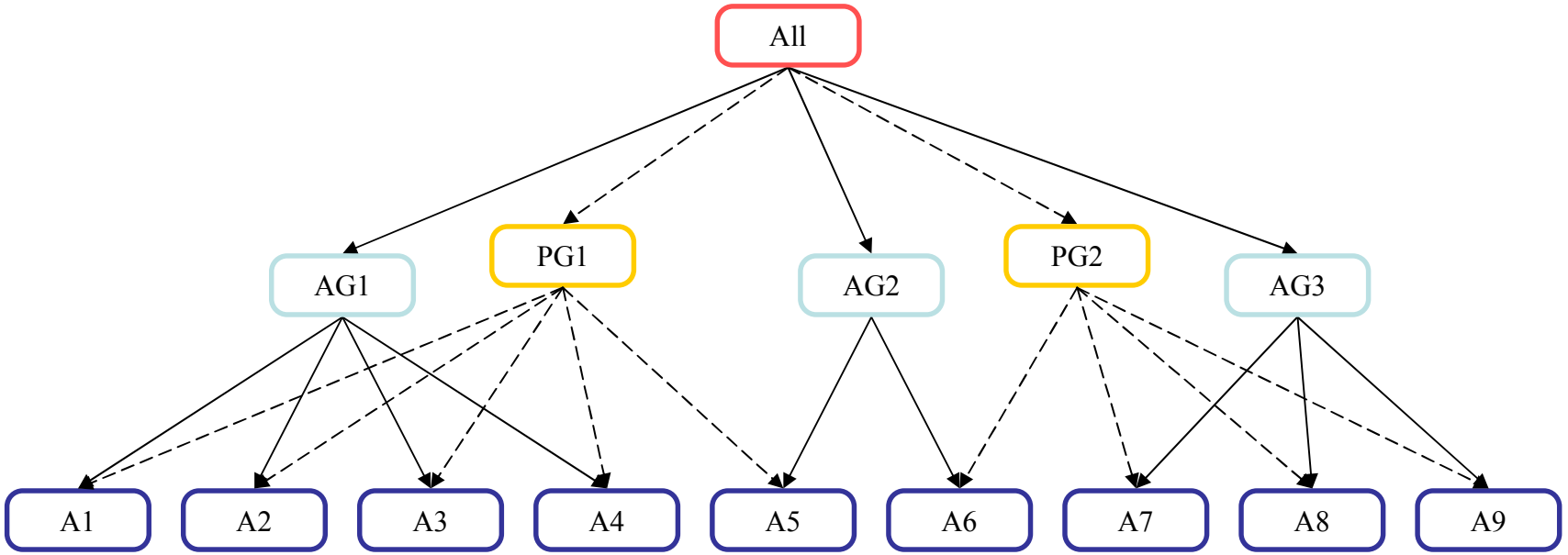
Unbalanced tree structure - variant II



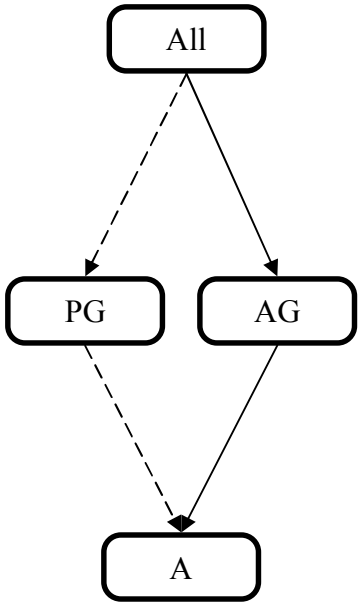
Typed unbalanced tree structure - variant II



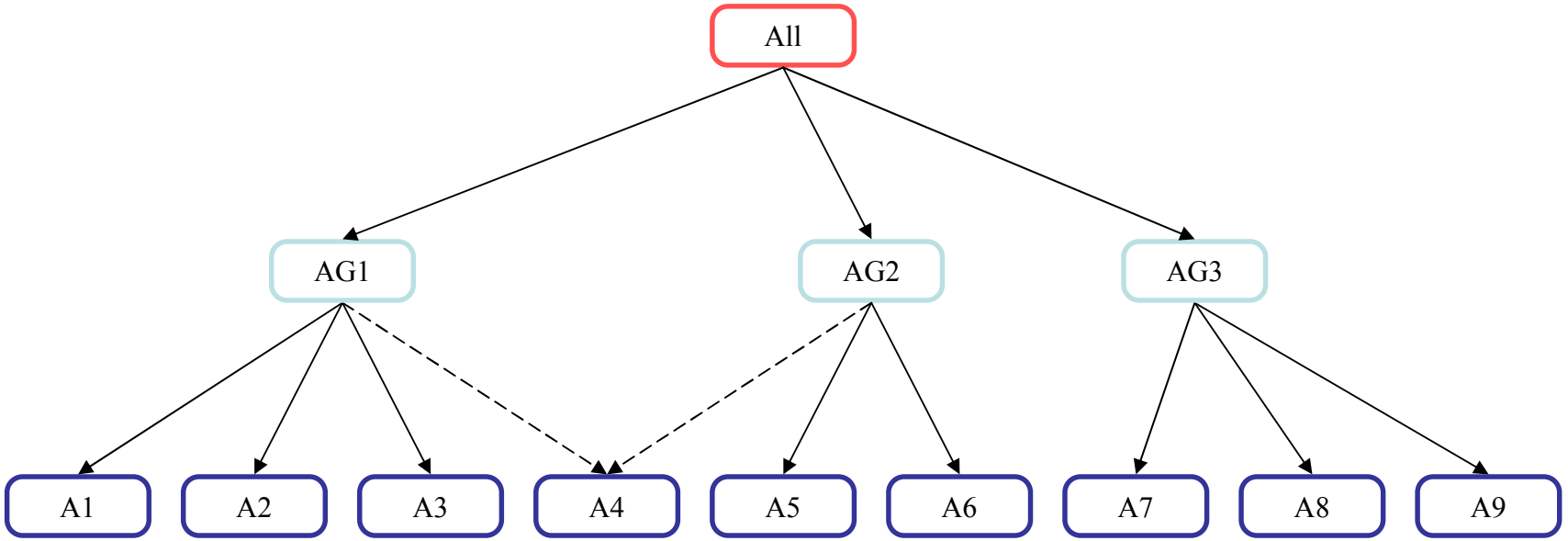
Parallel hierarchy



Typed parallel hierarchy

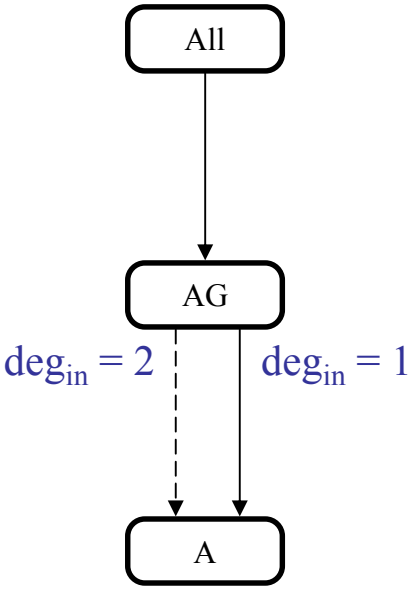


Heterarchy (many:many relationship)

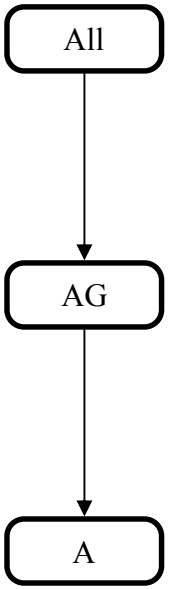


Typed heterarchy

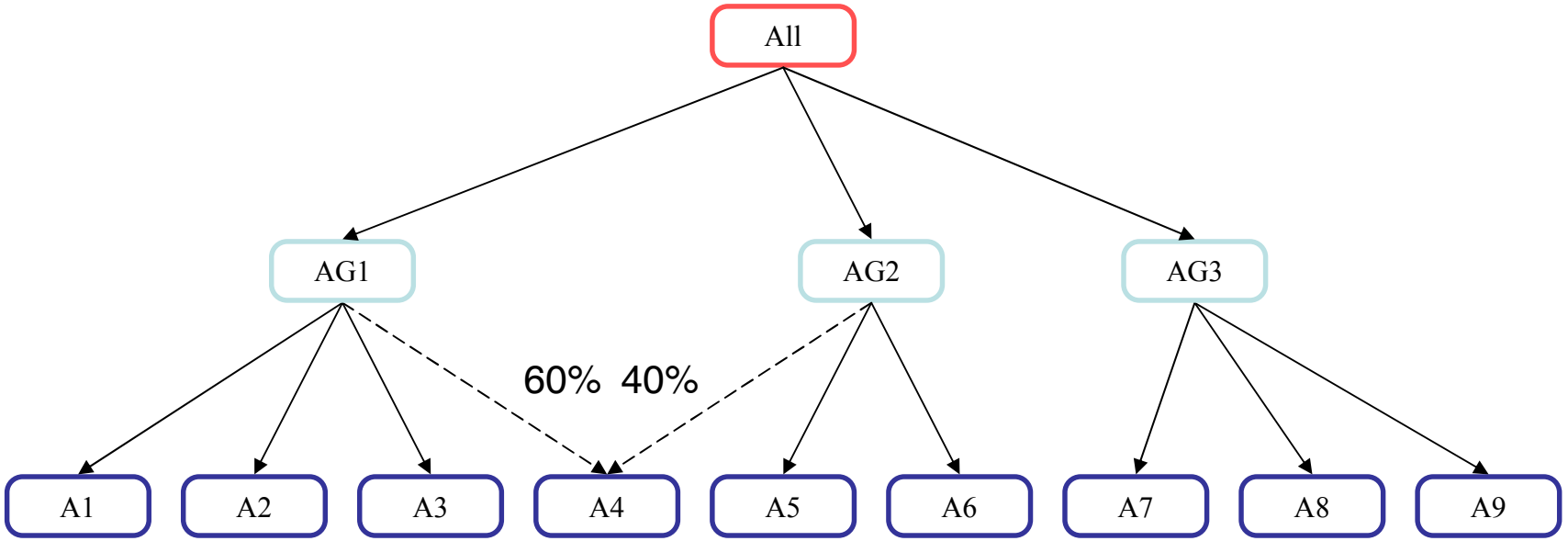
Distinguishing the "cardinality" of edges



Without distinguishing the "cardinality" of edges



Pro rata consolidation



Additivity

Aggregation in the case of nonadditive facts

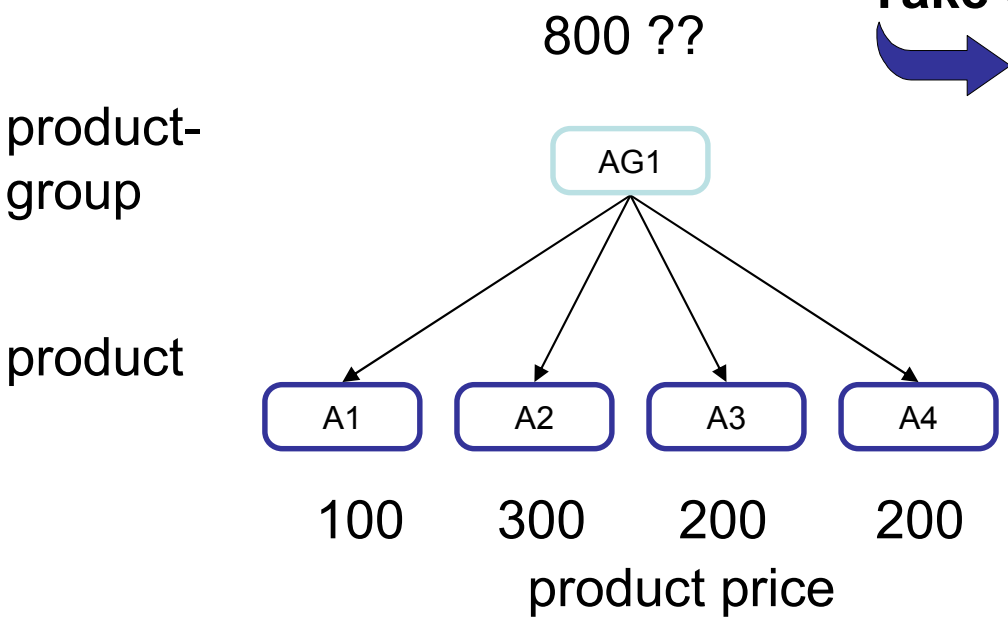
e.g. price, stock of inventory

May be reasonable:

calculation of average or last value

Take care:

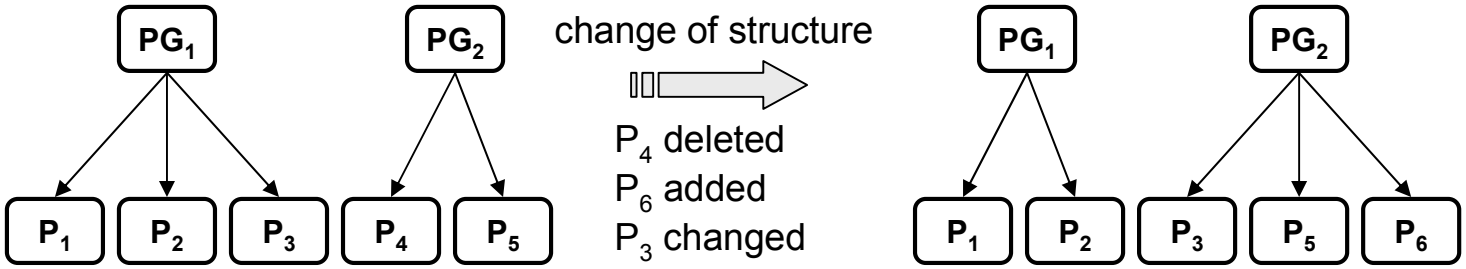
Semiadditive facts like closing stocks can be added along the storage location dimension but not along the time dimension



Agenda: Multidimensional Data Structures

- Steps of data modeling
- Elements of multidimensional data structures
- Hierarchical dimension structures
- Temporal aspects
- Guidelines for modeling

Slowly changing dimensions



Possible reporting requirements

- As is: report with current structure
- As of: report with the structure as it was on a fixed date (point in time)
- As posted: report showing the “historical truth” (transaction-oriented view)
- Report with comparable results (time slice)

Slowly changing dimensions - example

Product dimension as of 2005-04

Product	Productgroup
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Product dimension as of 2005-05

Product	Productgroup
P A	PG X
P B	PG Y (changed)
P C	PG Y
P D	PG Y
P E	PG Y (new)

fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Reporting requirements - example

Reporting current structure

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	100	100
PG Y	300	400

Reporting as of (2005-04)

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	200	200
PG Y	200	200

Reporting as posted

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	200	100
PG Y	200	400

Reporting comparable results

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	100	100
PG Y	200	200

As is: Report with the current structure

Product dimension as of 2005-05

Product	Productgroup
P A	PG X
P B	PG Y (changed)
P C	PG Y
P D	PG Y
P E	PG Y (new)

fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	100	100
PG Y	300	400

As Of:
Report with the structure as it was on a fixed date

Product dimension as of 2005-04

Product	Productgroup
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	200	200
PG Y	200	200

As posted: Report the "historical truth"

Product dimension as of 2005-04

Product	Productgroup
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Product dimension as of 2005-05

Product	Productgroup
P A	PG X
P B	PG Y (changed)
P C	PG Y
P D	PG Y
P E	PG Y (new)

fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	200	100
PG Y	200	400

Report with comparable results

Product dimension as of 2005-04

Product	Productgroup
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Product dimension as of 2005-05

Product	Productgroup
P A	PG X
P B	PG Y (changed)
P C	PG Y
P D	PG Y
P E	PG Y (new)

fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	100	100
PG Y	200	200

Solutions for structural changes in dimensions

- **Modifying historical data to match new structures**
  Pros: Small data sets; data structures remain clear
  Cons: Old structures are lost even if users want to analyze data with old structures
- **Save historical data separately in addition to all data matched to new structures**
  Pros : Old reports can be reproduced
  Cons : Higher volume of data; update processes are complicated and costly if new data needs to be analyzed according to old structures; confusing for end-users
- **Build parallel hierarchies with old and new structures**
  Pros : All data can be analyzed with any structure
  Cons : Dimension structure is very confusing
- **Temporal databases - timestamps**
  Pros : All data can be analyzed with any structure
  Cons : Poor performance; concepts not fully developed

Agenda: Multidimensional Data Structures

- Steps of data modeling
- Elements of multidimensional data structures
- Hierarchical dimension structures
- Temporal aspects
- Guidelines for modeling

Guidelines for conceptual design of dimensions

- Number of dimensions in a cube should be between four and ten (best is between six and eight)
- Number of levels should be small (a maximum of seven is good)
- Not too many children per consolidation element (a maximum of fifteen to twenty)
- Objects that define dimensions:
  - 1:1 relationship unsuitable
  - 1:many relationship most suitable for dimension hierarchy
  - many:many relationship should be modeled as two dimensions

Agenda: Conceptual Design

- Methods for conceptual design
- Conceptual modeling with ADAPT
- Tool support by Microsoft Visio™
- Tool support by SemTalk™

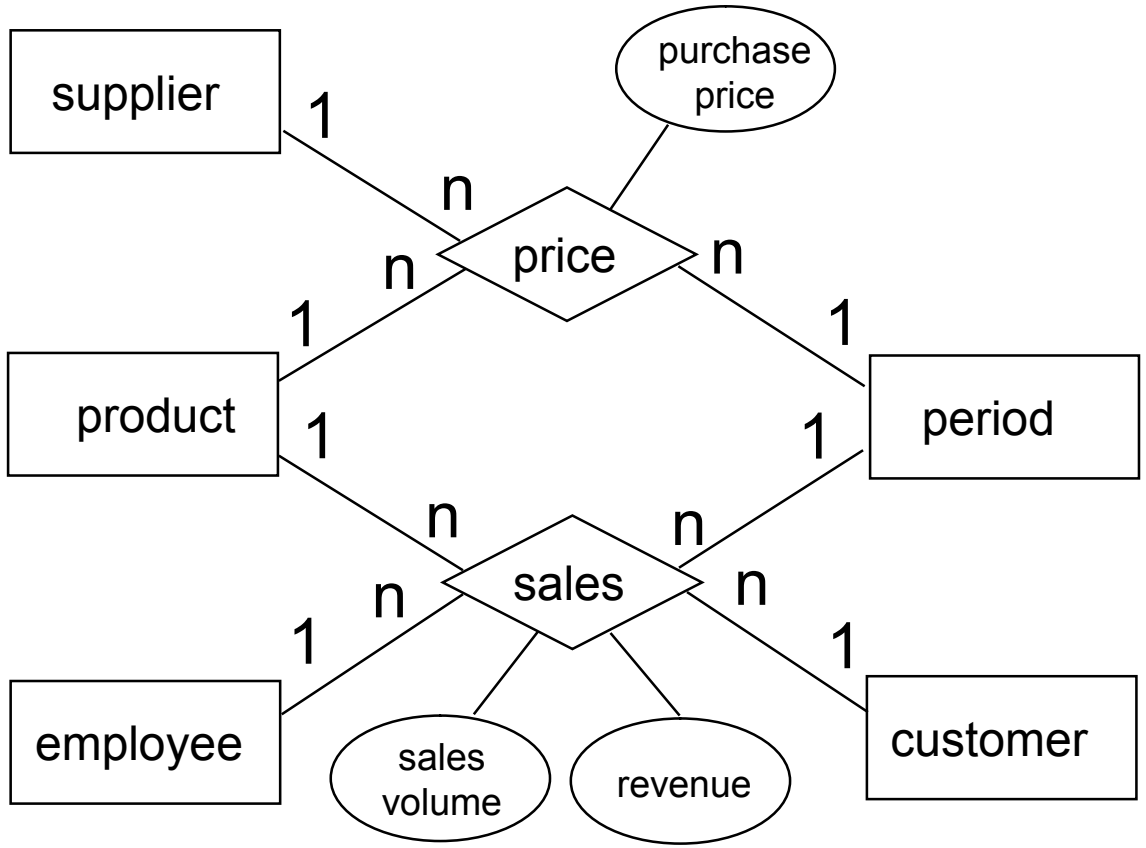
Agenda: Conceptual Design

- Methods for conceptual design
- Conceptual modelling with ADAPT
- Tool support by Microsoft Visio™
- Tool support by SemTalk™

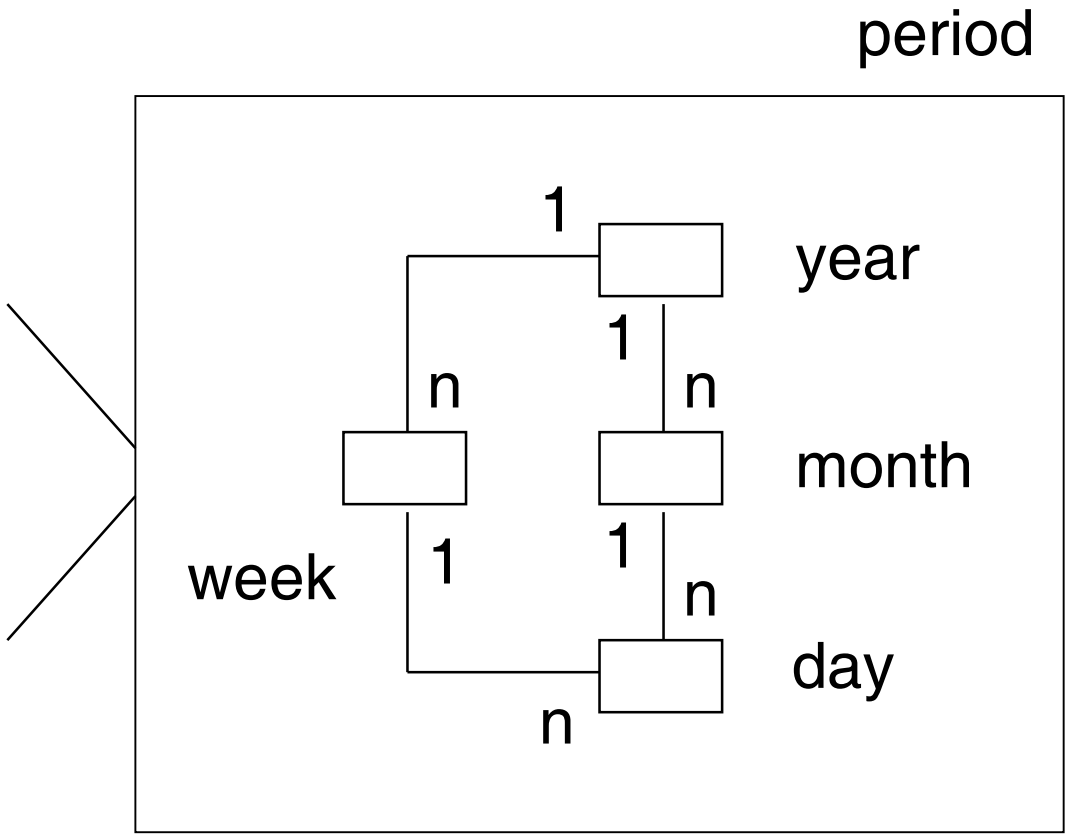
Conceptual modeling approaches

- Entity Relationship Model
- Multidimensional Entity Relationship Model (ME/R–Model)
- Dimensional Fact Modeling (Golfarelli)
- Dimensional Modeling (Kimball) / Star Schema
- Multidimensional Data Model (Cabibbo / Torlone)

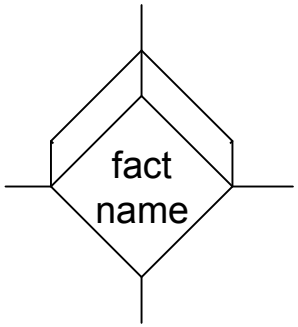
Entity Relationship Model



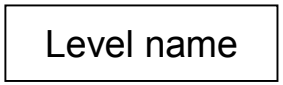
Extension eER for modeling hierarchies



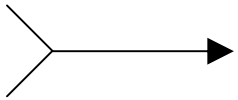
Multidimensional Entity Relationship Model (ME/R-Model)



Fact Relationship Type



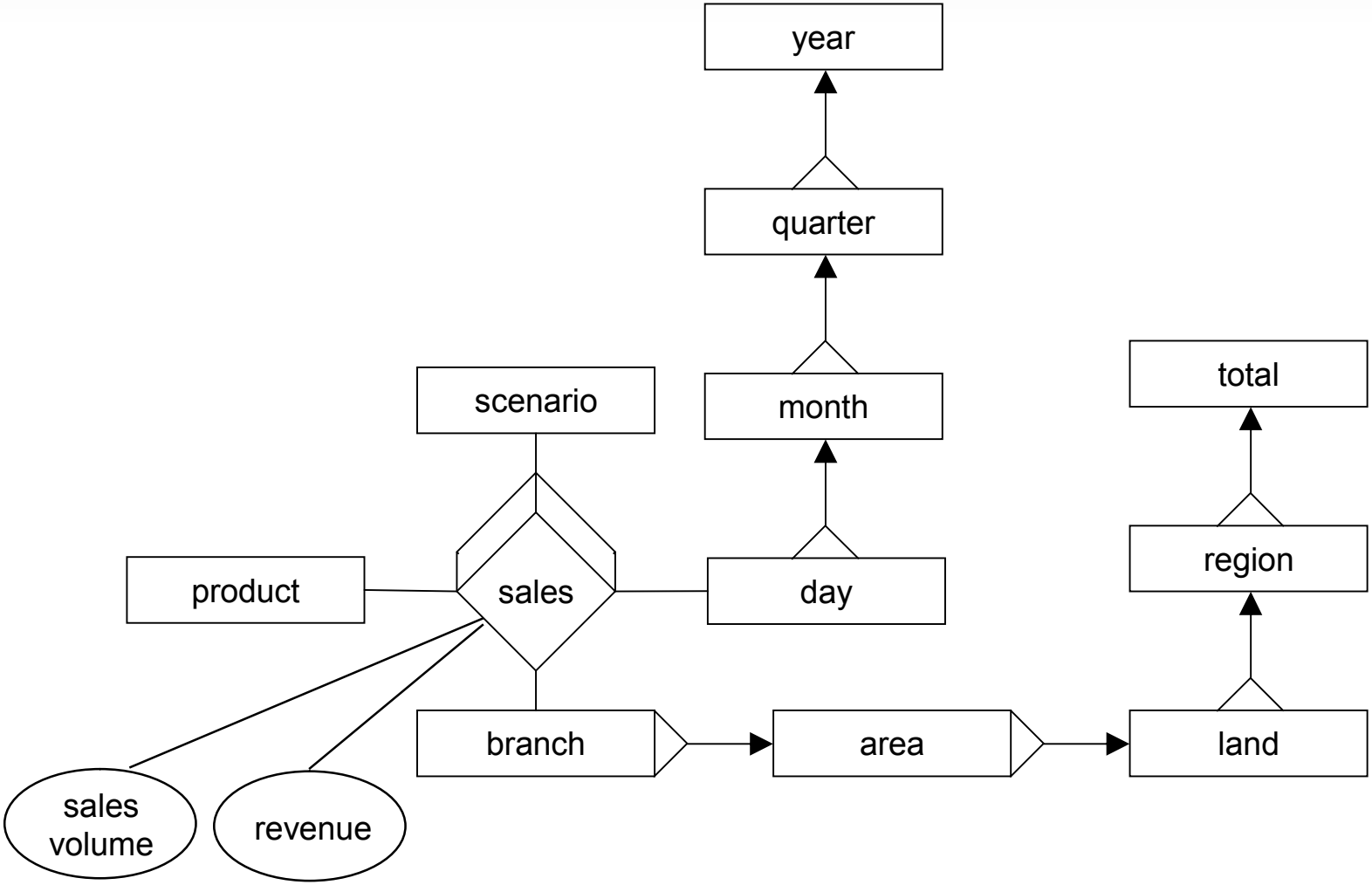
Dimension Level Entity Type



Roll-Up Relationship Type

reference: Sapia, Carsten; Blaschka, Markus; Höfling, Gabriele;, Dinter, Barbara: Extending the E/R Model for the Multidimensional Paradigm.
In: Proceedings of the Advances in Databases Technologies Conference, November 19-20, 1998, Singapore

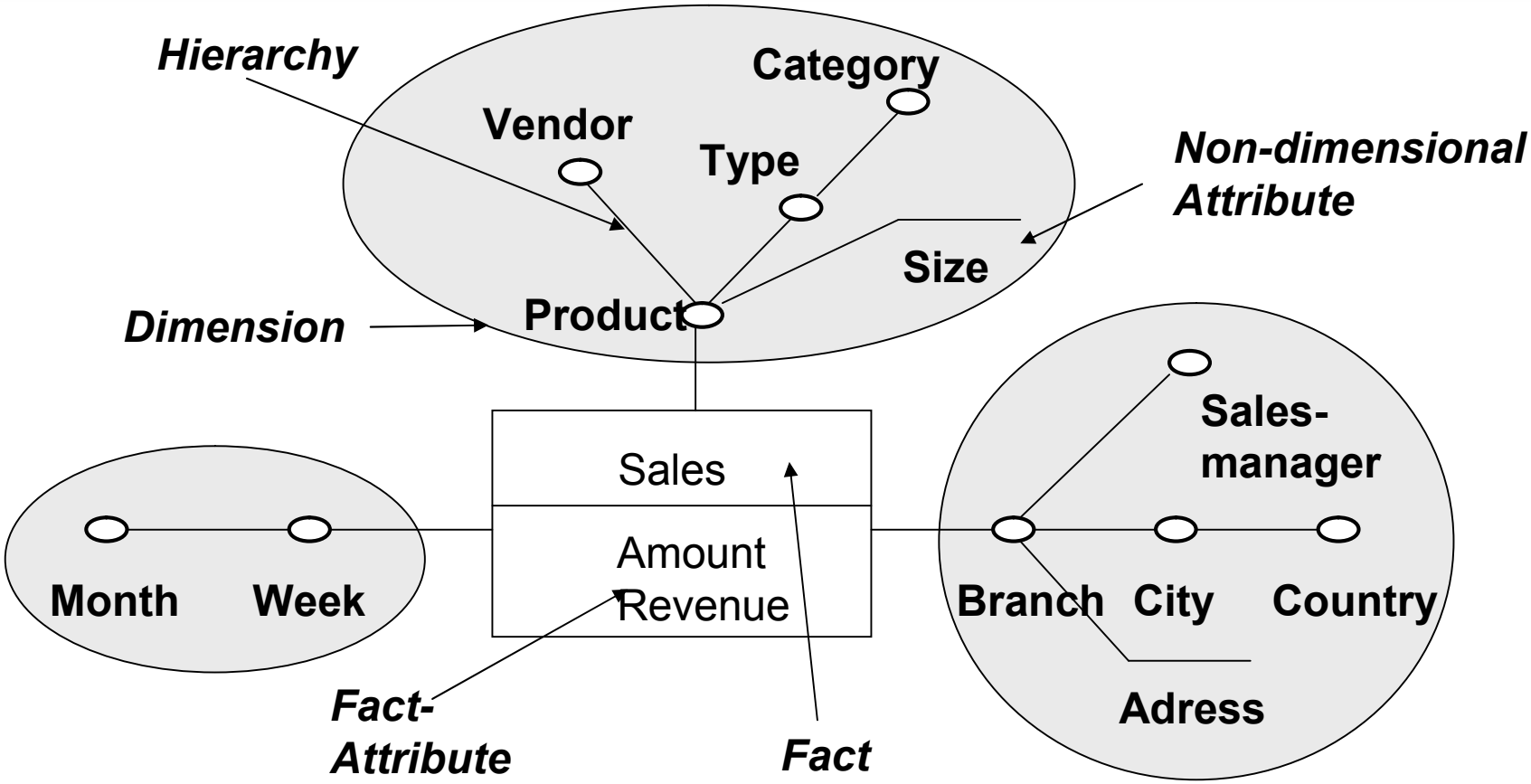
Example of a ME/R-Model



ERM: useful to a certain extent

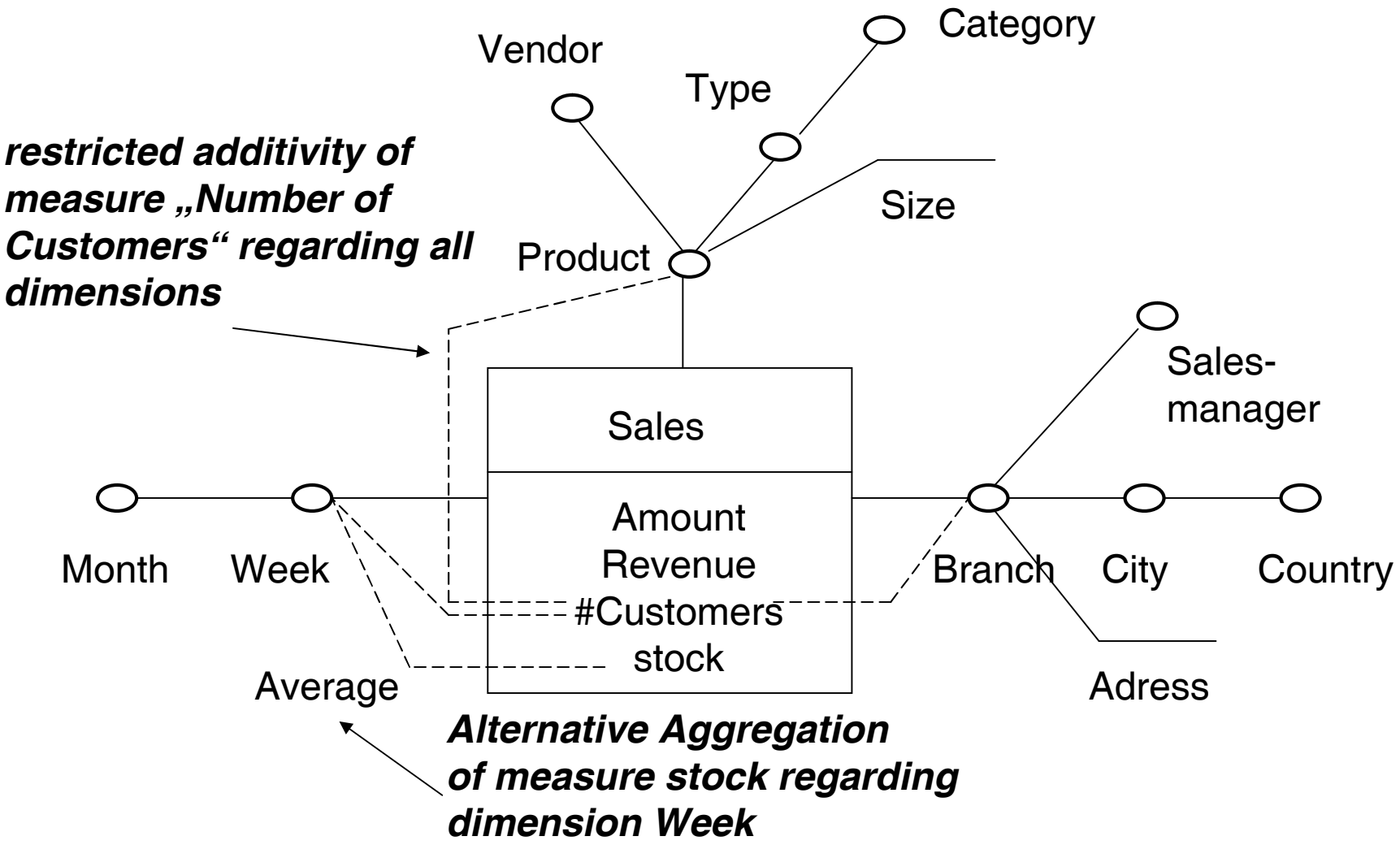
- Insufficient ability of handling single objects (Measures, actual/budget/forecast)
- Focus on classes of objects
- No support of connections between single objects (relative measures, calculation of measures)

Dimensional Fact Modeling

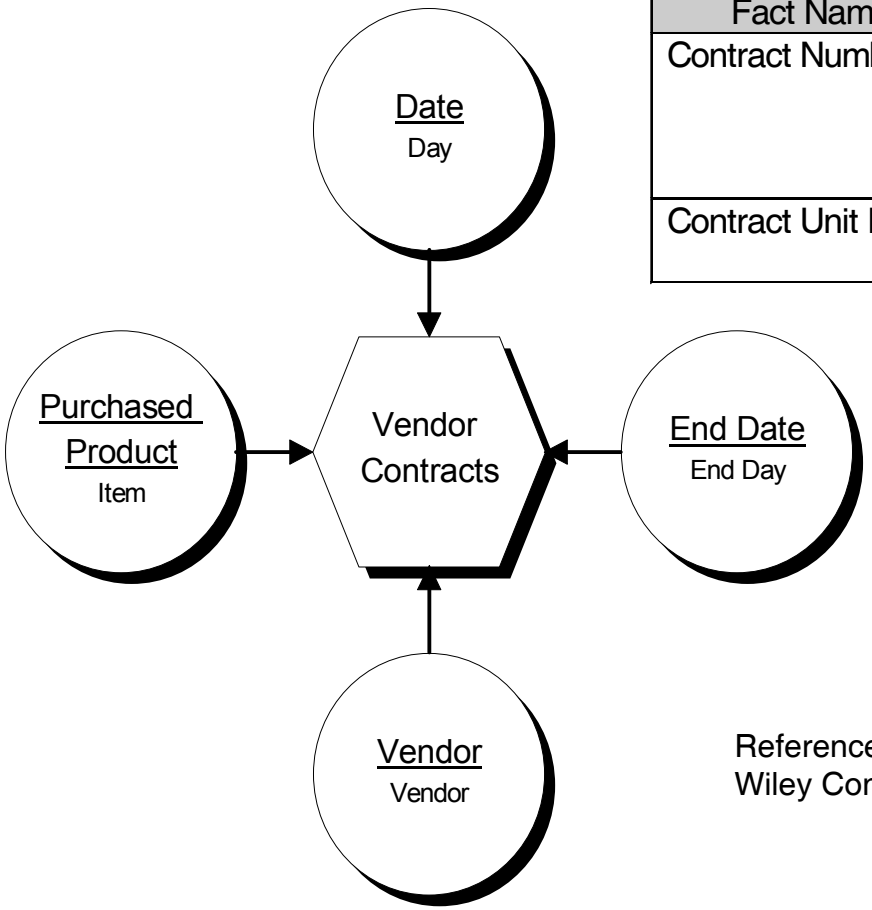


Reference: Golfarelli, Matteo; Maio, Dario; Rizzi, Stefano: Conceptual Design of Data Warehouses from E/R Schemas. In: Proceedings of the Hawaii International Conference On System Sciences, January 6-9, 1998, Kona, Hawaii.

Visualize aggregation with Dimensional Fact Modeling



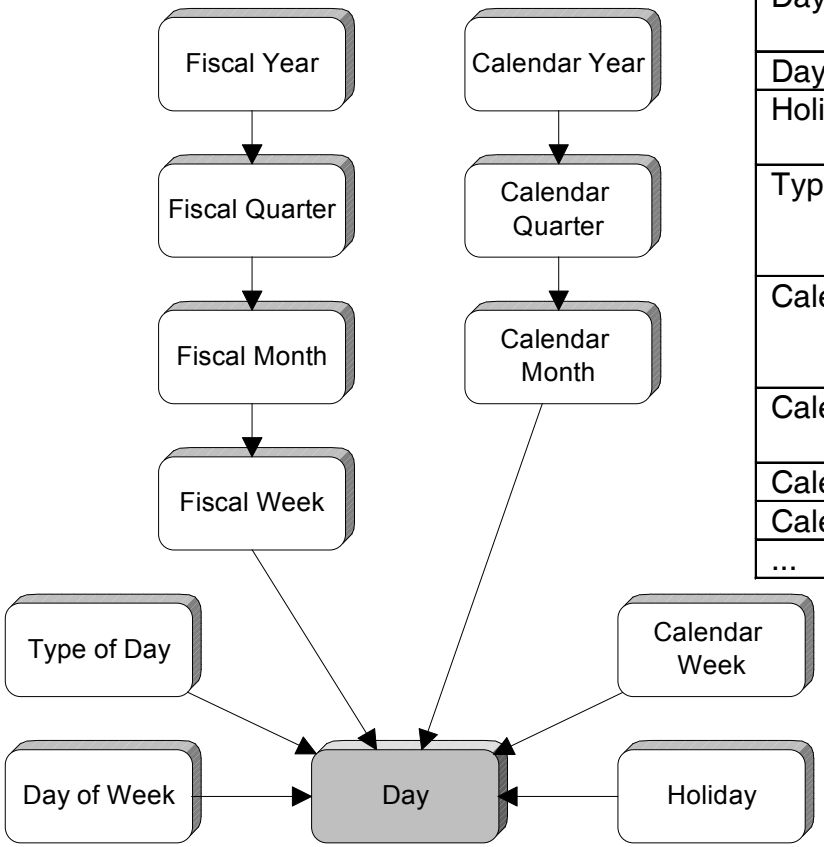
Dimensional Modeling (Facts)



Fact Name	Fact Description	Default Aggregation Rule
Contract Number	Indicates the specific contract number for the relationship between this vendor and Telco.	N/A
Contract Unit Price	The agreed upon unit price for this purchased item.	N/A

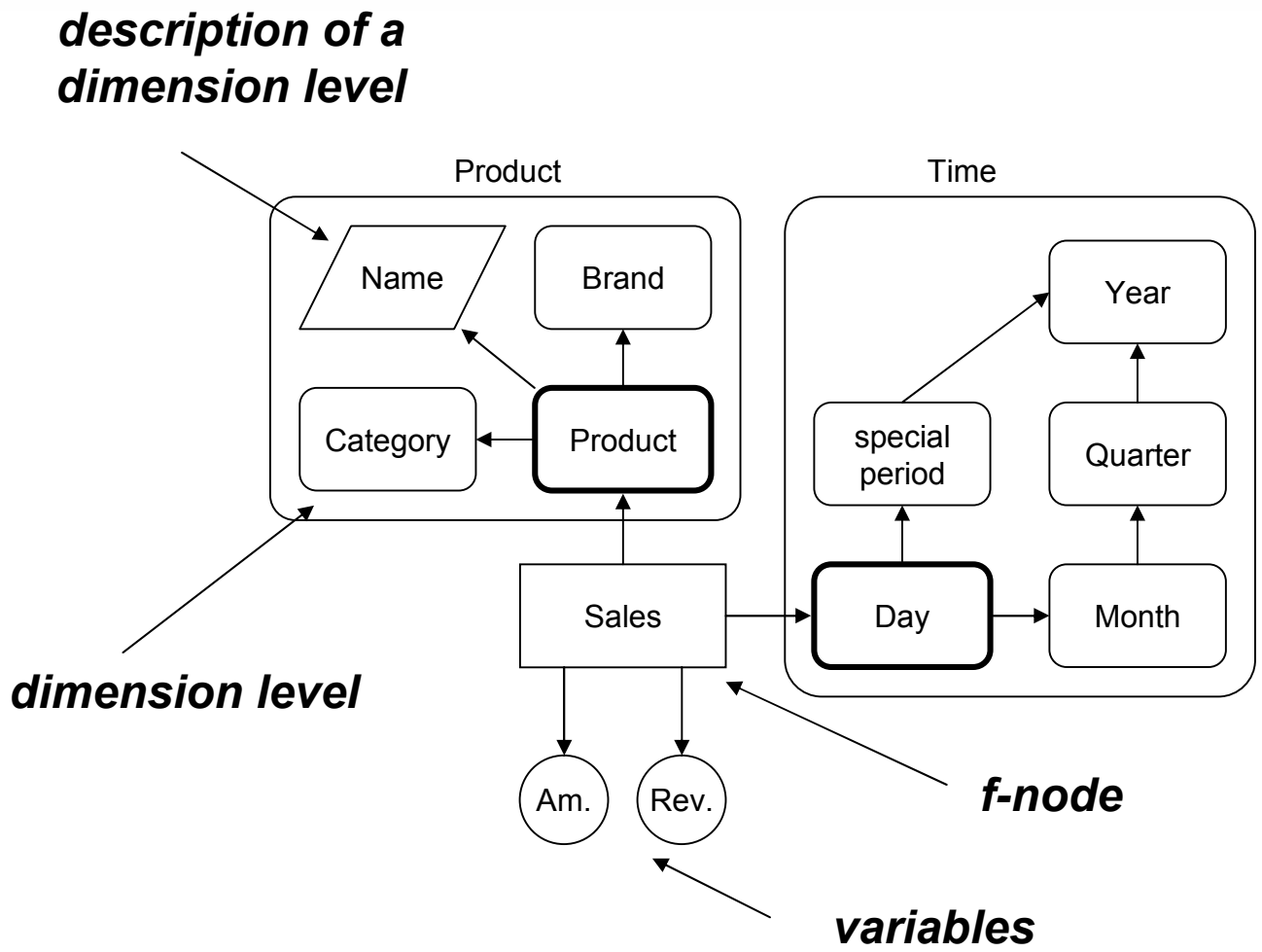
Reference: Ralph Kimball, The Data Warehouse Lifecycle Toolkit, Wiley Computer Publishing, New York et.al. 1998.

Dimensional Modeling (Dimensions)



Attribute Name	Attribute Description	Sample Values
Day	The specific day that an activity took place.	06/04/1998; 06/05/1998
Day of Week	The specific name of the day.	Monday; Tuesday
Holiday	Identifies that this day is a holiday.	Easter; Thanksgiving
Type of Day	Indicates whether or not this day is a weekday or a weekend day.	Weekend; Weekday
Calendar Week	The week ending date, always a Saturday. Note that WE denotes week ending.	WE 06/06/1998; WE 06/13/1998
Calendar Month	The calendar month.	January,1998; February, 1998
Calendar Quarter	The calendar quarter.	1998Q1; 1998Q4
Calendar Year	The calendar year.	1998
...	...	...

Multidimensional Data Model ($\mathcal{MD}$)



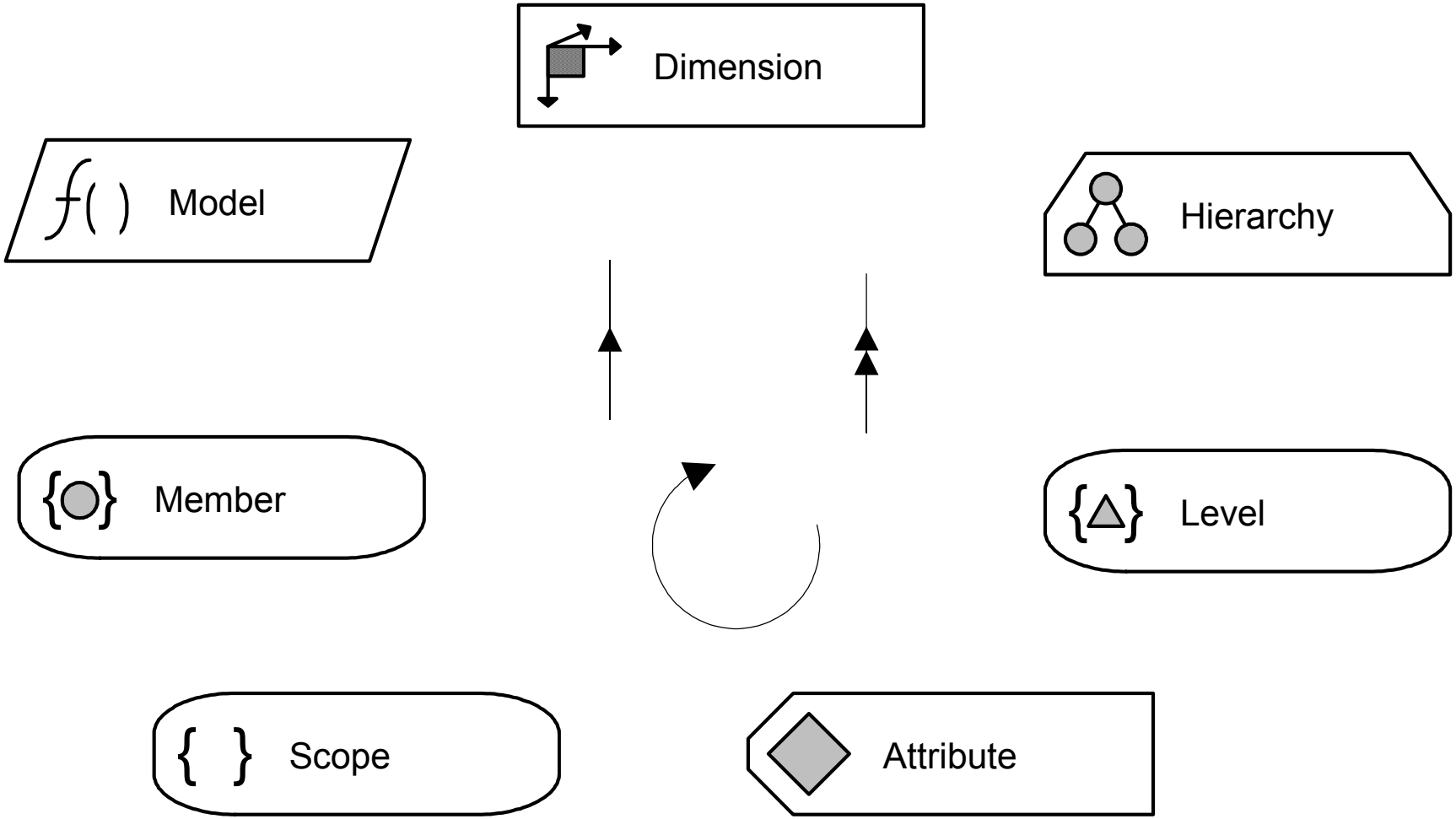
Agenda: Conceptual Design

- Methods for conceptual design
- Conceptual modeling with ADAPT
- Tool support by Microsoft Visio™
- Tool support by SemTalk™

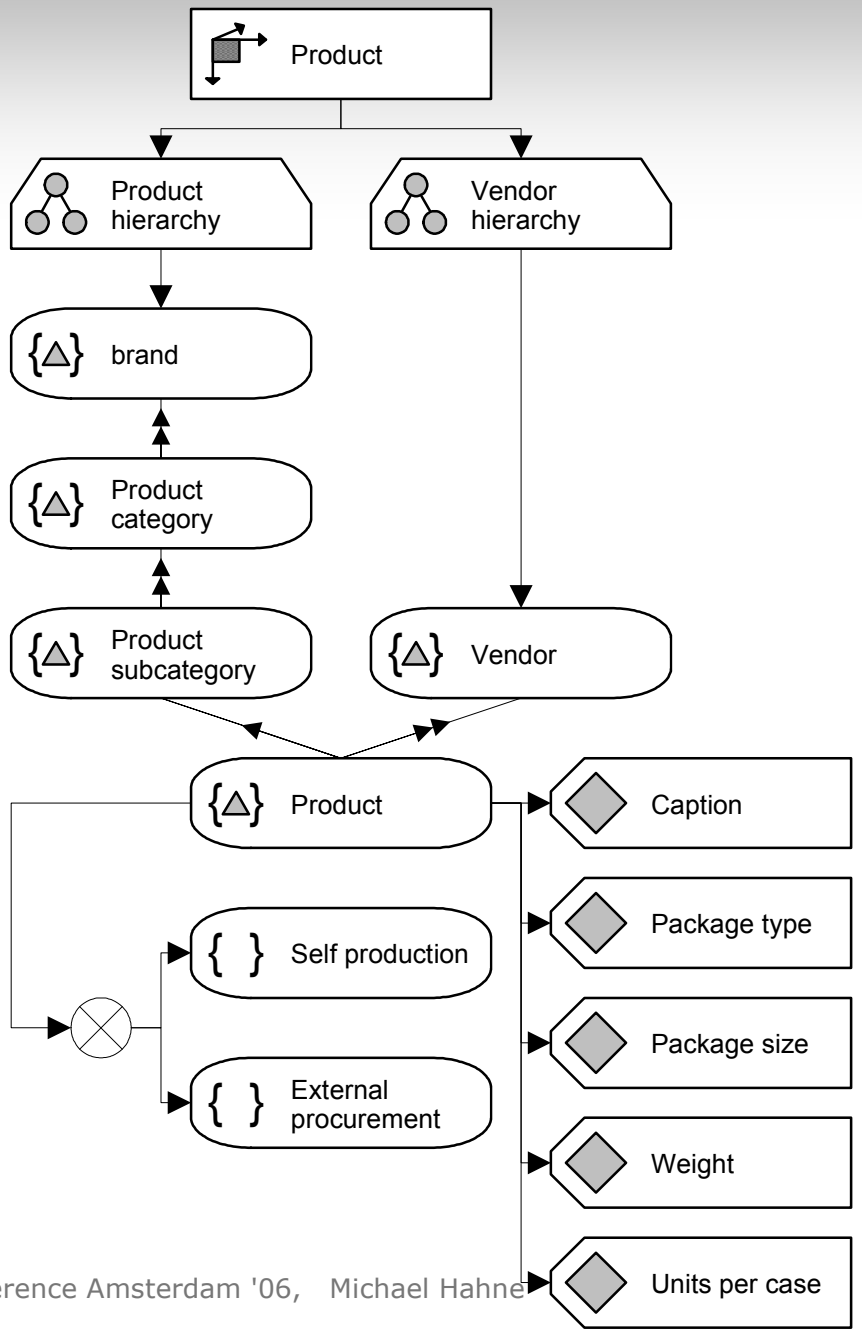
Objects in ADAPT

- ADAPT: Application Design for Analytical Processing Technologies
- Objects for modeling of dimensions
- Objects for modeling of cubes
- Objects representing relationships

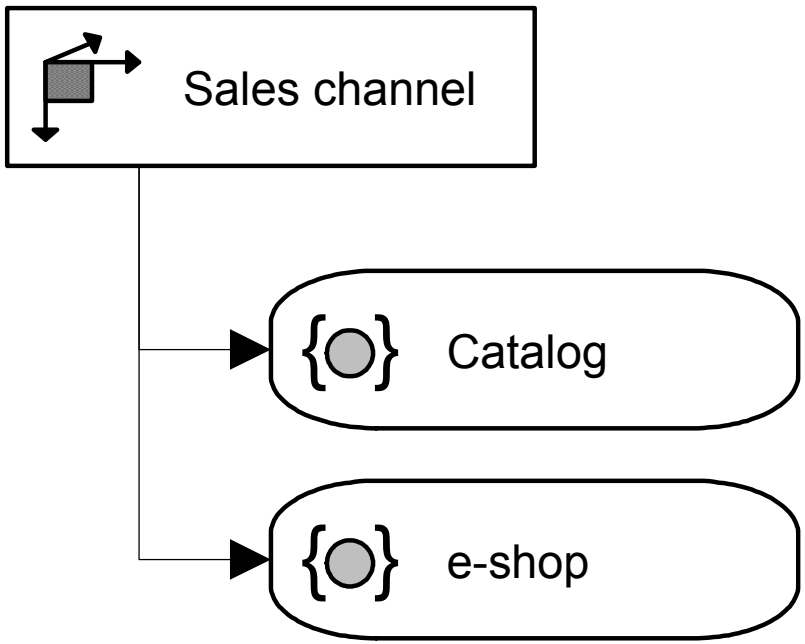
Objects for modeling of dimensions in ADAPT



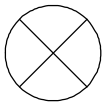
Typed dimension



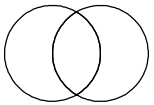
Non-typed dimension



Types of relationships in ADAPT

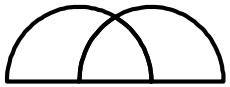


Fully exclusive



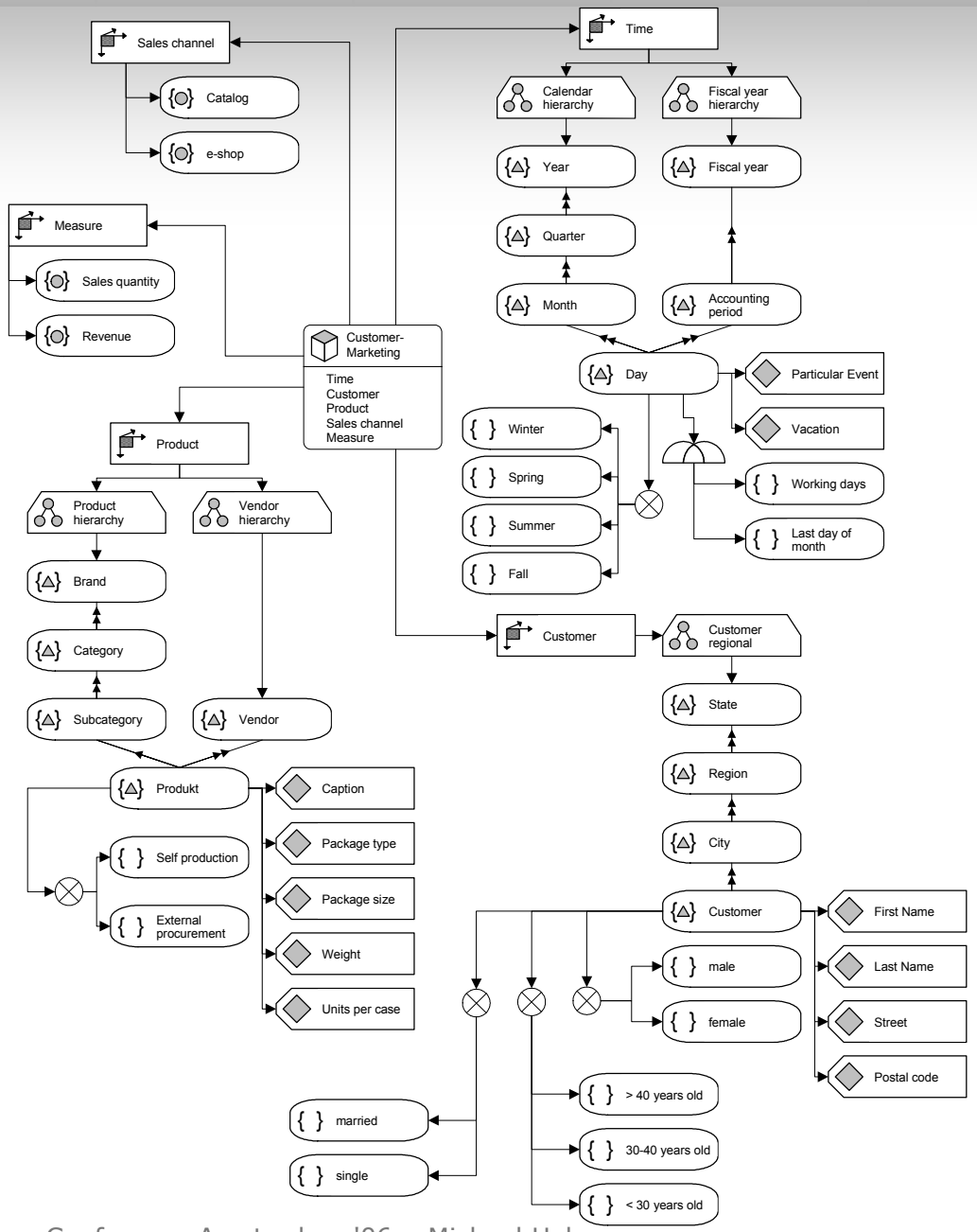
Fully overlapping

Partially exclusive

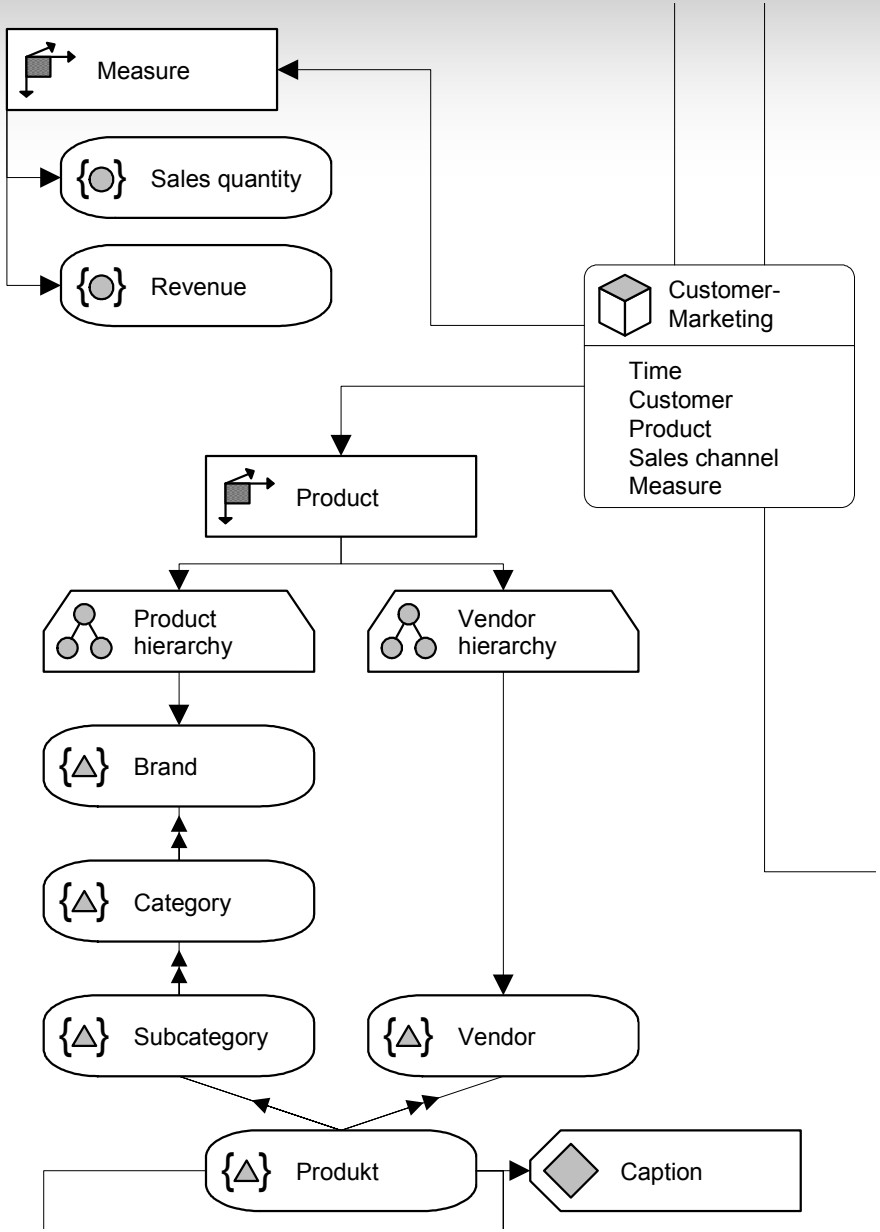


Partially overlapping

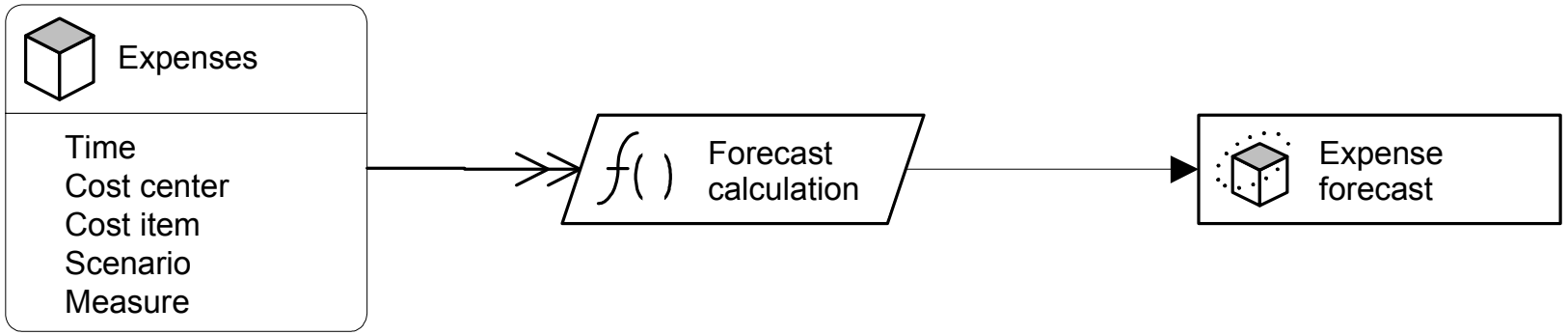
Cube in ADAPT



Cube in ADAPT (Zoom)



Derived view of a cube (model)



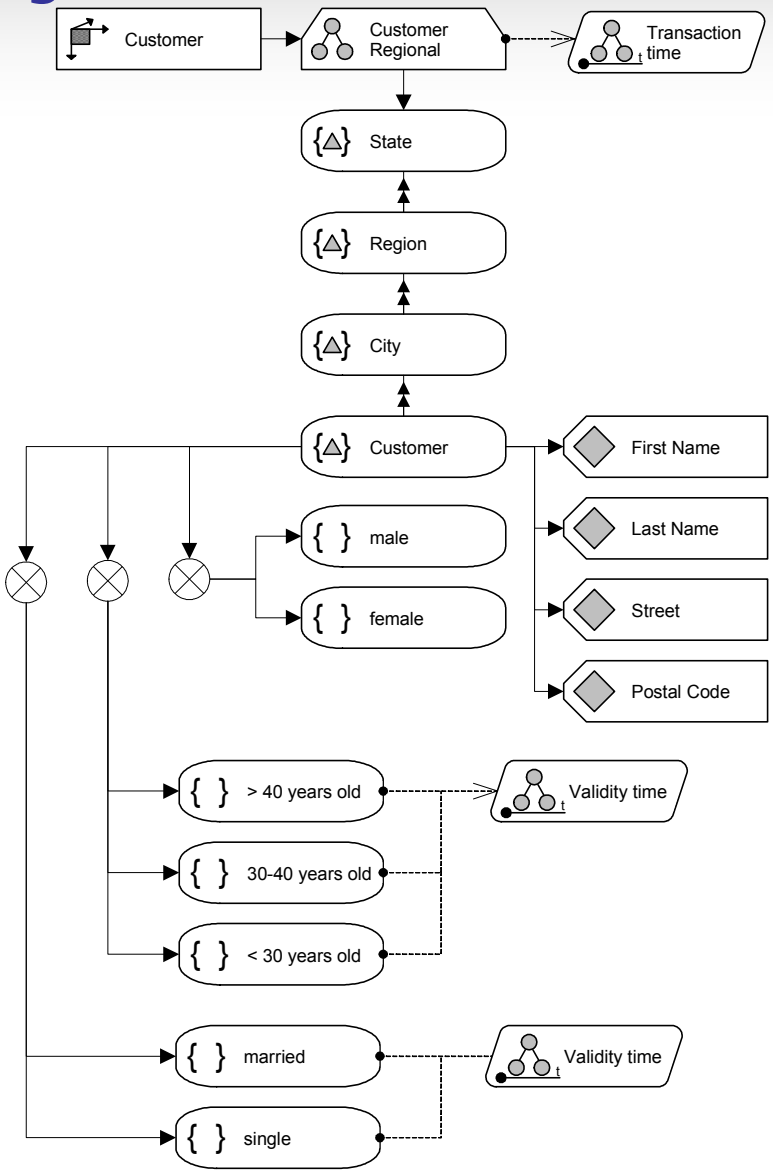
Extension coming up with T-ADAPT

•In addition to the objects in ADAPT there are the following objects in T-ADAPT (*M. Hahne*):

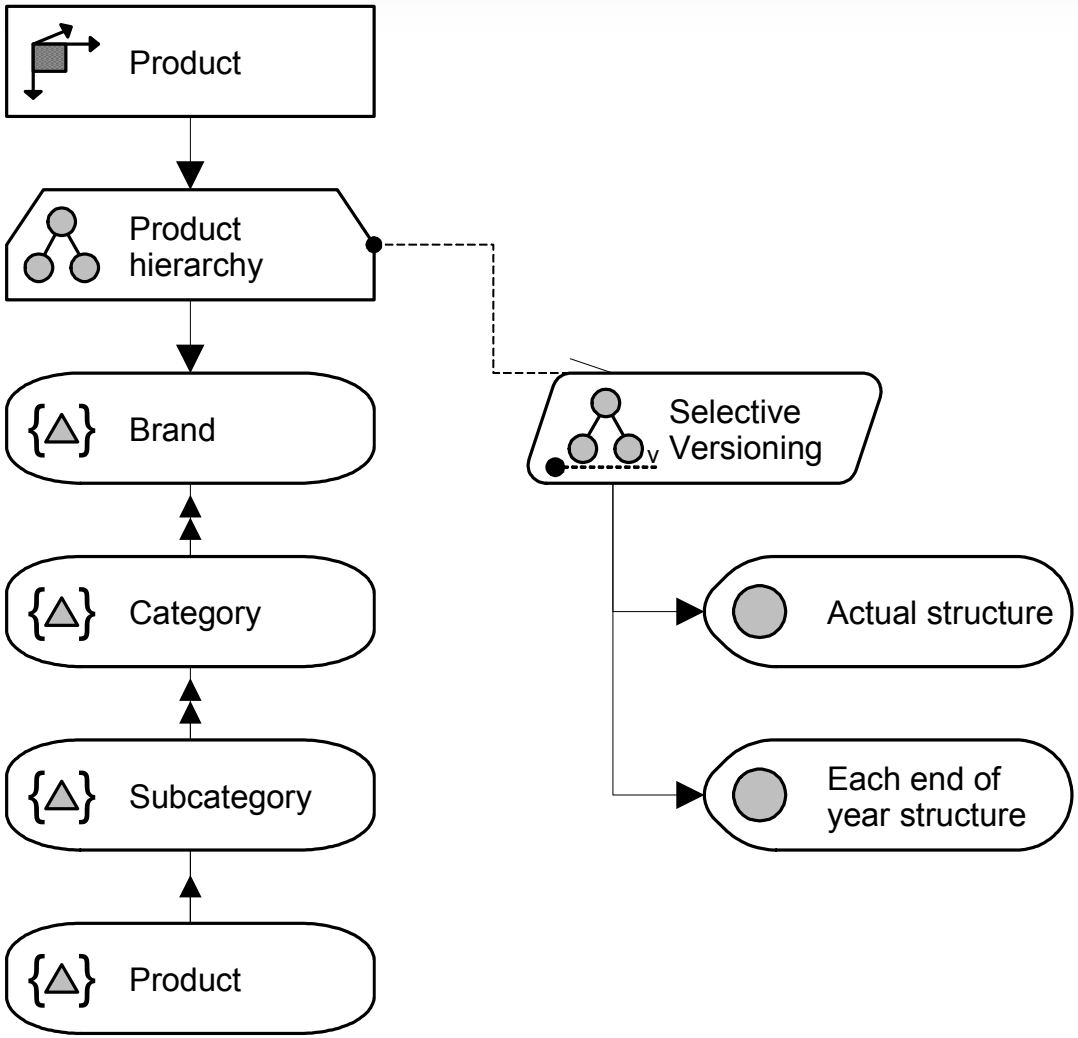
- Complete versioning
  - Special case of transactional versioning
- Selective Versioning

- New objects:

Complete versioning on the basis of valid time and transaction time



Separate versions



Agenda: Conceptual Design

- Methods for conceptual design
- Conceptual modeling with ADAPT
- Tool support by Microsoft Visio™
- Tool support by SemTalk™

Agenda: Conceptual Design

- Methods for conceptual design
- Conceptual modeling with ADAPT
- Tool support by Microsoft Visio™
- Tool support by SemTalk™

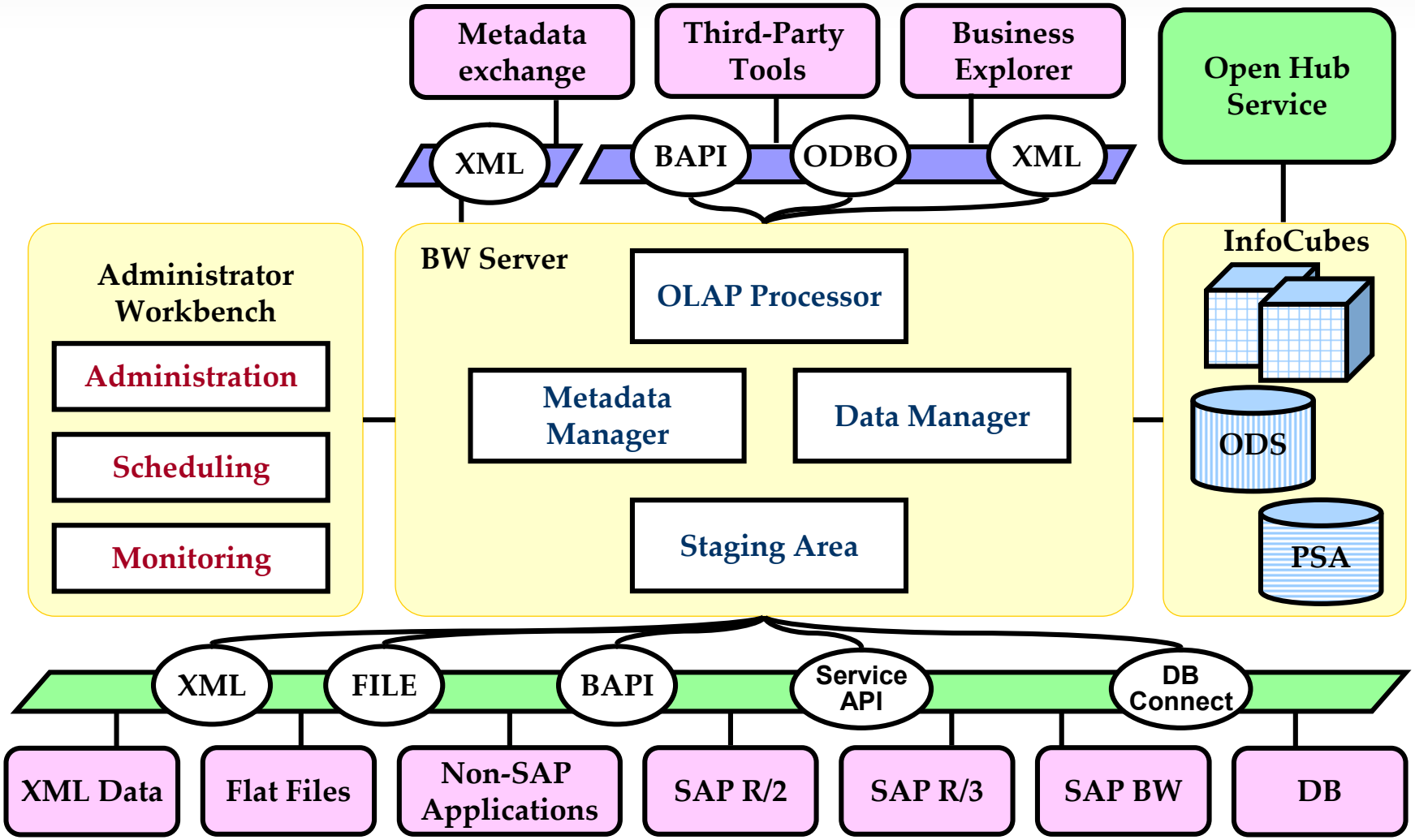
Agenda: Data Model of SAP BW

- Architecture of the SAP Business Information Warehouse
- Extended Star Schema of the SAP AG

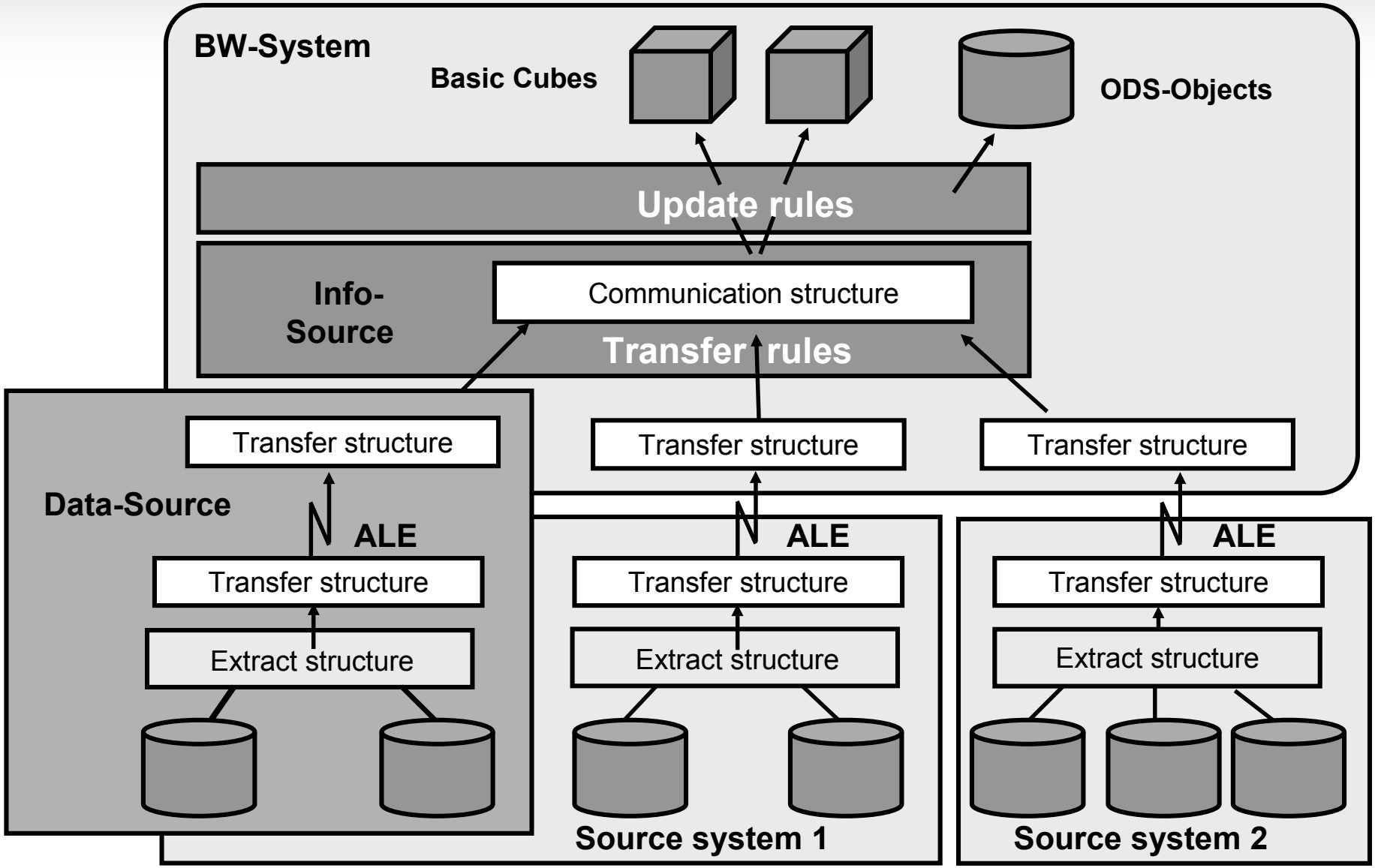
Agenda: Data Model of SAP BW

- Architecture of the SAP Business Information Warehouse
- Extended Star Schema of the SAP AG

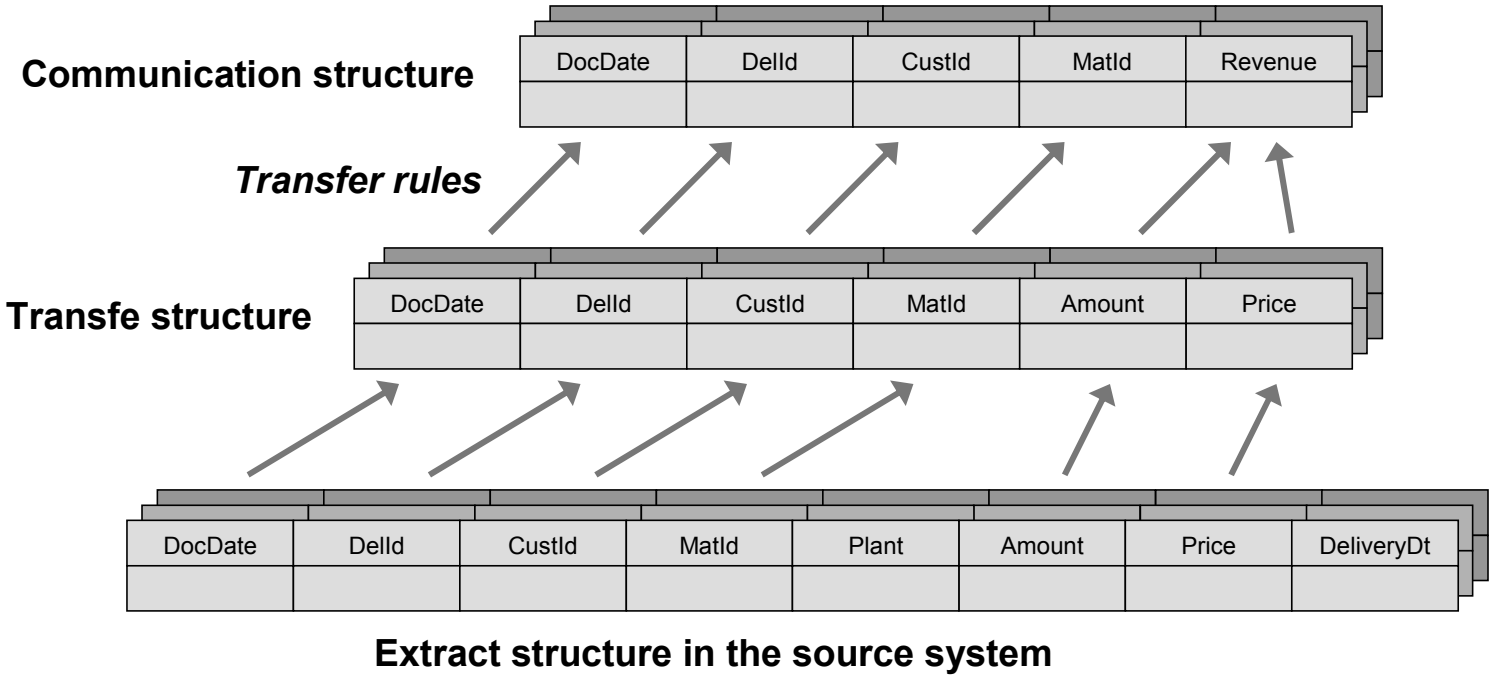
Architecture of SAP® Business Information Warehouse



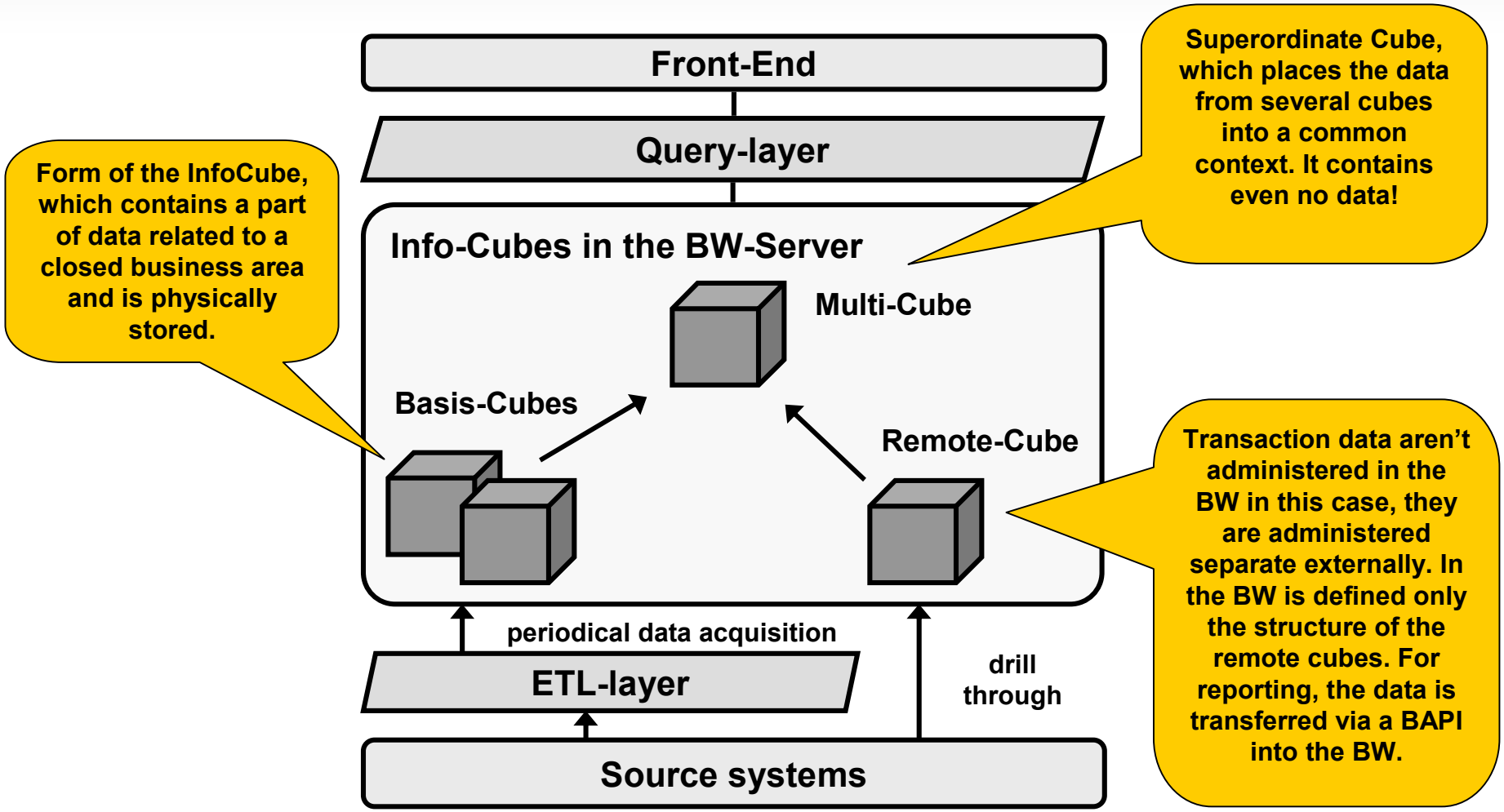
Transformation steps



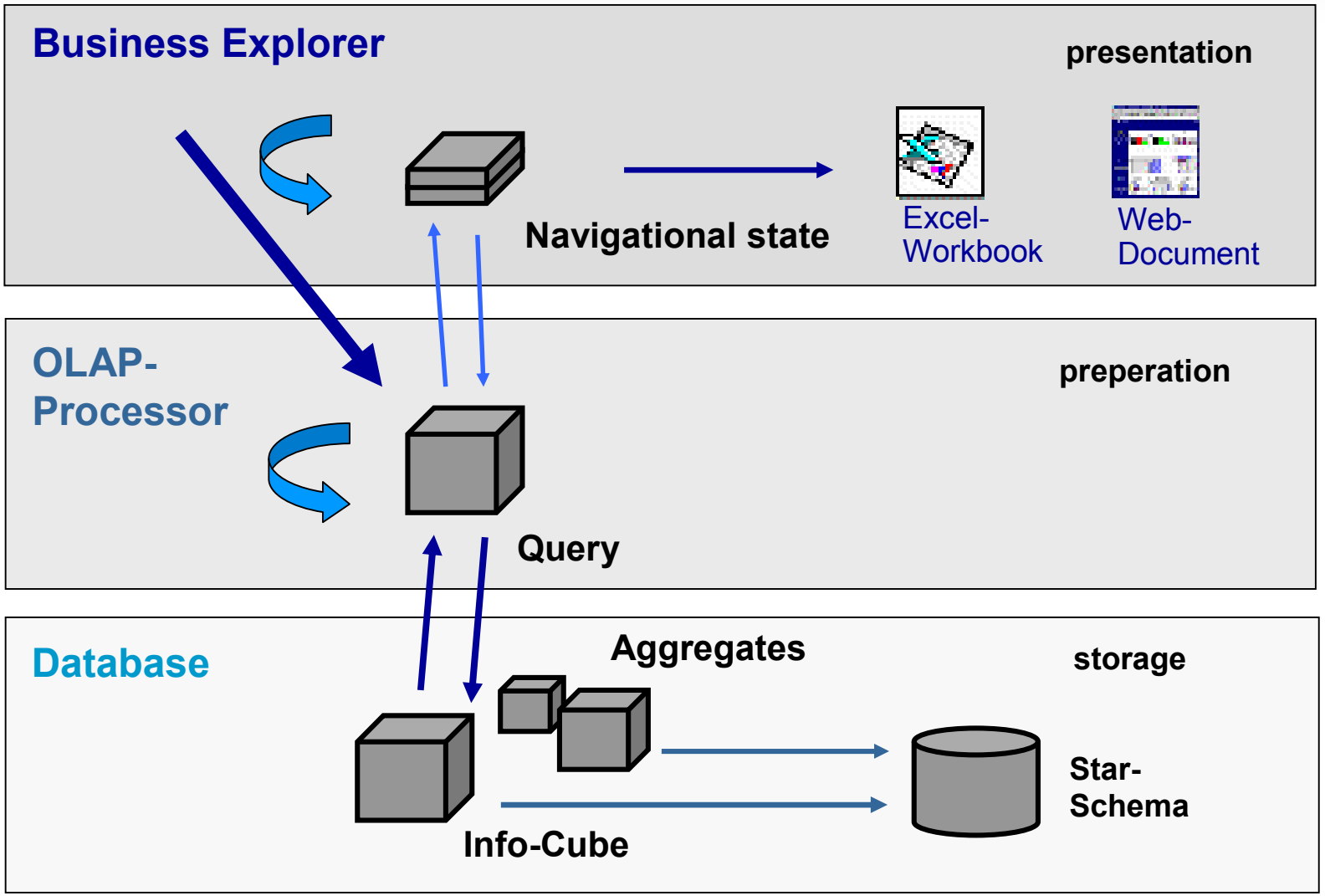
Transformation within transfer rules



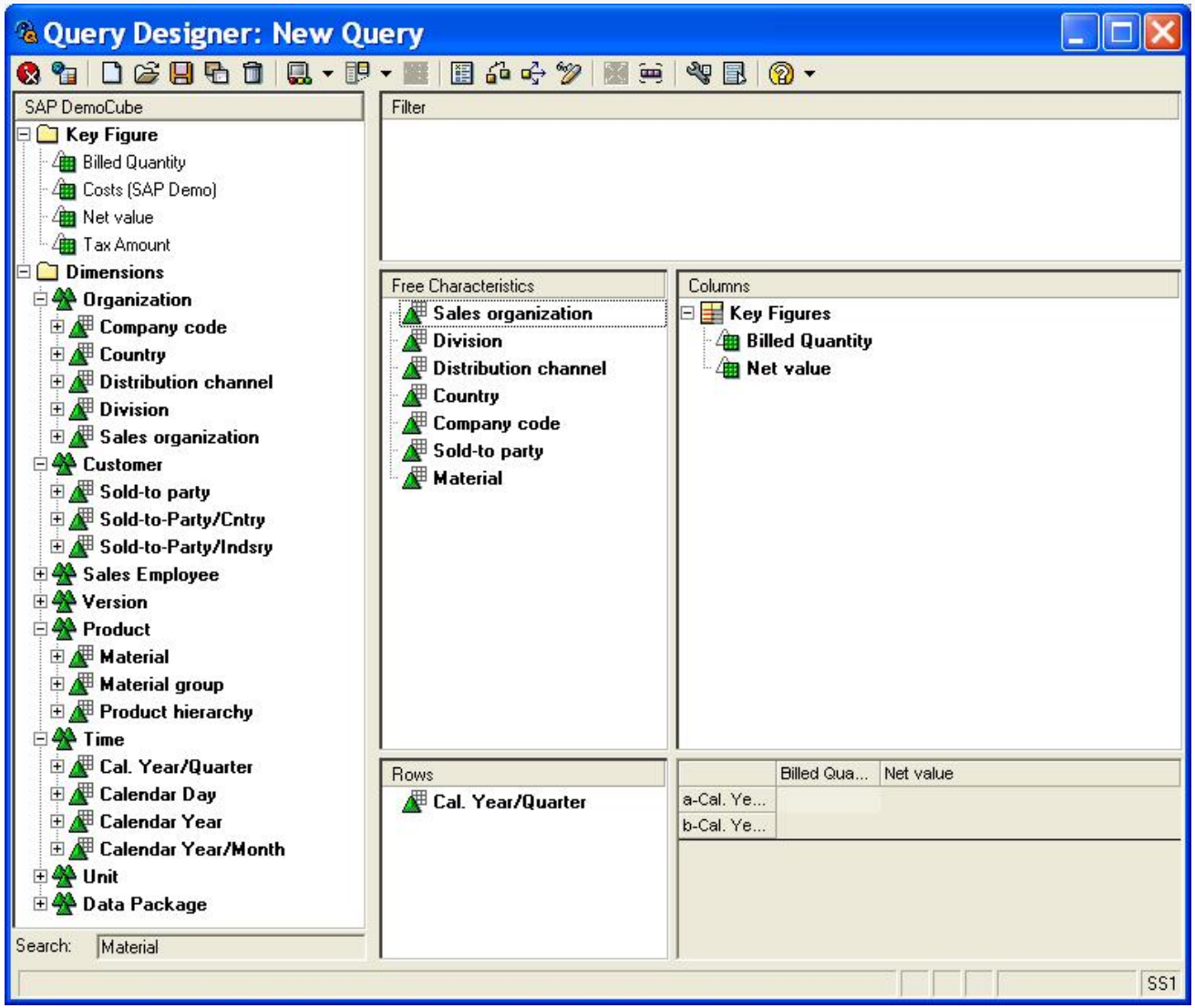
Types of Info-Cubes



Steps of query execution



Query definition in QueryDesigner

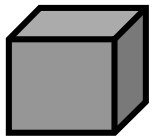


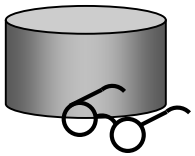
Agenda: Data Model of SAP BW

- Architecture of the SAP Business Information Warehouse
- Extended Star Schema of the SAP AG

Data targets in SAP® BW

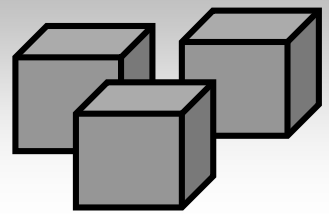
Data targets are objects in which transactional data is stored for the purpose of reporting and analysis. The most important data targets are:

InfoCubes 

ODS Objects 

Additionally, there are further data targets in BW, which enable (for example) direct master data reporting (Info Sets, Info Objects)

Definition of InfoCube



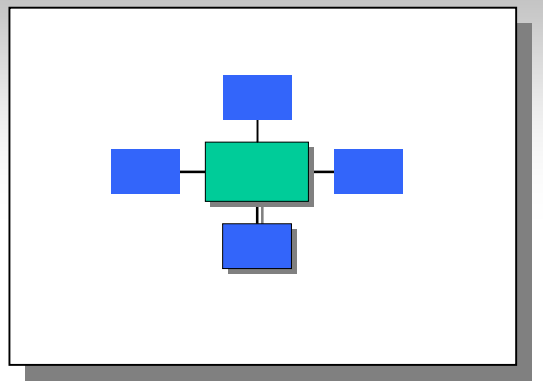
- The InfoCube is central data storage serving as the basis for reports and analyses in SAP® BW. It contains a restricted data volume, such as might relate for example to a specific well-defined business area or business unit.

- InfoCubes contain two types of data: measures and characteristics.

- The term "InfoCube" designates a table structure in which some relational tables are linked according to the so-called Star Schema (multidimensional data storage)

- Star Schema: Dimension tables are grouped in a star formation around a central fact table

Star Schema



- The Star Schema is the most typical way of representing multidimensional data structures in relational data bases.

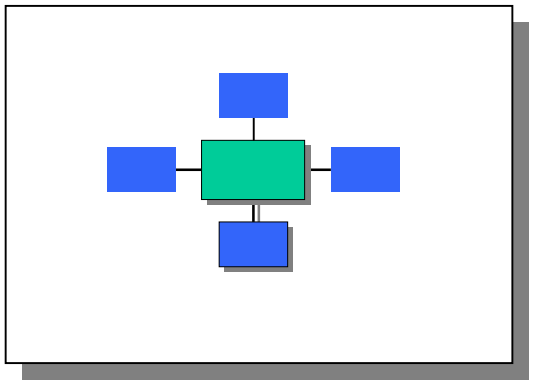
- In the Star Schema, facts are stored in a separate fact table, while characteristics are grouped in Dimension tables. The dimension tables are joined to the fact table via foreign key and primary key relationships (DIM ID).

- In this way, all data records from the fact table are marked uniquely by a combination of these foreign key values from the dimension tables.

Pros and cons of the general Star Schema

☺ Good performance for data analysis

☺ Very flexible when adding characteristics and measures



😐 Problems come along with:

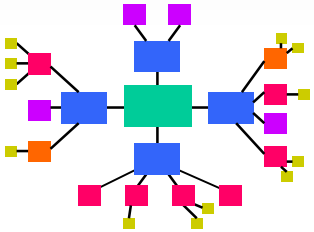
- many-many relationships and
- unbalanced hierarchies

because of the uniqueness of the primary keys in the dimension tables

Therefore, SAP AG decided to extend the Star Schema.

Master data is stored separately and independently from InfoCubes, in the so-called Extended Star Schema.

Extended Star Schema

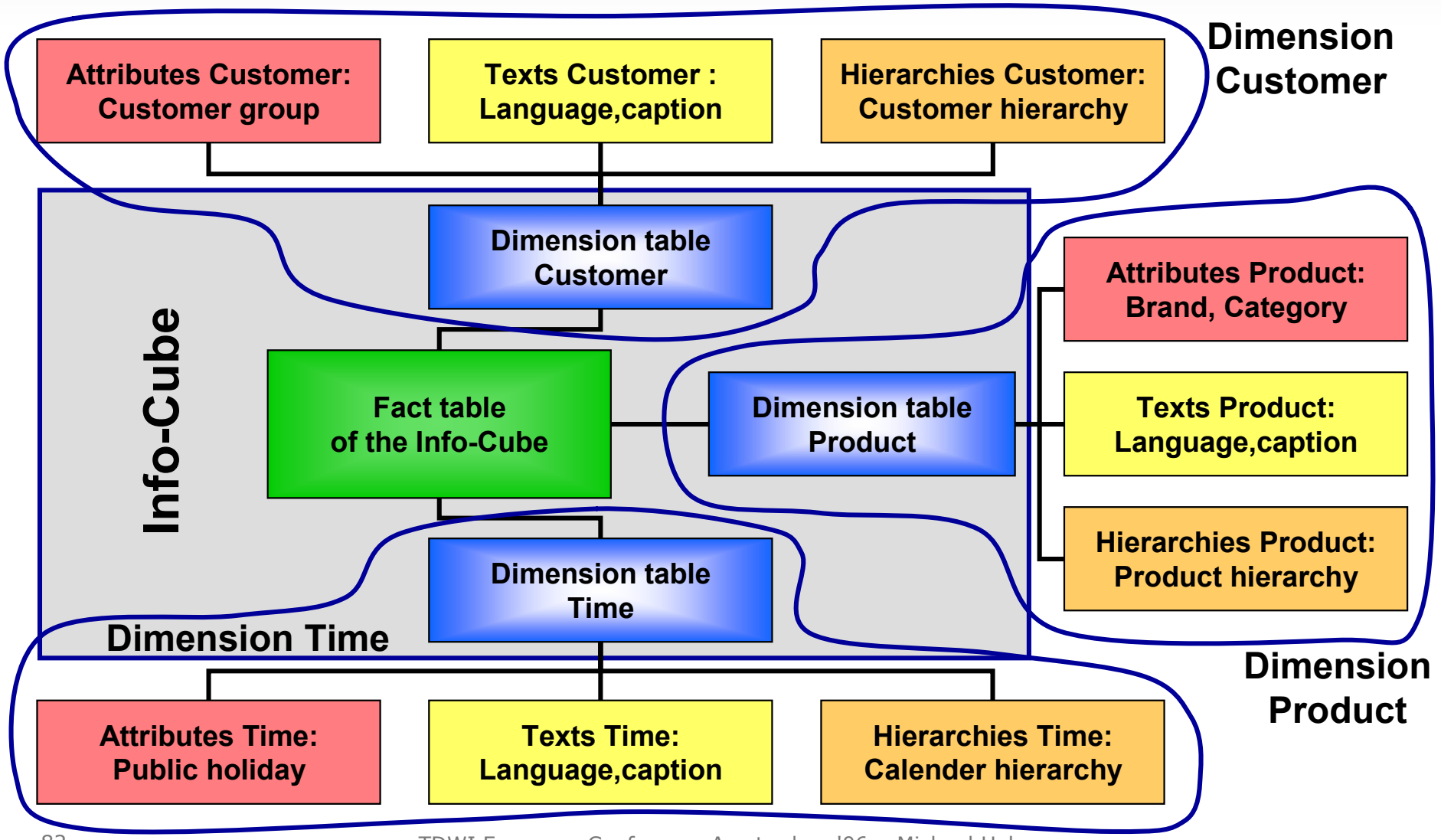


The Extended Star Schema gives access to:

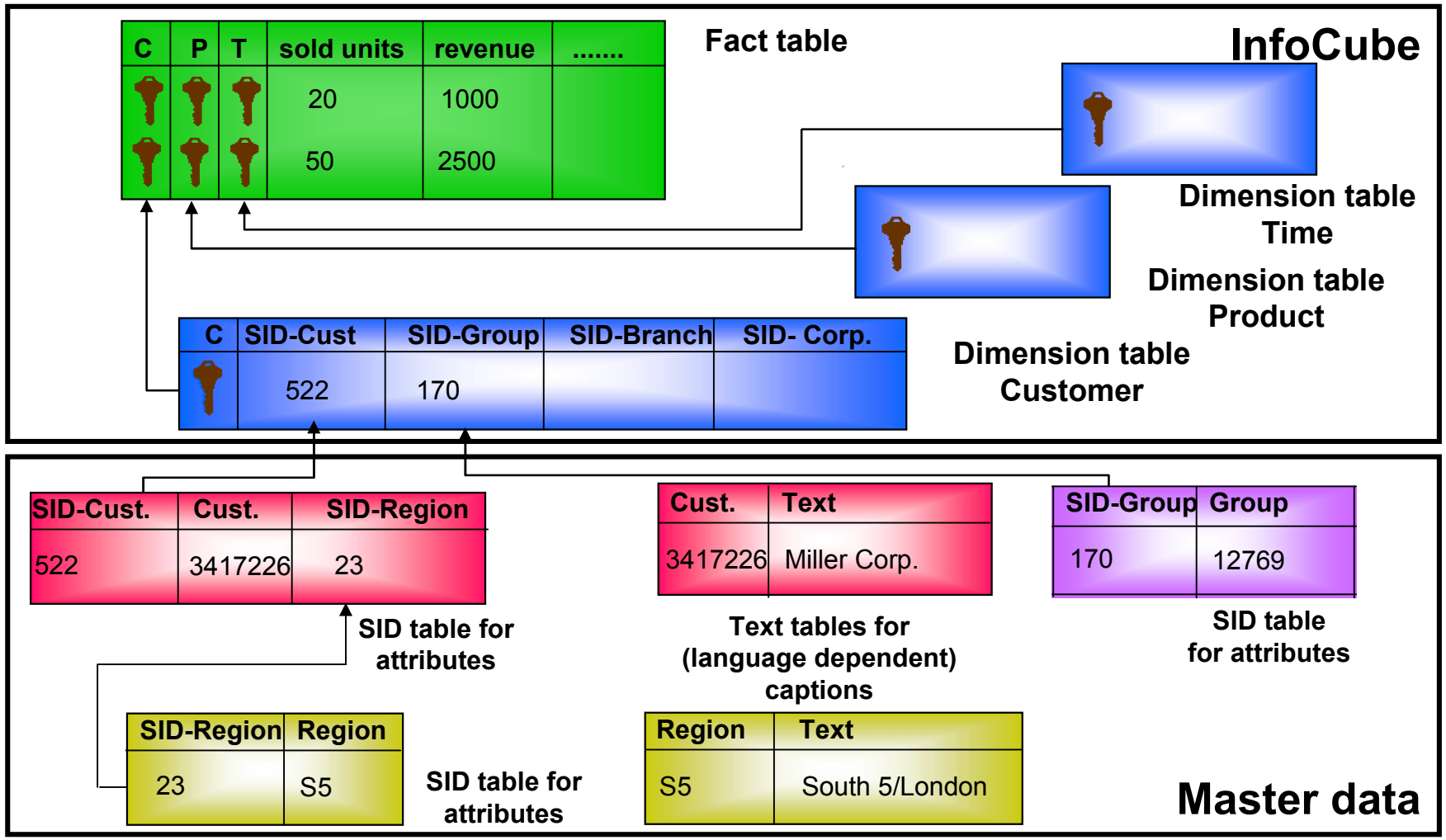
- Master data tables and their corresponding attributes
- Text tables with extensive multilingual captions
- External hierarchy tables for structured data access

Master data and hierarchy tables are joined to the fact table via SID Tables (pointer tables) and dimension tables

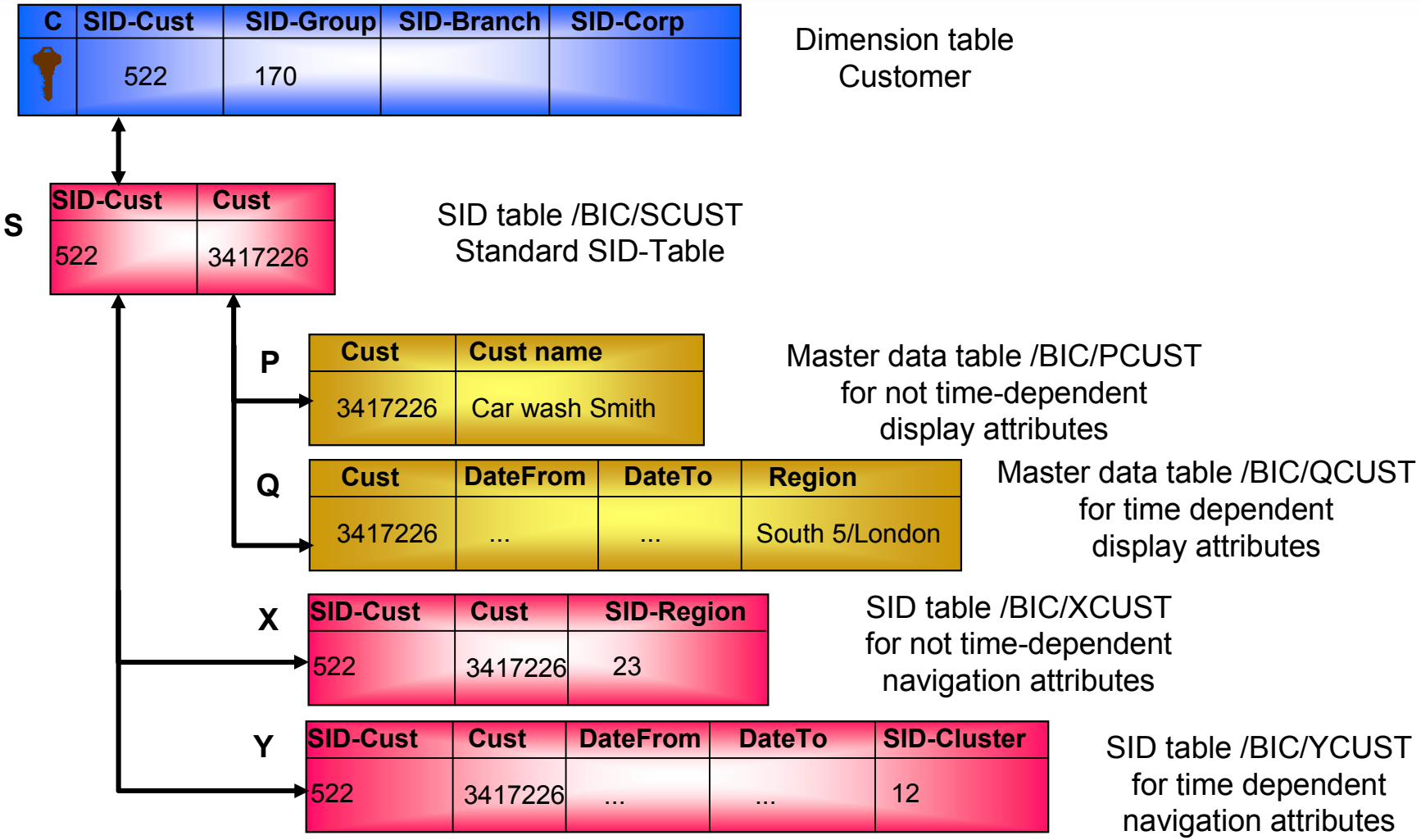
Concept of master data



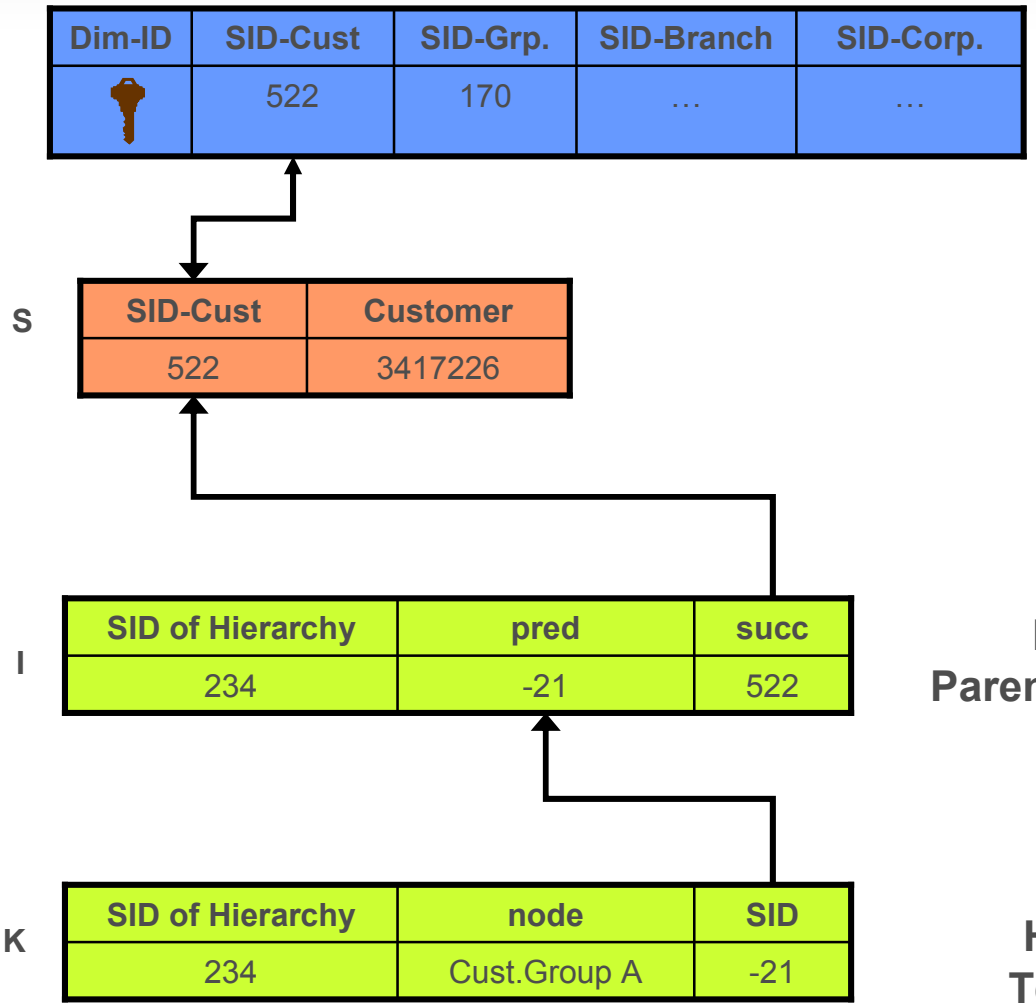
Connecting master data to cubes via SID Tables



Different master data tables



Hierarchy tables



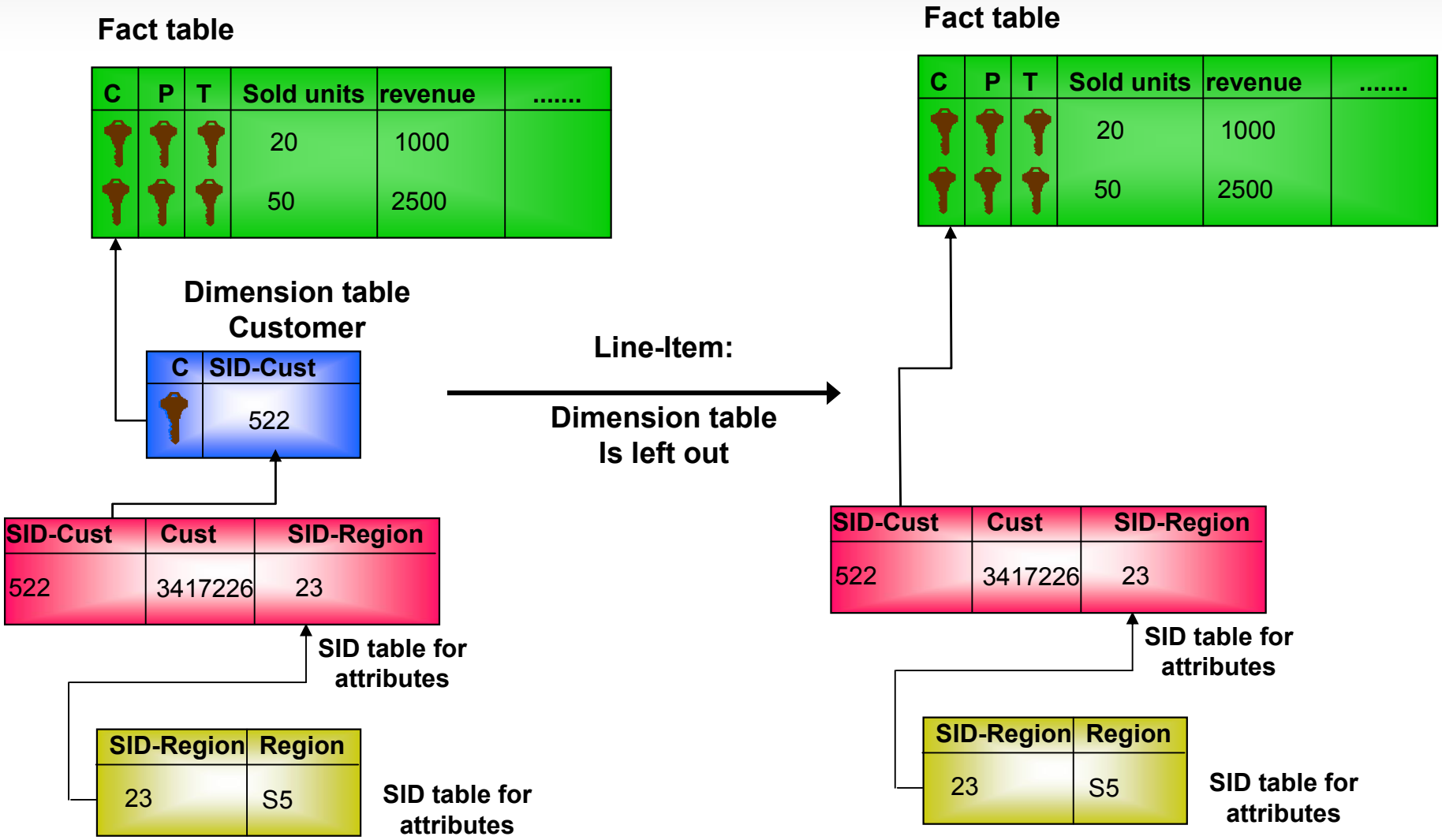
Dimension table Customer

SID Table /BIC/SCUST
Standard SID Table

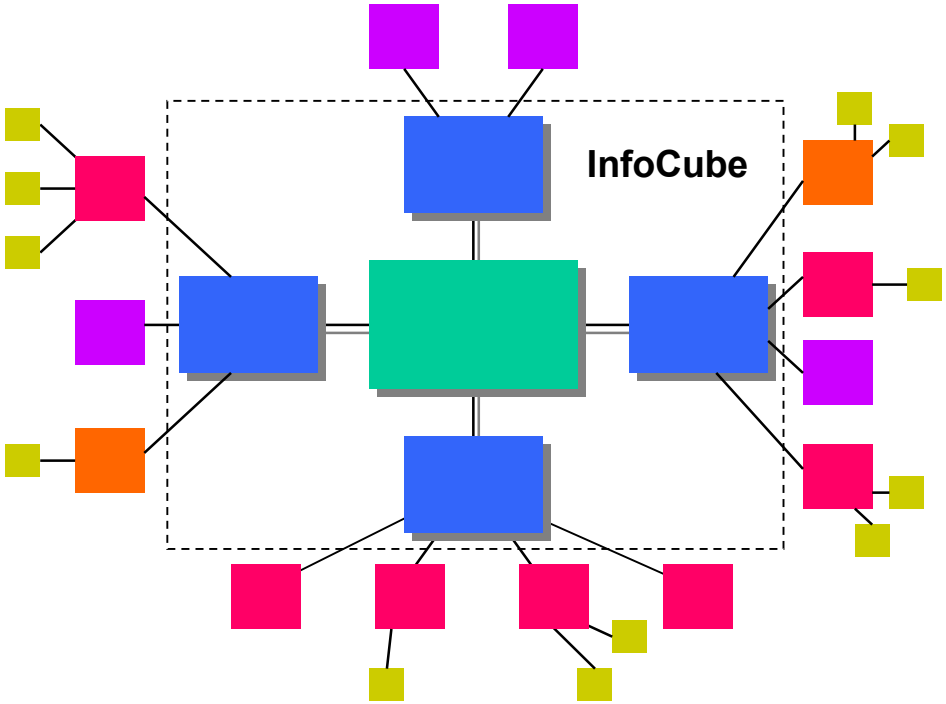
Hierarchy table /BIC/ICUST
Parent-Child-Tuple of the hierarchies

Hierarchy table /BIC/KCUST
Text nodes of the hierarchies

Line-Item dimensions



Complexity of the extended Star Schema, at a glance



(1) Fact Table

(2) Dimension tables

(3) Time-independent SID
Time-dependent SID
Conventional SID

(4) SID Attributes

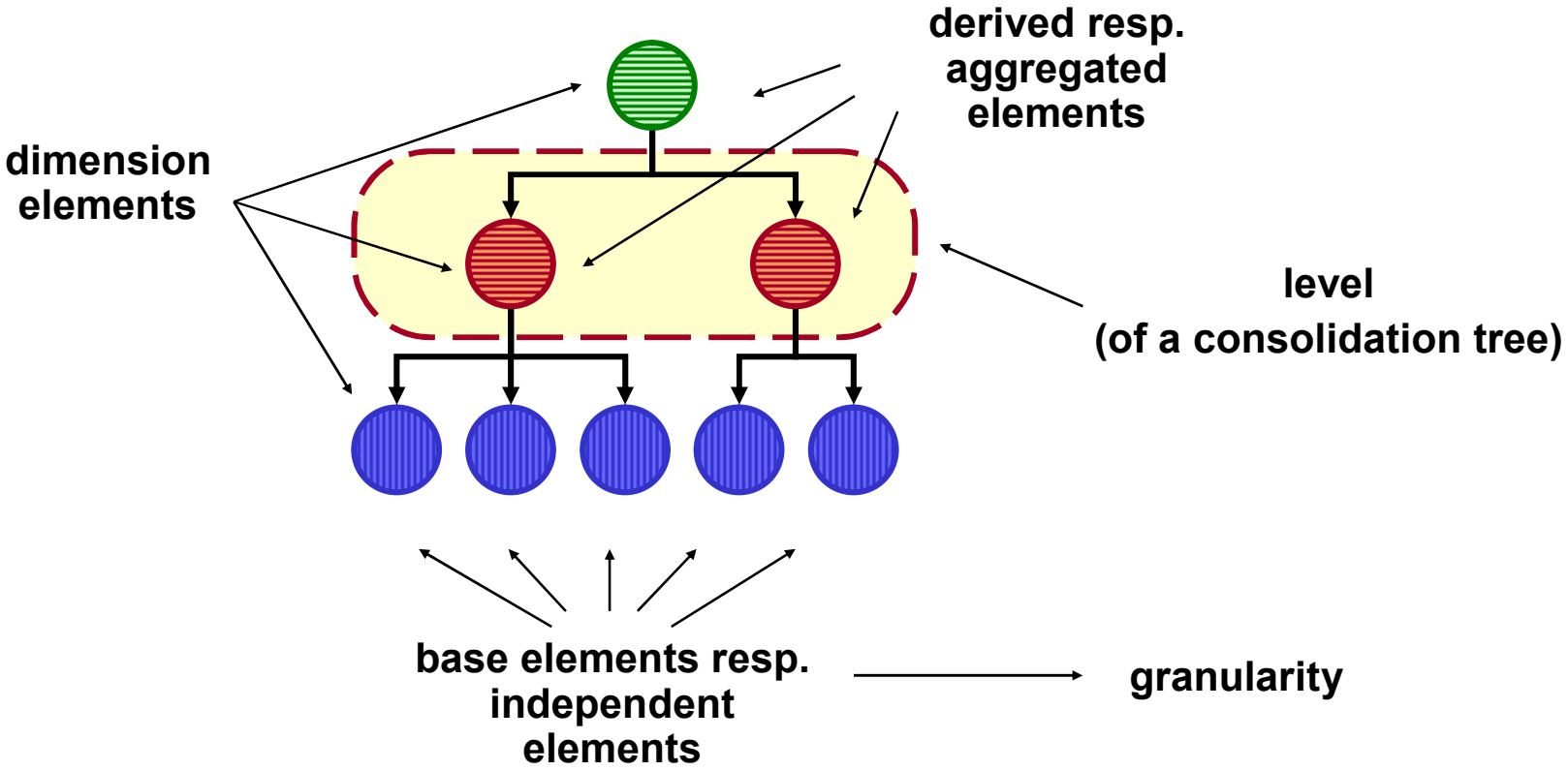
Agenda: Modeling hierarchies in SAP BW

- Variants for modeling hierarchical dimension structures
- Temporal aspects and time stamping
- Modeling guidelines

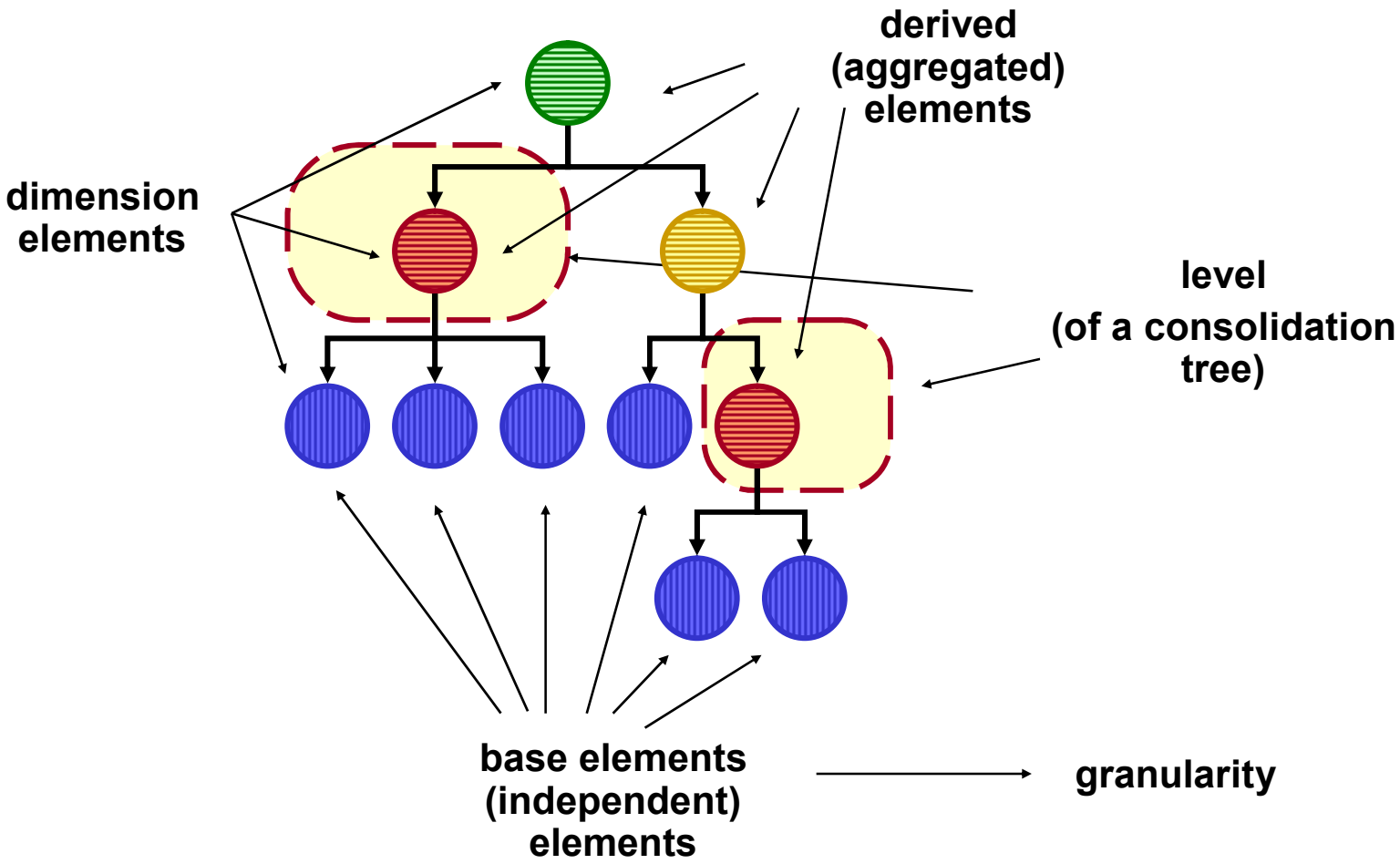
Agenda: Modeling hierarchies in SAP BW

- Variants for modeling hierarchical dimension structures
- Temporal aspects and time stamping
- Modeling guidelines

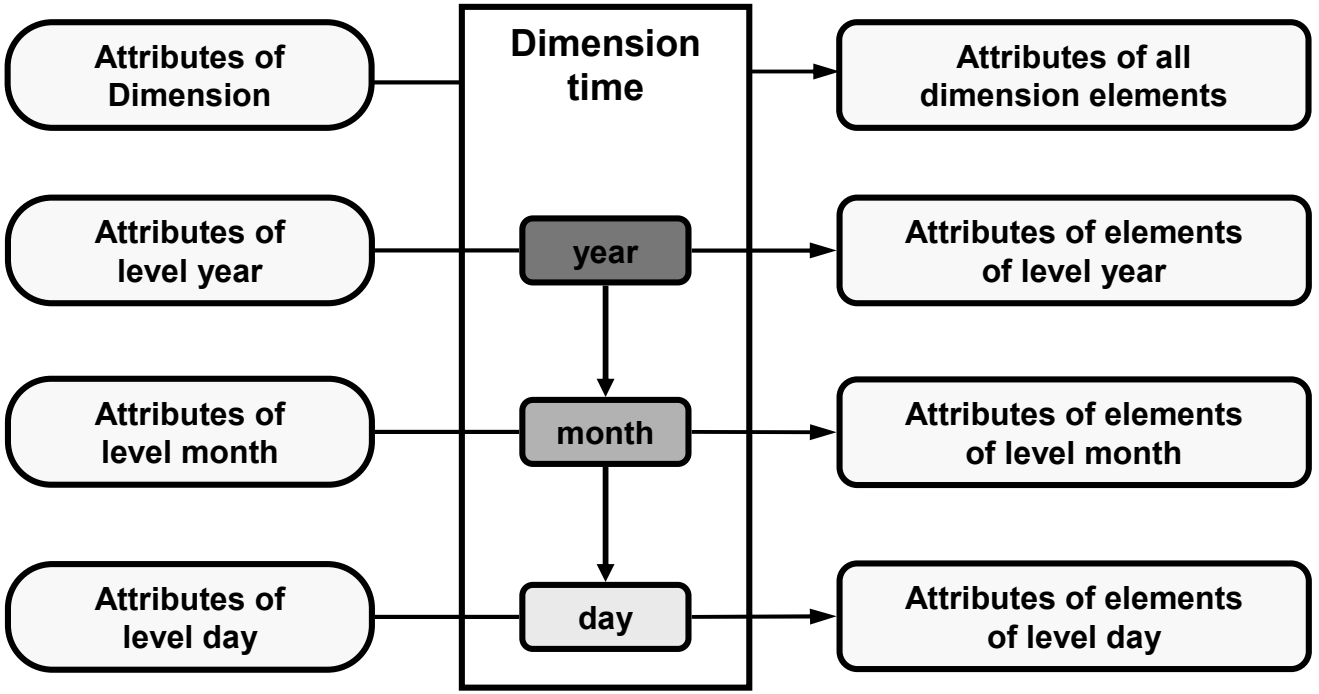
Dimensions and balanced hierarchies



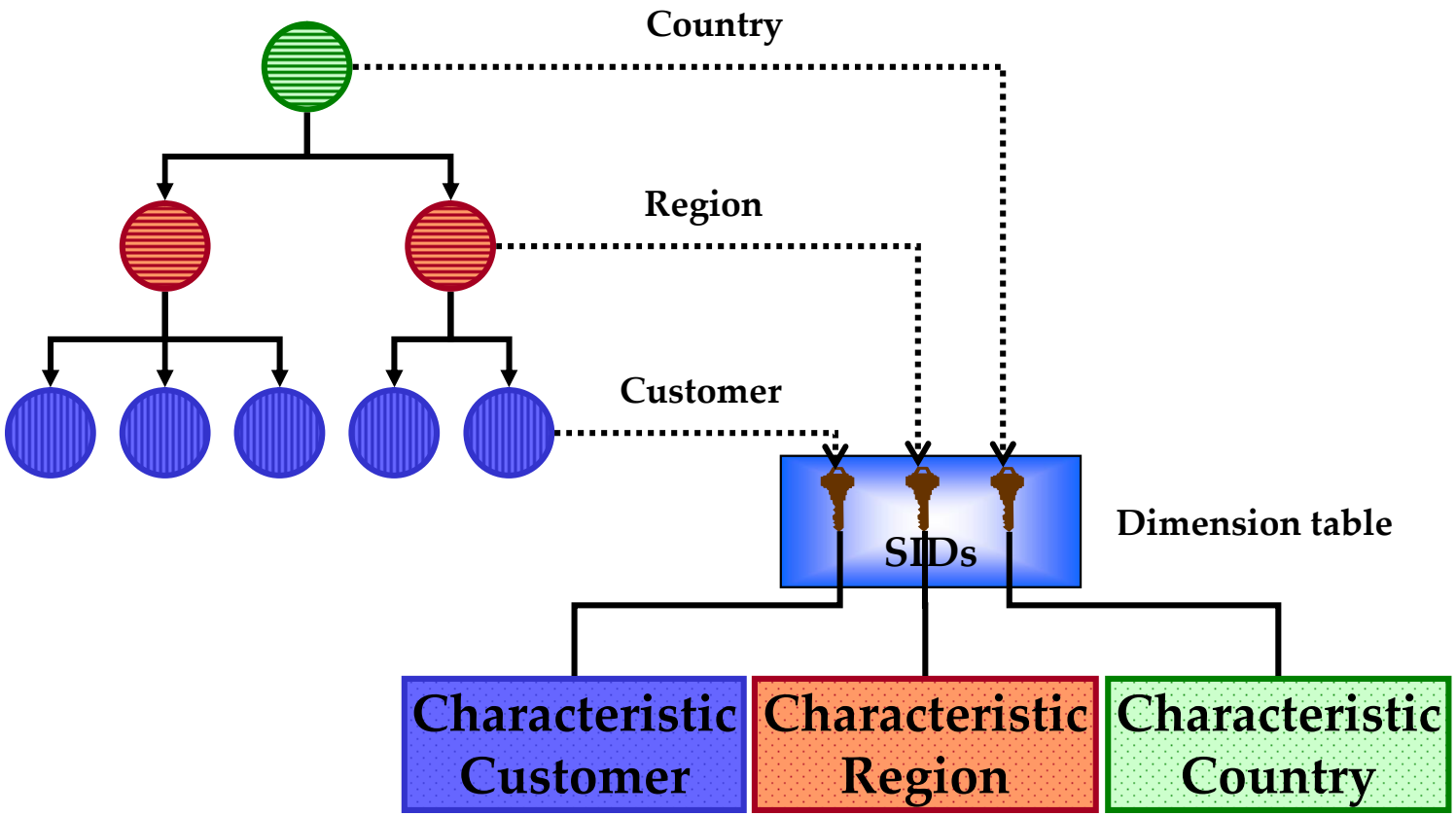
Unbalanced hierarchies



Attributes


Attributes of Dimension
Dimension time
Attributes of all dimension elements
Attributes of level year
year
Attributes of elements of level year
Attributes of level month
month
Attributes of elements of level month
Attributes of level day
day
Attributes of elements of level day

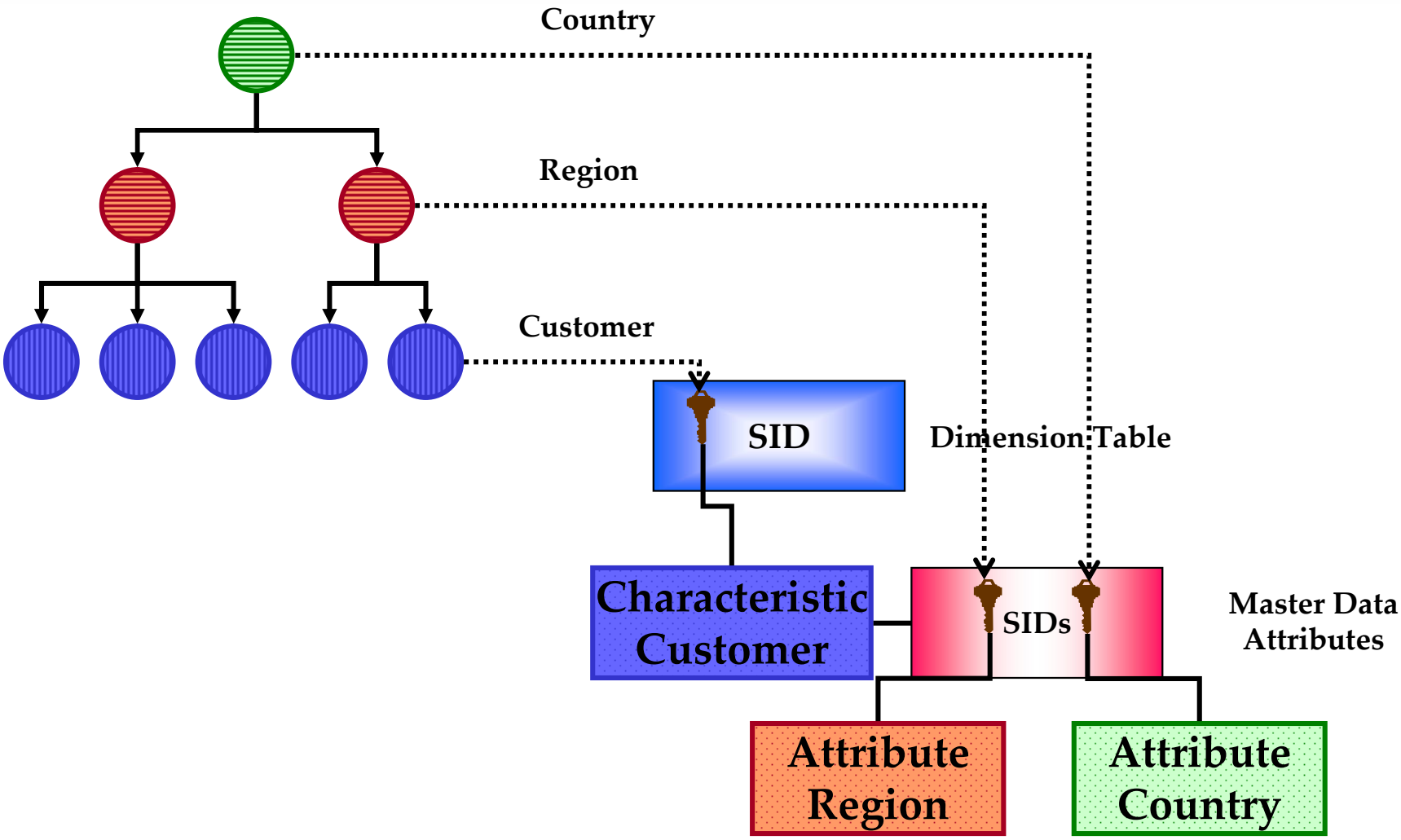
Hierarchies within a dimension via characteristics



Hierarchies within a dimension via characteristics

- Each level is represented by an InfoObject → number of levels should be fixed
- Generally faster than attributes and external hierarchies
- Include the higher hierarchical levels to aggregates
- No predefined drill-down paths

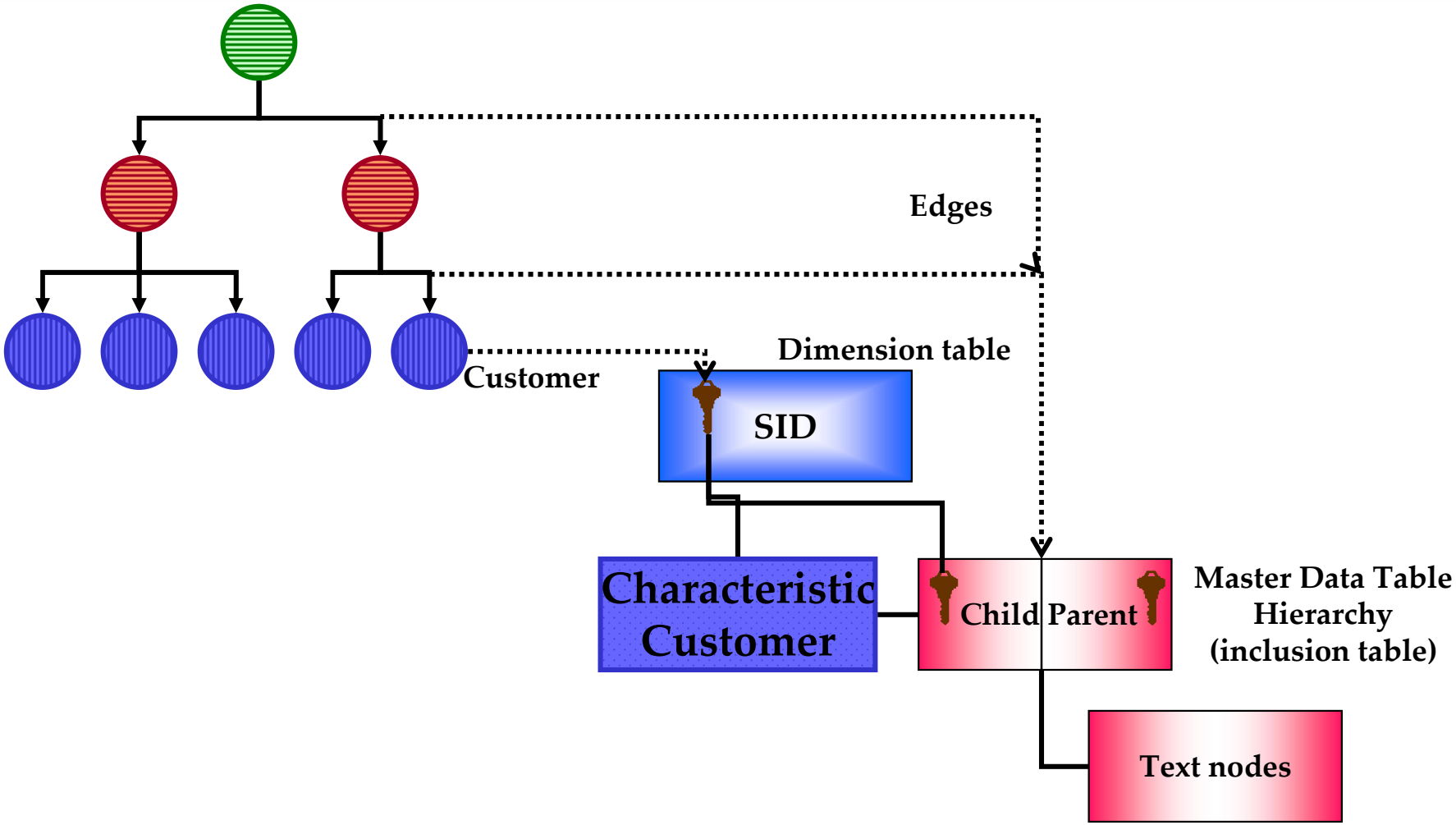
Navigational attributes as basis of hierarchical structures



Navigational attributes as the basis of hierarchical structures

- Each level is represented by an InfoObject → number of levels should be fixed
- Include the higher hierarchical levels to aggregates
- No predefined drill-down paths
- Poor performance without aggregates
- Increased flexibility for reorganization

External hierarchies in BW



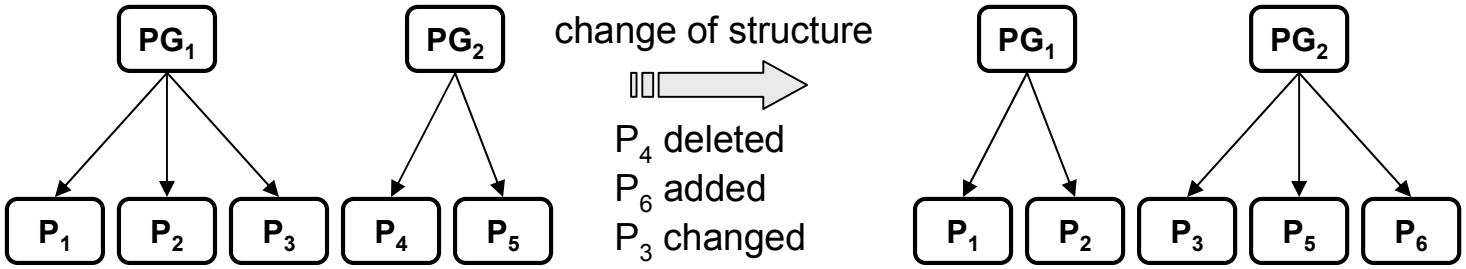
External hierarchies in BW

- Reasonable alternative when there are frequent changes to dimension structure
- Enables unbalanced structures
- Several hierarchies per Info Object are possible
- Poor performance, similar to navigational attributes
- Problems with big hierarchies
- In the case of time dependency, only time stamping the entire structure will enable aggregates

Agenda: Modeling hierarchies in SAP BW

- Variants for modeling hierarchical dimension structures
- Temporal aspects and time stamping
- Modeling guidelines

Time dependency: changes to consolidation trees



Example of "Slowly Changing Dimensions"

Product dimension in 2005-04

Product	Productgroup
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Product dimension in 2005-05

Product	Productgroup
P A	PG X
P B	PG Y (changed)
P C	PG Y
P D	PG Y
P E	PG Y (new)

Fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Reporting requirements - Scenarios

Reporting scenario: current structure

Productgroup	Rev. 2005-04	Rev. 2005-05
PG X	100	100
PG Y	300	400

Reporting scenario: old structure

Productgroup	Rev. 2005-04	Rev. 2005-05
PG X	200	200
PG Y	200	200

Reporting scenario: “historical truth”

Productgroup	Rev. 2005-04	Rev. 2005-05
PG X	200	100
PG Y	200	400

Reporting scenario: comparable results

Productgroup	Rev. 2005-04	Rev. 2005-05
PG X	100	100
PG Y	200	200

Scenario I : Report with current structure

Product dimension in 2005-05

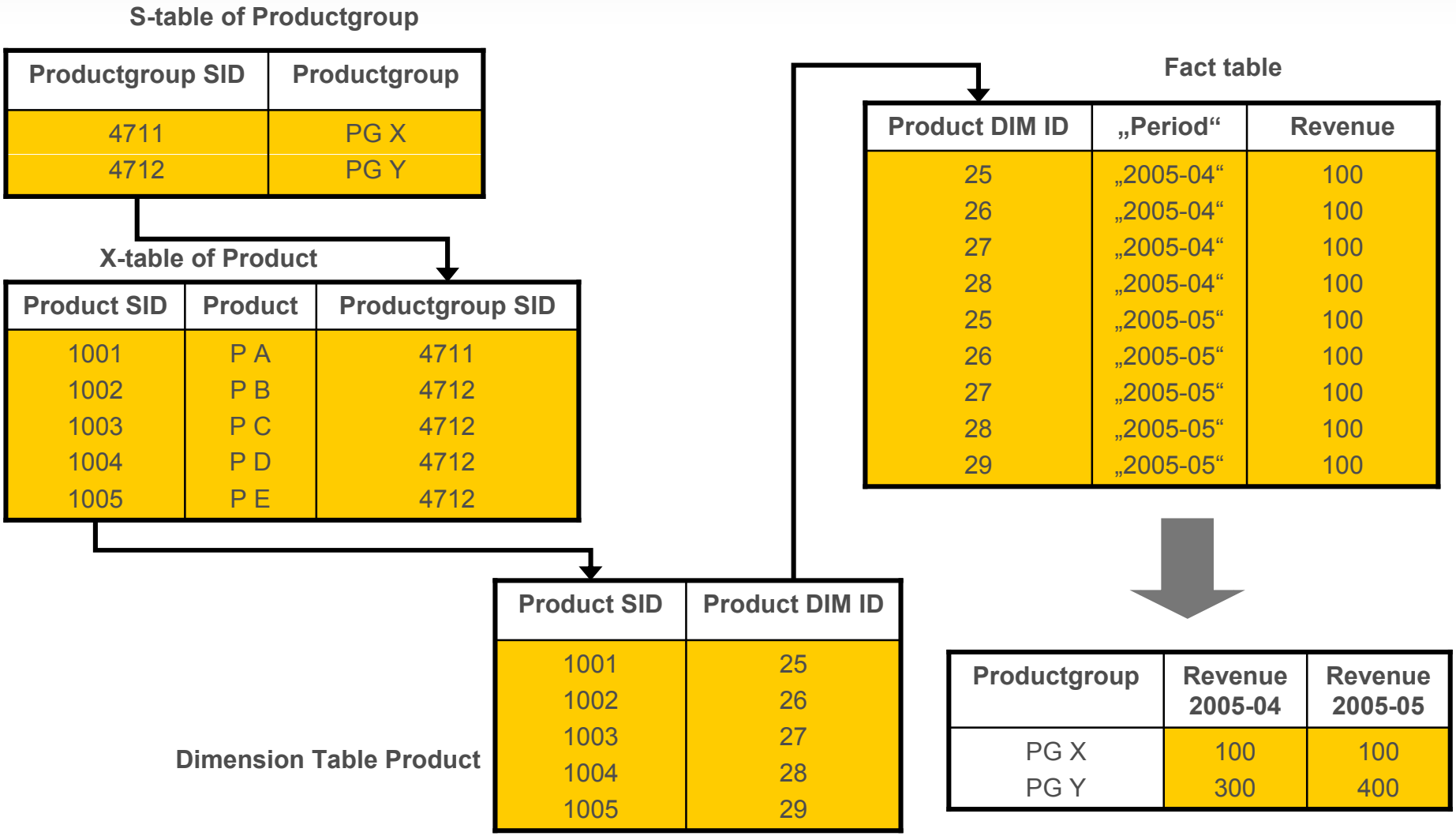
Product	Productgroup
P A	PG X
P B	PG Y (changed)
P C	PG Y
P D	PG Y
P E	PG Y (new)

Fact table

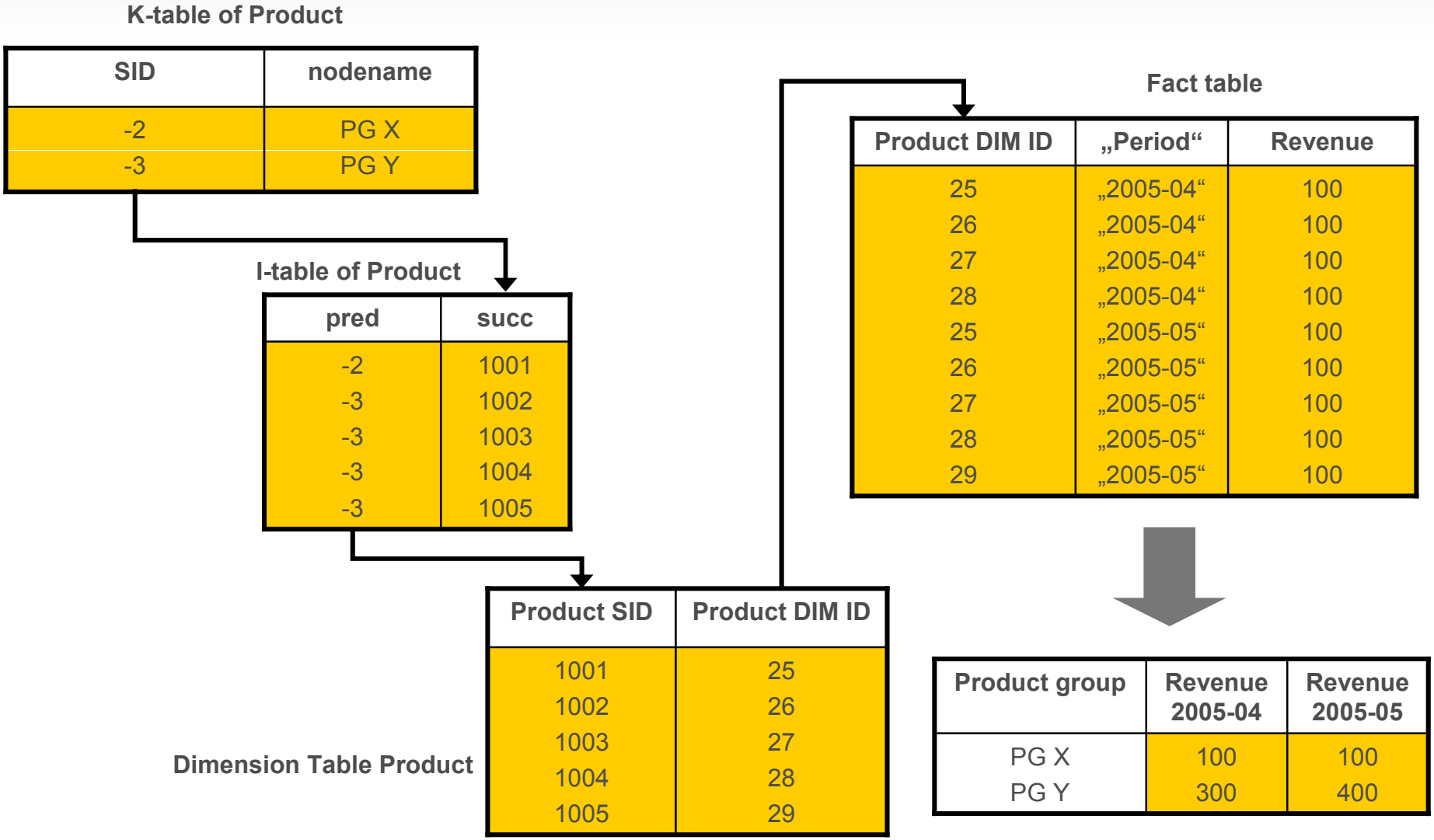
Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	100	100
PG Y	300	400

Query path "current structure" with navigational attributes



Query path "current structure" with external hierarchy



Scenario II : Report with old structure

Product dimension in 2005-04

Product	Productgroup
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Productgroup	Rev. 2005-04	Rev. 2005-05
PG X	200	200
PG Y	200	200

Query path "old structure" with time-dependent navigational attributes

S-table of productgroup

Productgroup SID	Productgroup
4711	PG X
4712	PG Y

Query key date 2005-04

Y-table of product

Product SID	Product	Productgroup SID	DateFrom	DateTo
1001	P A	4711	1000-01	9999-12
1002	**P B**	**4711**	**1000-01**	**2005-04**
1002	**P B**	**4712**	**2005-05**	**9999-12**
1003	P C	4712	1000-01	9999-12
1004	P D	4712	1000-01	9999-12
1005	**P E**	**4712**	**2005-05**	**9999-12**

Dimension Table Product

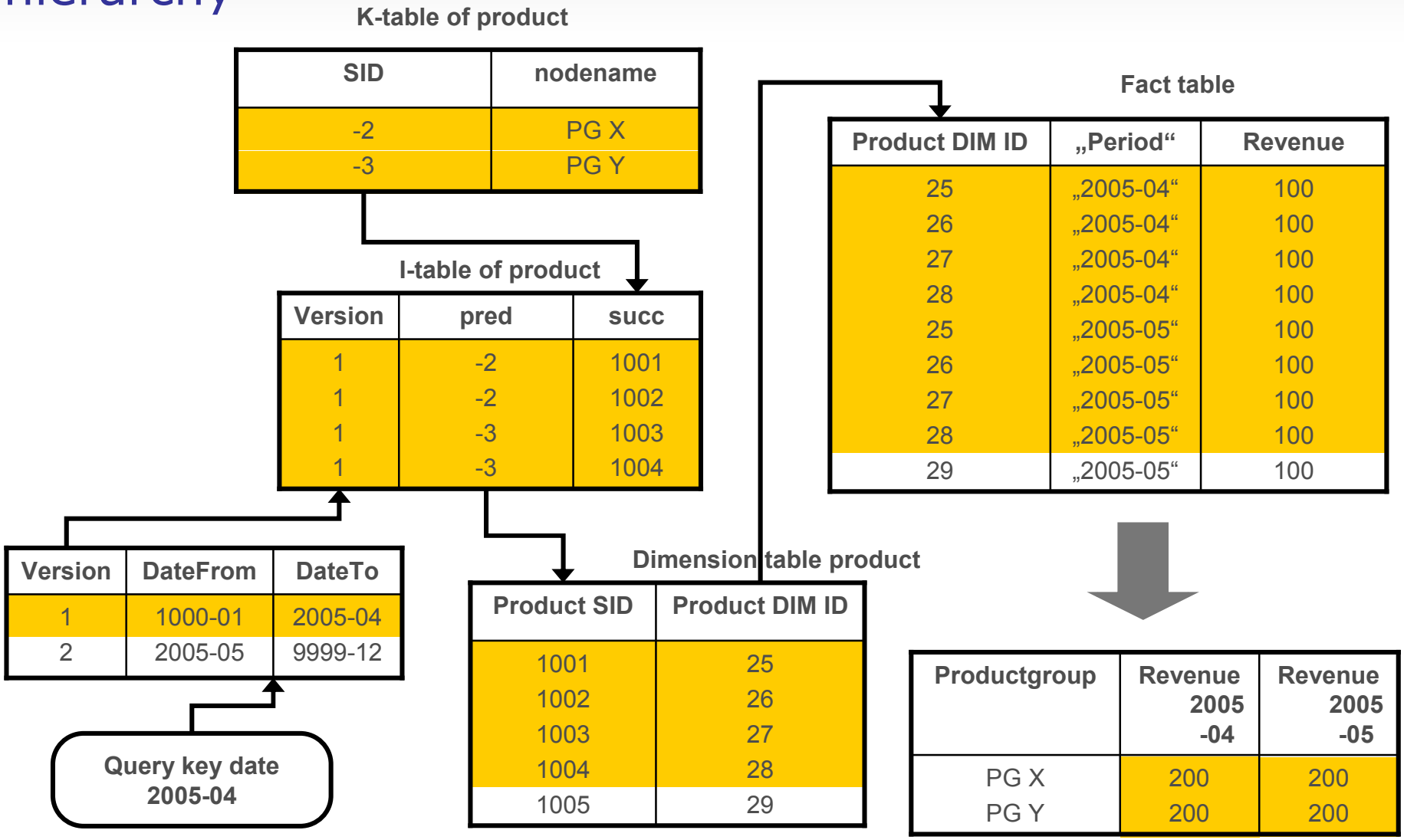
Product SID	Product DIM ID
1001	25
1002	26
1003	27
1004	28
1005	29

Fact table

Product DIM ID	„Period“	Revenue
25	„2005-04“	100
26	„2005-04“	100
27	„2005-04“	100
28	„2005-04“	100
25	„2005-05“	100
26	„2005-05“	100
27	„2005-05“	100
28	„2005-05“	100
29	„2005-05“	100

Productgroup	Revenue 2005-04	Revenue 2005-05
PG X	200	200
PG Y	200	200

Query path "old structure" with time-dependent hierarchy



Scenario III : Report with "historical truth "

Product dimension in 2005-04

Product	Productgroup
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Product dimension in 2005-05

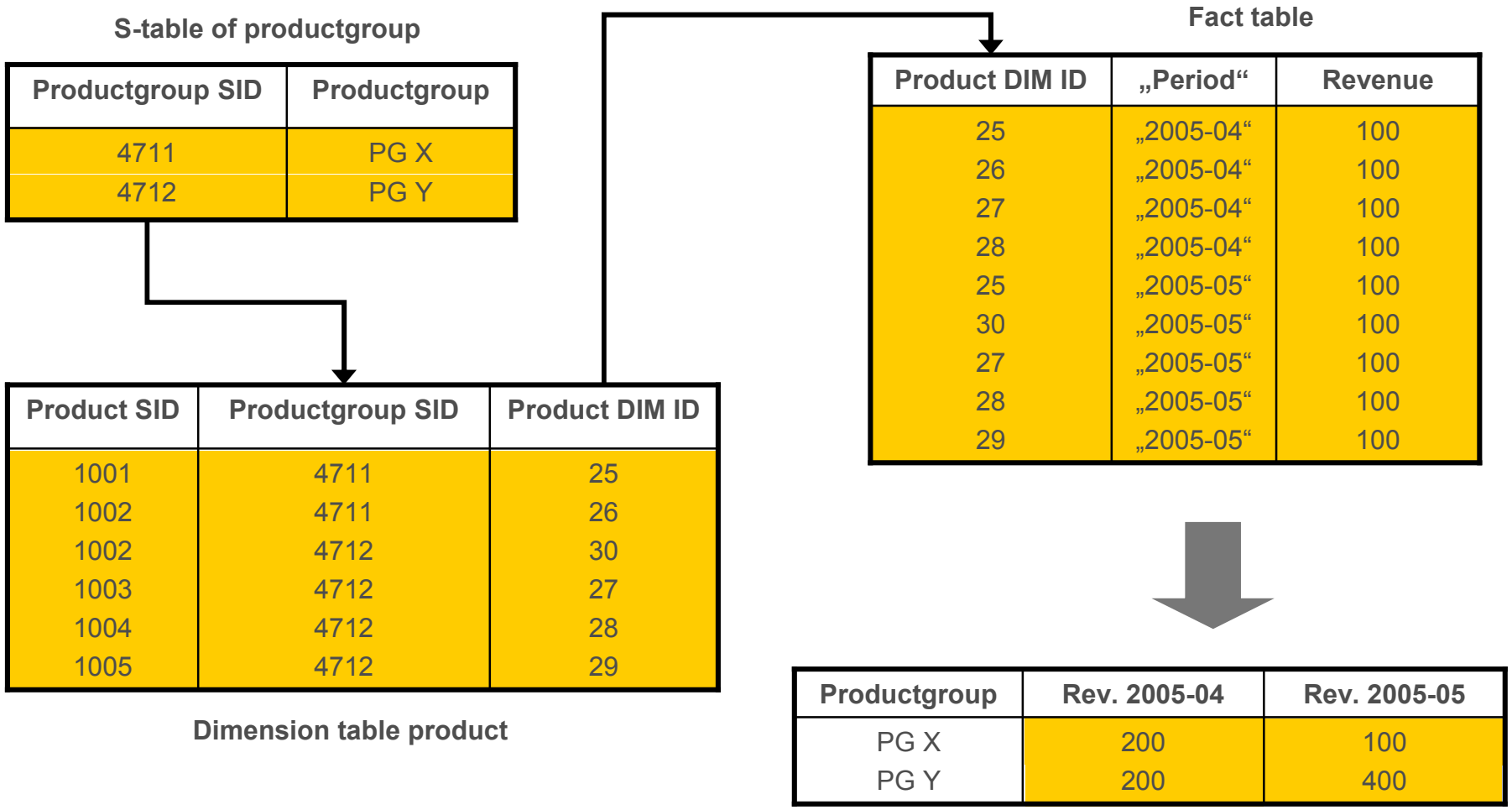
Product	Productgroup
P A	PG X
P B	PG Y (changed)
P C	PG Y
P D	PG Y
P E	PG Y (new)

Fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Productgroup	Rev. 2005-04	Rev, 2005-05
PG X	200	100
PG Y	200	400

Query path "historical truth" with characteristics



Scenario IV : Report with Comparable Results

Product dimension in 2005-04

Product	Productgroup
P A	PG X
P B	PG X
P C	PG Y
P D	PG Y

Product dimension in 2005-05

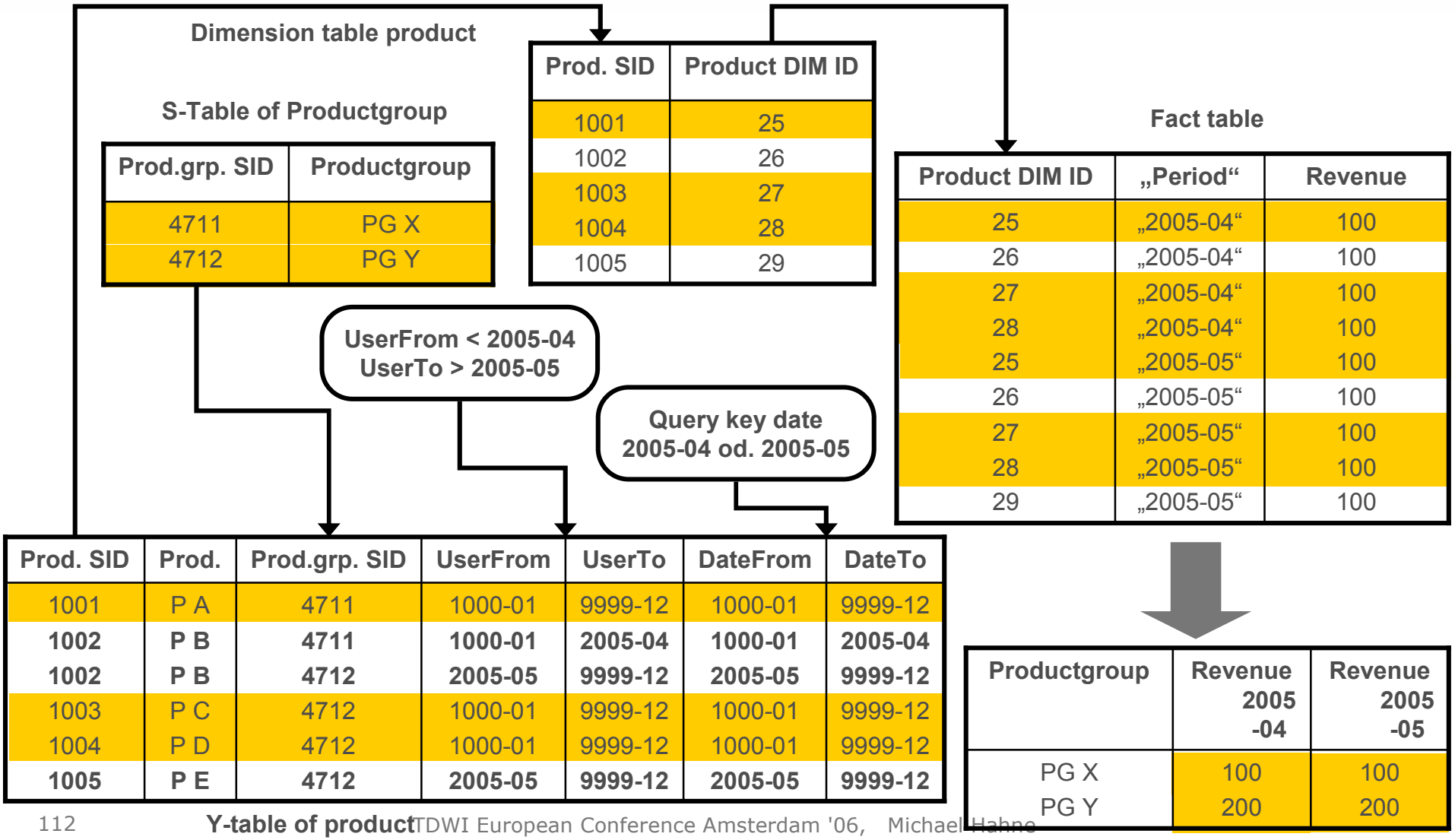
Product	Productgroup
P A	PG X
P B	PG Y (changed)
P C	PG Y
P D	PG Y
P E	PG Y (new)

Fact table

Product	Period	Revenue
P A	2005-04	100
P B	2005-04	100
P C	2005-04	100
P D	2005-04	100
P A	2005-05	100
P B	2005-05	100
P C	2005-05	100
P D	2005-05	100
P E	2005-05	100

Productgroup	Rev. 2005-04	Rev. 2005-05
PG X	100	100
PG Y	200	200

Query path "comparable results" with time-dependent navigational attributes



Agenda: Modeling hierarchies in SAP BW

- Variants for modeling hierarchical dimension structures
- Temporal aspects and time stamping
- Modeling guidelines

Modeling Guidelines

- Modeling of dimensions
- Design of InfoProvider

Guidelines for conceptual modeling of dimensions

- Number of dimensions should be between four and ten (optimally between six and eight)
- Number of hierarchy levels (seven at most)
- Number of elements per consolidation element (a maximum of fifteen to twenty elements is advisable)
- Determination of dimensions:
  - 1:1 relationship unsuitable (-> attributes)
  - 1:N relationship determines dimension hierarchy
  - M:N relationship rather two different dimensions

Guidelines for logical modeling of dimensions

- Model characteristics with high cardinality as line-item dimensions
- Attributes that change frequently should be modeled with their own dimensions (use line item where possible!)
- Group characteristics with very low cardinality (e.g. scenario) into one dimension, in order to reduce the number of dimensions and to restrict the number of dimensions to 16 at most
- Distribute characteristics of a hierarchy that have high cardinality into separate dimensions (parent characteristics in their own dimension)

Criteria to help decision-making in logical modeling of dimension structures in BW

- Versioning
- Scope
- Performance
- Navigational paths
- Unbalanced dimension structures
- Leaves with multiple parent elements
- Structural changes and reorganization

Hierarchy Guidelines: Versioning

External hierarchy	Hierarchy within a Dimension (characteristics)	Hierarchy defined by navigational attributes
• Transactional view isn't possible (no "as posted") • Different types of views are possible (hierarchy versions and time-dependent hierarchies)	• Only the transactional view ("as posted") is possible	• Transactional view isn't possible (no "as posted") • Time-dependent attributes enable different views

Hierarchy Guidelines: Scope

External hierarchy	Hierarchy within a Dimension (characteristics)	Hierarchy defined by navigational attributes
• Hierarchy is part of master data and is valid for each InfoCube in the system (where the underlying InfoObject is used)	• Only valid in the InfoCube	• Hierarchy is part of master data and is valid for each InfoCube in the system (where the underlying InfoObject is used)

Hierarchy Guidelines: Performance

External hierarchy	Hierarchy within a Dimension (characteristics)	Hierarchy defined by navigational attributes
• Aggregates should be used for good query performance	• Good performance (even without aggregates)	• Aggregates should be used for good query performance

Hierarchy Guidelines: Navigational paths

External hierarchy	Hierarchy within a Dimension (characteristics)	Hierarchy defined by navigational attributes
• Drill-down path is predefined by the structure of the consolidation tree	• Levels can be skipped because there isn‘t a predefined drill-down path (all characteristics in a dimension are equal)	• Levels can be skipped because there isn‘t a predefined drill-down path (all navigational attributes of a characteristic are equal)

Hierarchy Guidelines: Unbalanced dimension structure

External hierarchy	Hierarchy within a Dimension (characteristics)	Hierarchy defined by navigational attributes
• Unbalanced hierarchies are possible	• Each characteristic corresponds to a certain level of the hierarchy, therefore only balanced structures are possible	• Each characteristic corresponds to a certain level of the hierarchy, therefore only balanced structures are possible

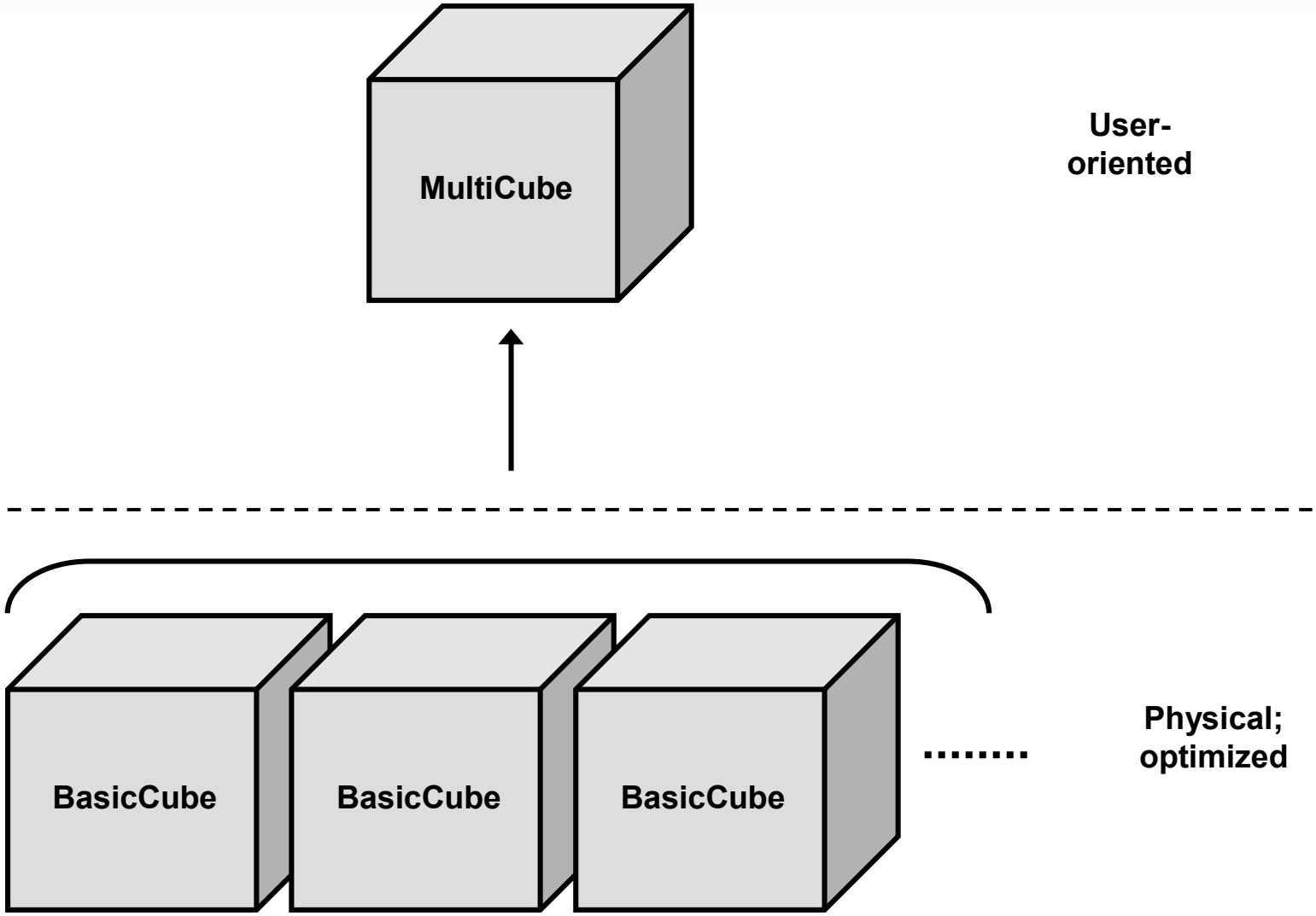
Hierarchy Guidelines: Leaves with multiple parent elements

External hierarchy	Hierarchy within a Dimension (characteristics)	Hierarchy defined by navigational attributes
• Many-many relationships between the levels of the hierarchy are possible and consolidated correctly	• Many-many relationships between hierarchy levels are only possible as they are defined by the transactions • (“as posted“ view)	• Many-many relationships between levels are impossible

Hierarchy Guidelines: Structural changes and reorganization

External hierarchy	Hierarchy within a Dimension (characteristics)	Hierarchy defined by navigational attributes
• Quick changes and reorganization possible	• Reloading cube(s) is required for reorganization	• Reorganization is possible (additional attributes and/or changes of master data)

Two-layer concept of Cube Modeling



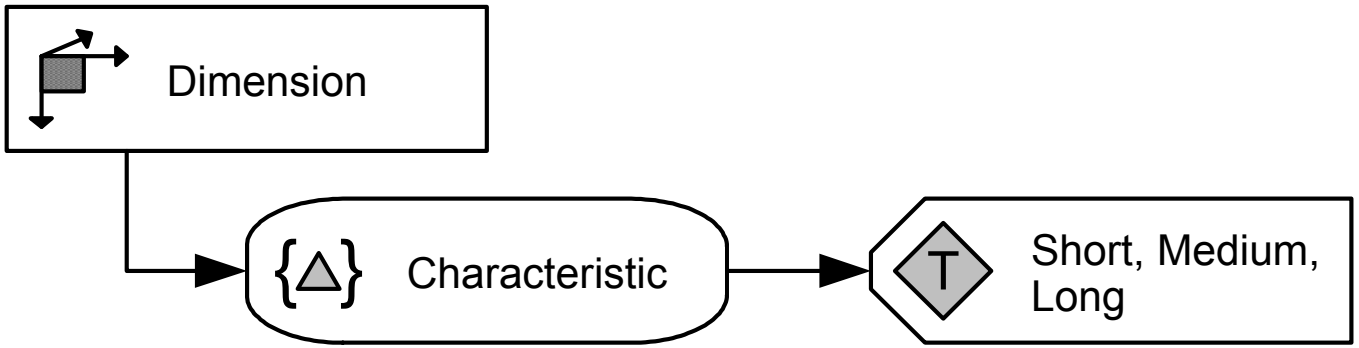
Agenda: Model Management

- Graphical model representation with Visio™
- Metadata in Data Warehousing
- Common Warehouse Metamodel (CWM)
- Implementation of the CWM in BW

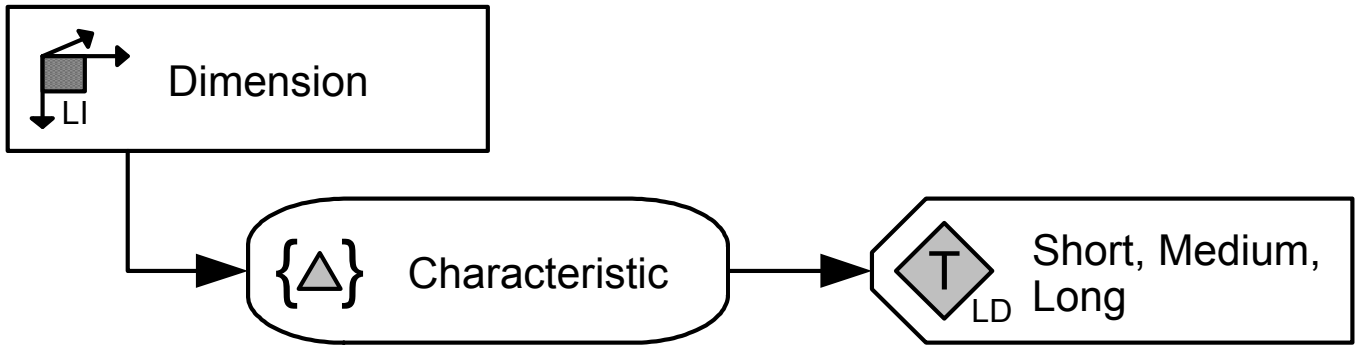
Agenda: Model Management

- Graphical model representation with Visio™
- Metadata in Data Warehousing
- Common Warehouse Metamodel (CWM)
- Implementation of the CWM in BW

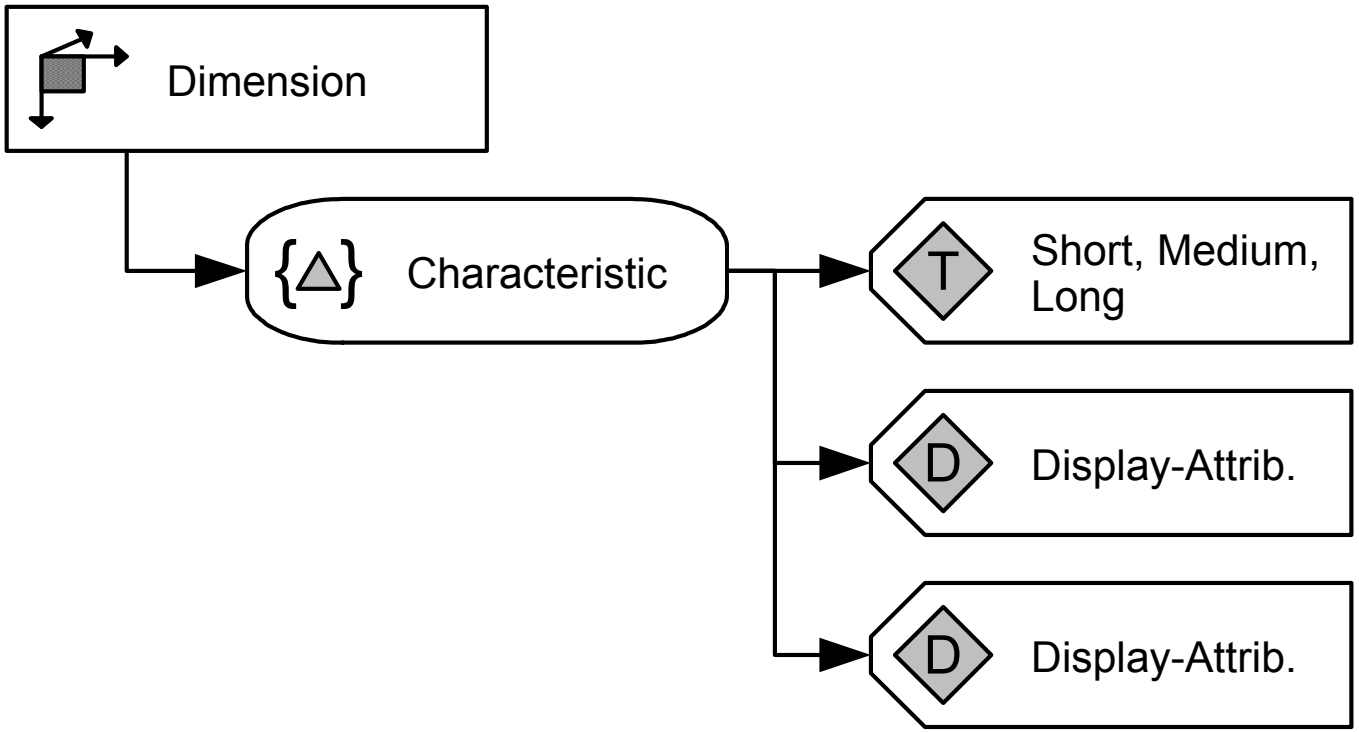
Dimension with one characteristic



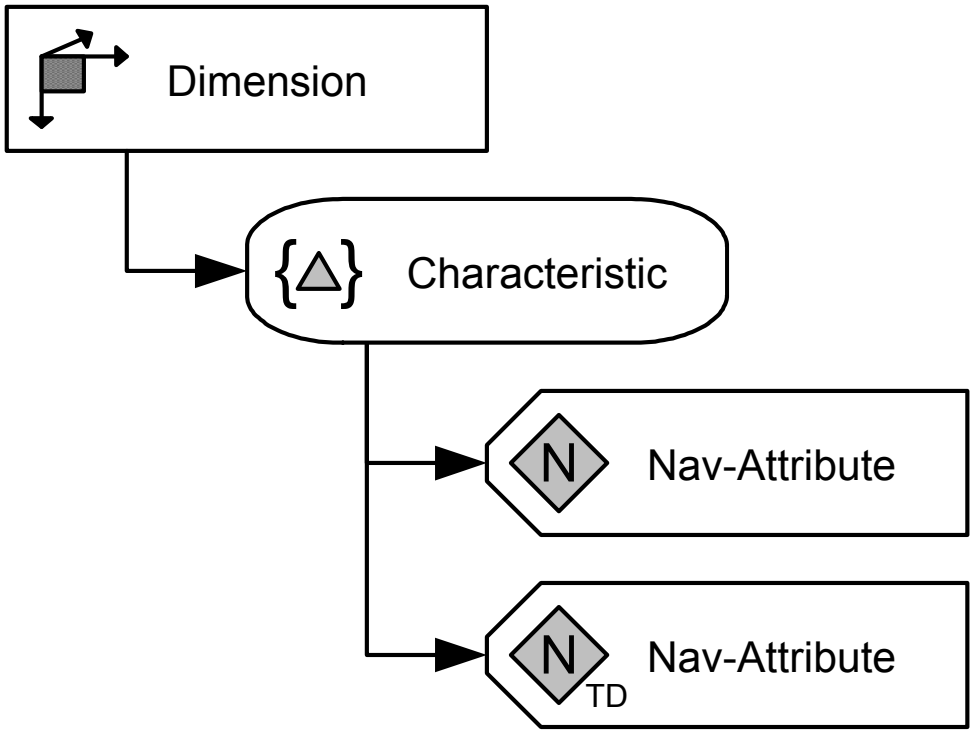
Line-Item-Dimension



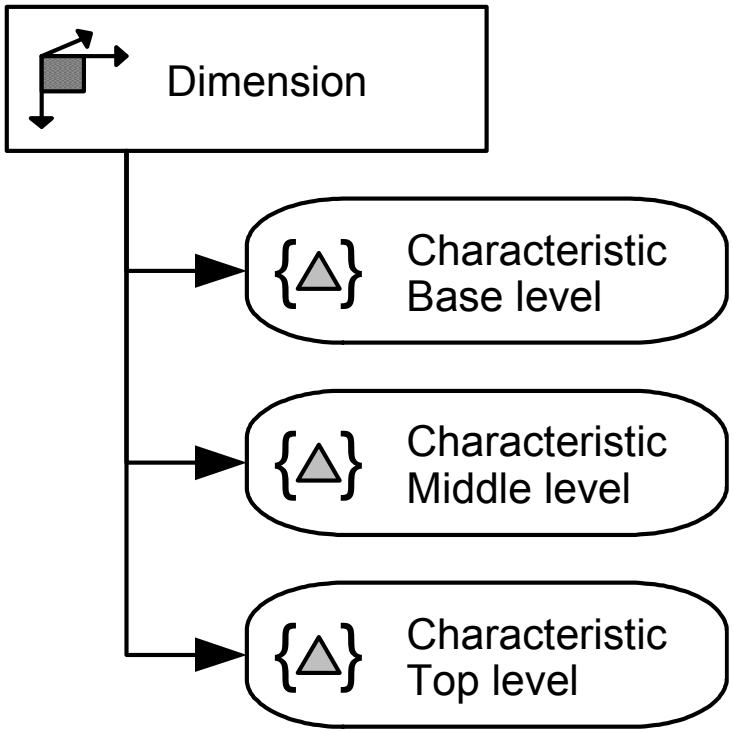
Display attributes of a characteristic



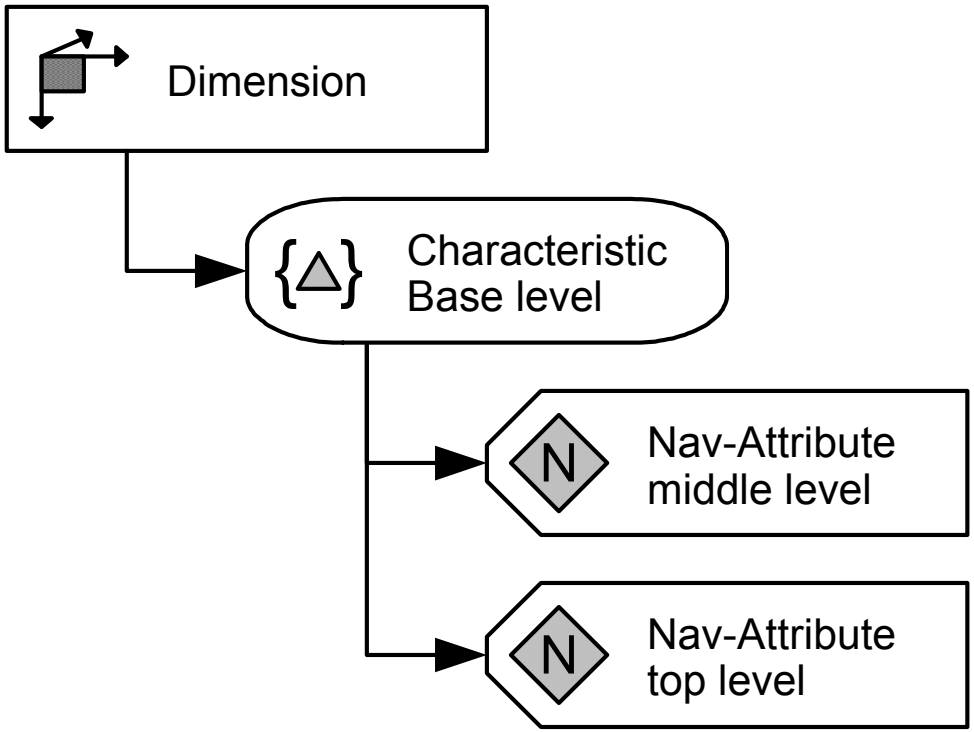
Navigational attributes of a characteristic



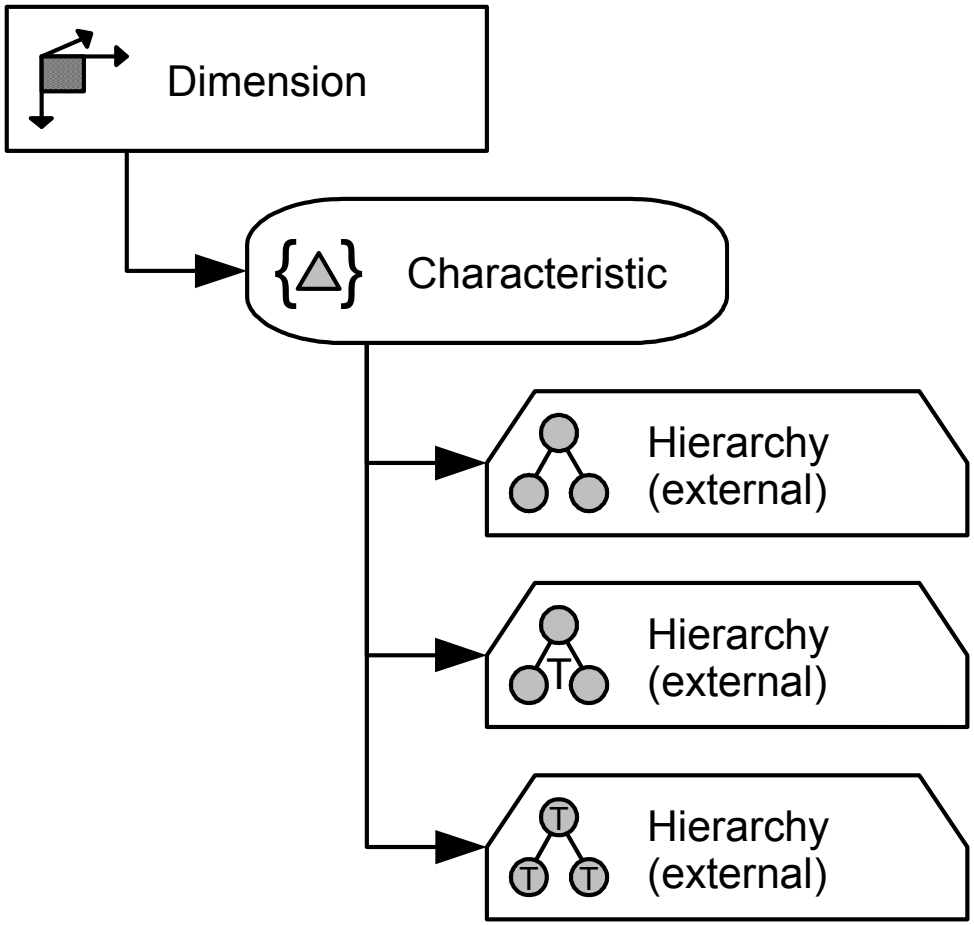
Hierarchy within a dimension with characteristics



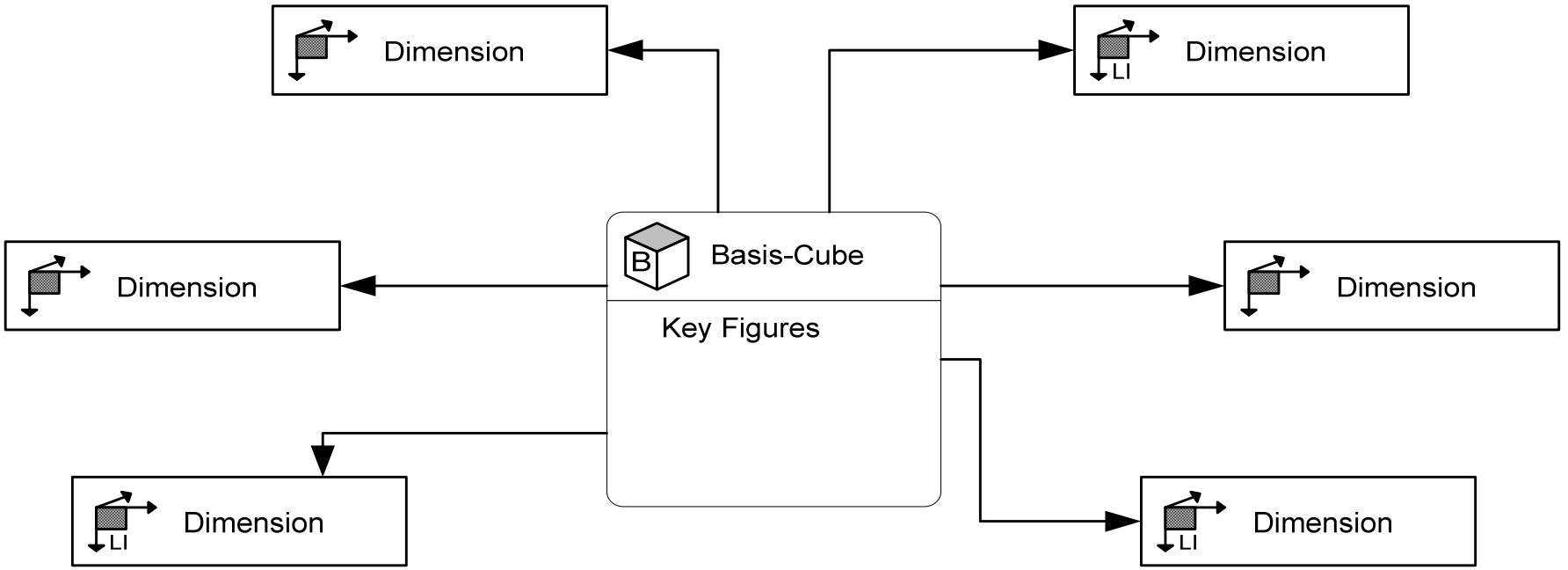
Hierarchical structure with navigational attributes



External hierarchy



Modeling of BasicCubes



Agenda: Model Management

- Graphical model representation with Visio™
- Metadata in Data Warehousing
- Common Warehouse Metamodel (CWM)
- Implementation of the CWM in BW

Definition of metadata

- Concise:
"**Data about data**"
not sufficiently precise

- More precise:
"**Data that describes the meaning and structure of business data, as well as how it is created, accessed and used**"
[Devlin, 1997]

- Commonly used:
"**Each kind of information that is required for planning, building and running information systems**"
[Bauer/Günzel, 2001]

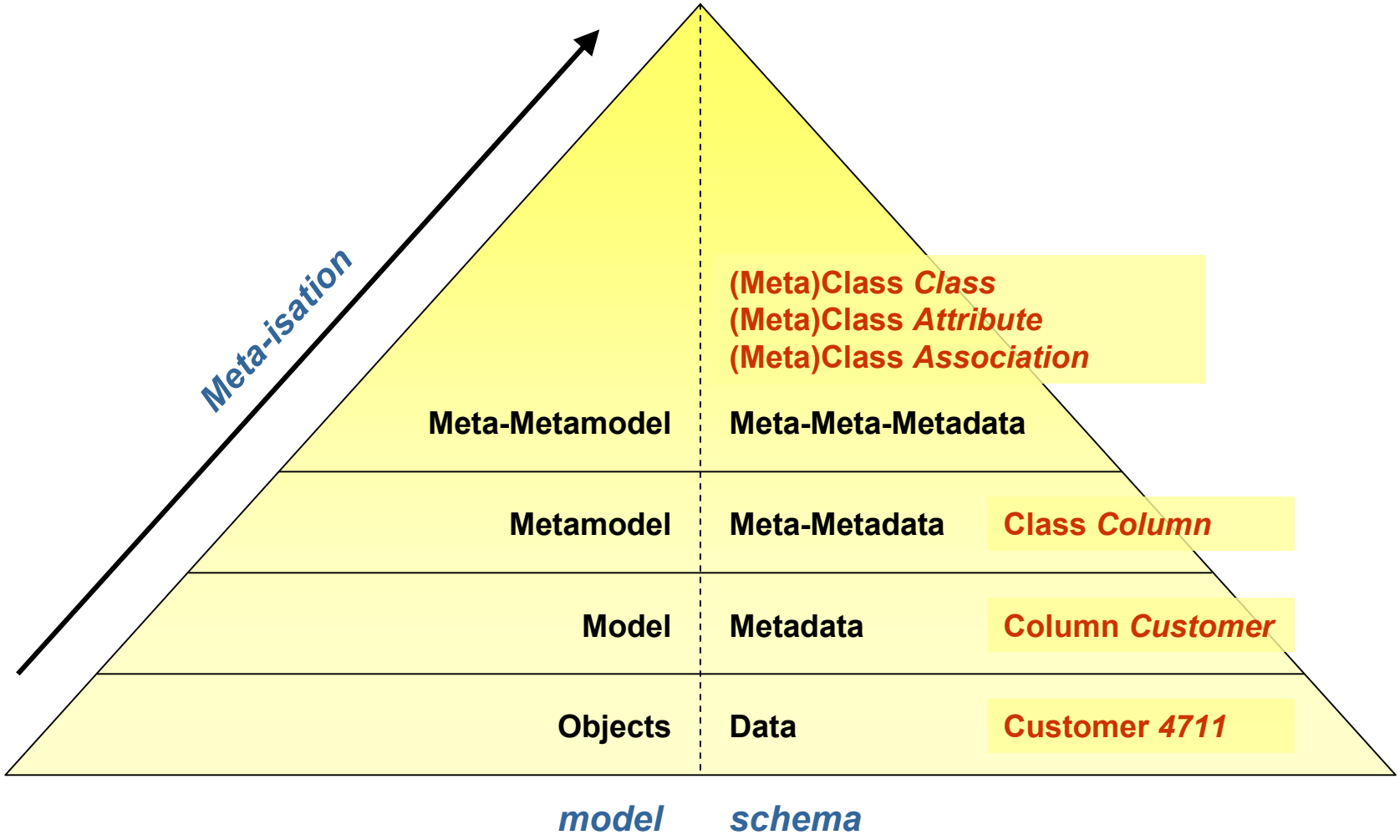
Models

- Constructivist approach to modeling:
Models are the result of a process of modeling that generates a "mini-world" by abstracting from the real world

- Classification:
Models for description, explanantion and decision making

- Sometimes confused usage of the word "model":
sometimes: language of modeling (**model**)
other times: concrete instance (**schema**)

Abstraction layers of modeling



Meta-isation

- Metamodels:
  When models and modeling itself become the subject of modeling
- Further abstraction:
  Proceed to models that focus on the description of metamodels
- This process is called meta-isation and can be continued indefinitely
- Objective of creating a model:
  Representing an area of interest in a descriptive model, by means of language.

Relevance of metadata in Data Warehousing

- For end users: permits use of information in the Data Warehouse in a relatively autonomous way
  - Definition of used business terms (glossary)
  - Mapping of business terms to data objects in the Data Warehouse
  - Description of reports, contact persons, access rights
- For development and administration: technically oriented metadata
  - Data sources, rules for improving data quality
  - Transformation rules, steps of data consolidation
  - Mapping of data sources to data warehouse models

Relevance of metadata in Data Warehousing

- „best-of-breed“ vs. „end-to-end“
  - Heterogenous complex system landscape → uniform Data Warehouse solution
  - Improved Interoperability by means of meta data exchange
- Meta data management is crucial to improving efficiency and effectivity of a data warehouse solution

Classification of meta data

- DSS vs. operativ:
  - operational meta data:
    - Source systems
    - data structures
    - field definitions
    - source data structures
    - transformation process
    - target system
  - DSS meta data:
    - Data storage
    - Data structures in the data warehouse
    - Objects of the Business Model
    - Navigation possibilities

Classification of meta data

- Main functions of a Data Warehouse:
  - Meta-Data: Meta data management, basis of the other functions
  - Source: Identification of data sources
  - Load: Data transfer and transformation
  - Storage: Storage type, data distribution, archiving
  - Query: Reporting and analysis, OLAP, simulation and Data Mining
- Threedimensional classification:
  - Processes: Design, Populate, Administer, Analyze
  - Users: Technical User oder Business User
  - Data: Operational, Data Warehouse, Data Mart

Agenda: Model Management

- Graphical model representation with Visio™
- Metadata in Data Warehousing
- Common Warehouse Metamodel (CWM)
- Implementation of the CWM in BW

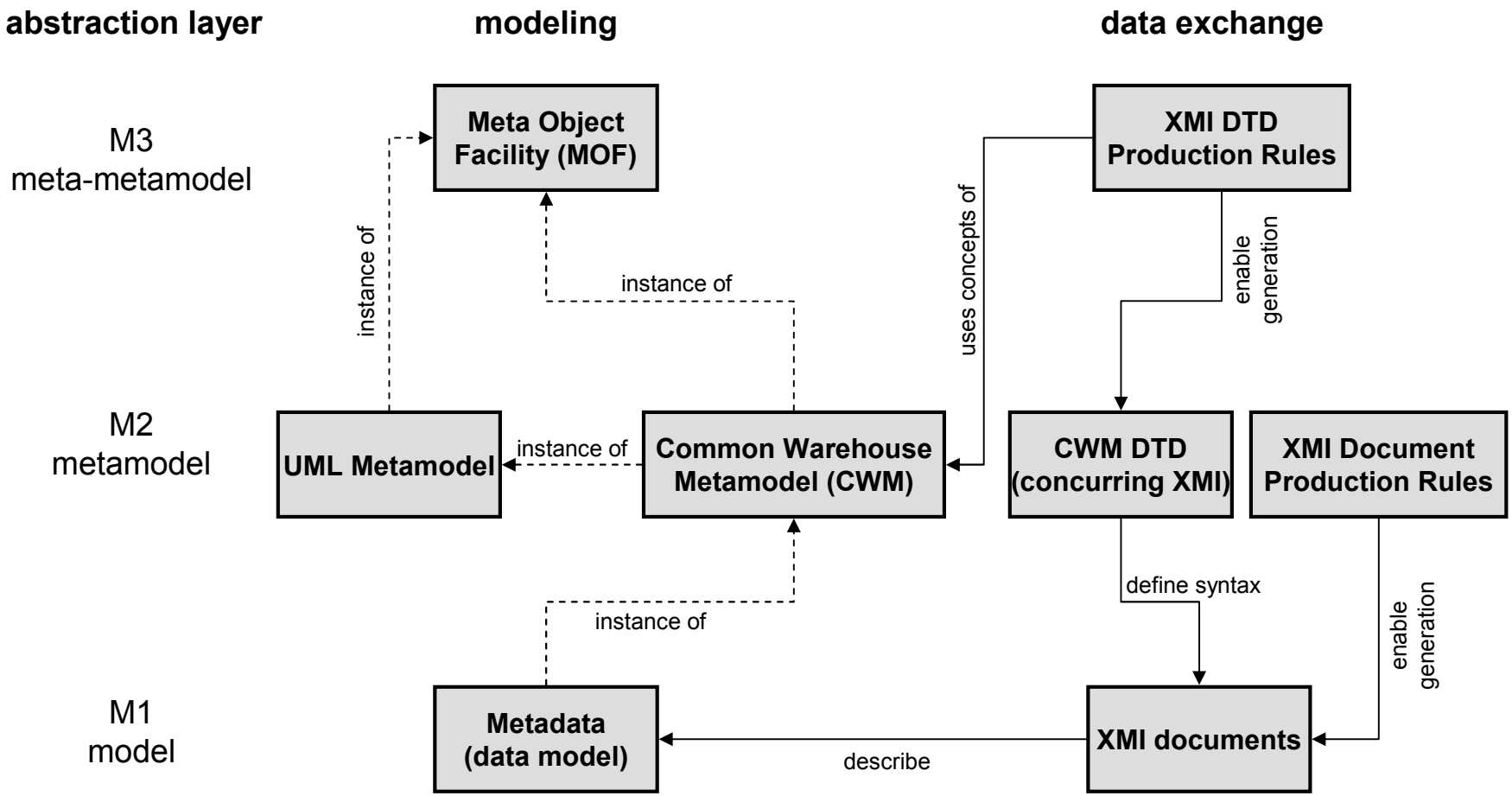
Agenda: Model Management

- Graphical model representation with Visio™
- Metadata in Data Warehousing
- Common Warehouse Metamodel (CWM)
- Implementation of the CWM in BW

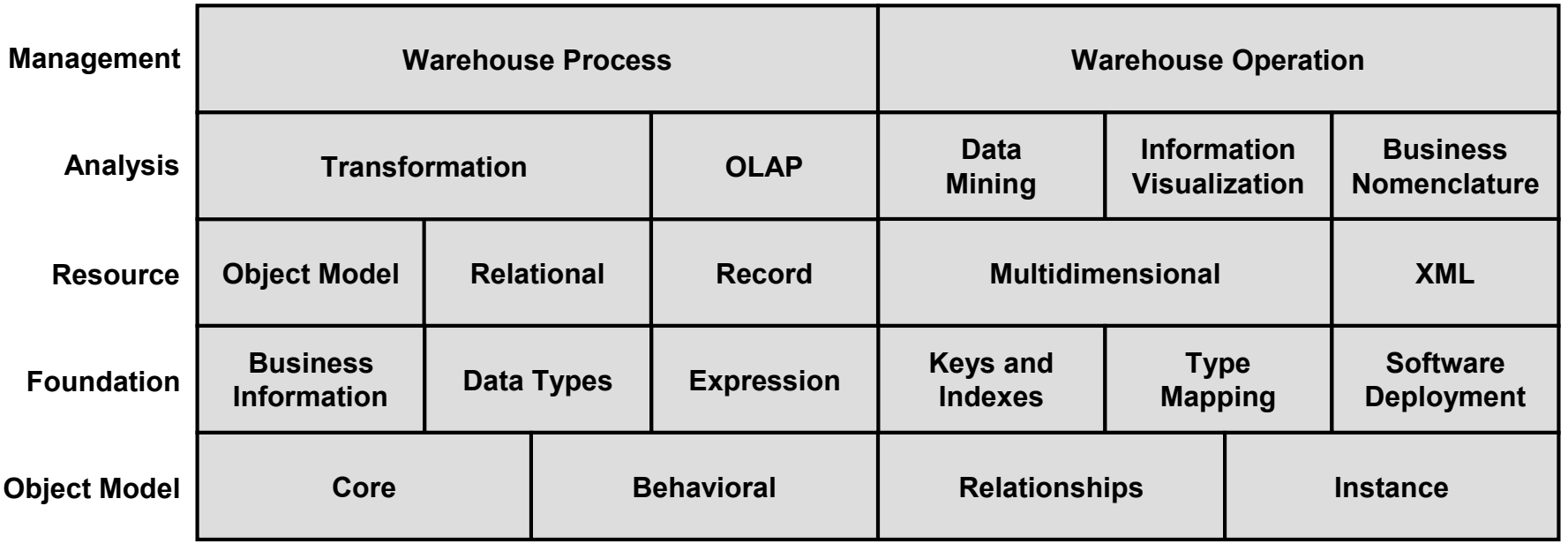
Common Warehouse Metamodel (CWM)

- Model for Data Warehouse metadata
- Developed since 1998: active participation of IBM, NCR, Hyperion Solutions and Oracle among others
- Certified by the Object Management Group (OMG) as a Data Warehouse standard in 2000 for the first time
- Embedded in OMG‘s Model Driven Architecture (MDA)
- Based upon more than 200 Classes und 150 Associations (UML class diagram)
- Metadata exchange via XMI (XML Metadata Interchange) and via CORBA interfaces
- Basic principles: completeness, comprehensibility, independence, compatibility, interoperability, extensibility, availability

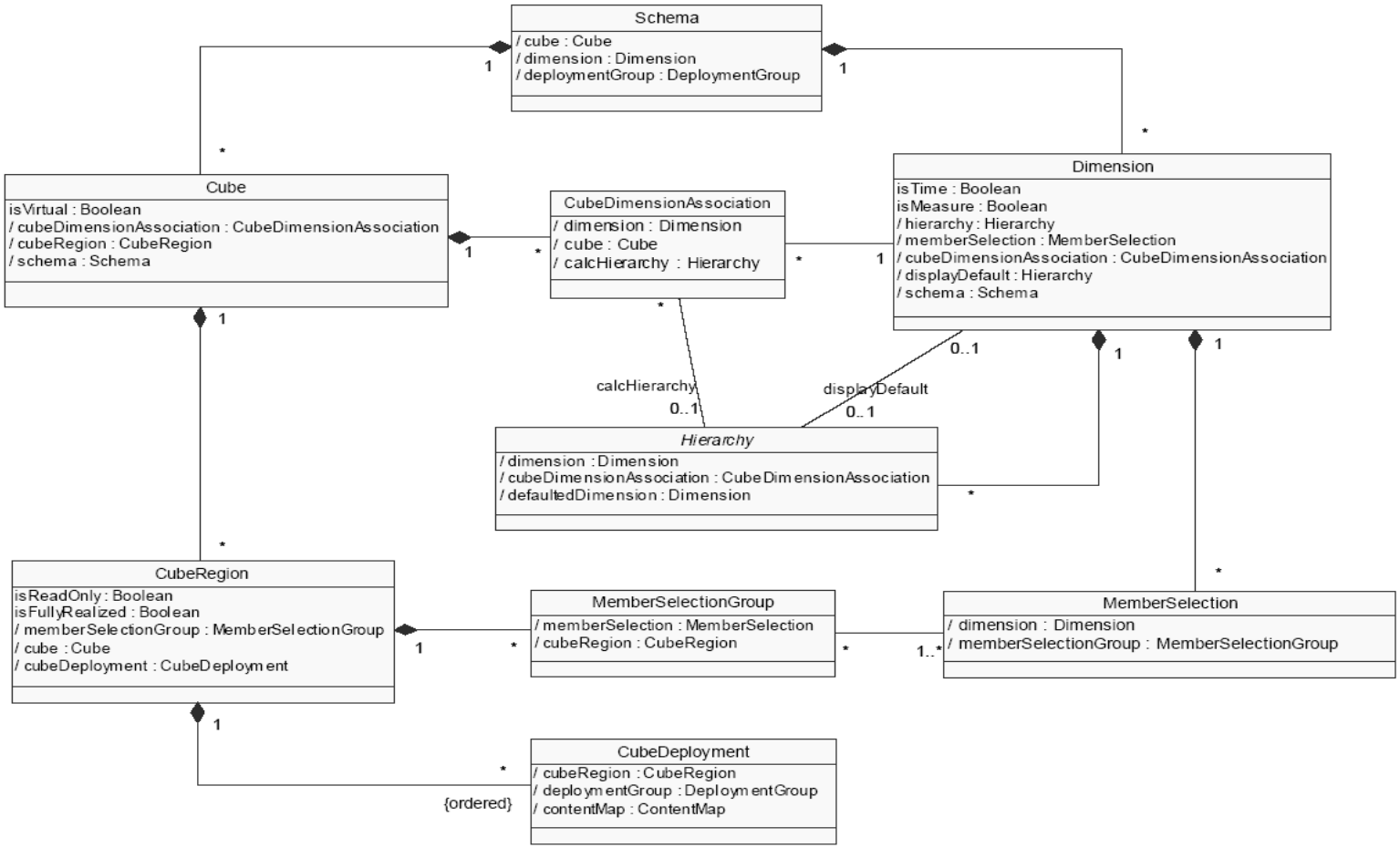
OMG metamodel architecture



CWM metamodel layering

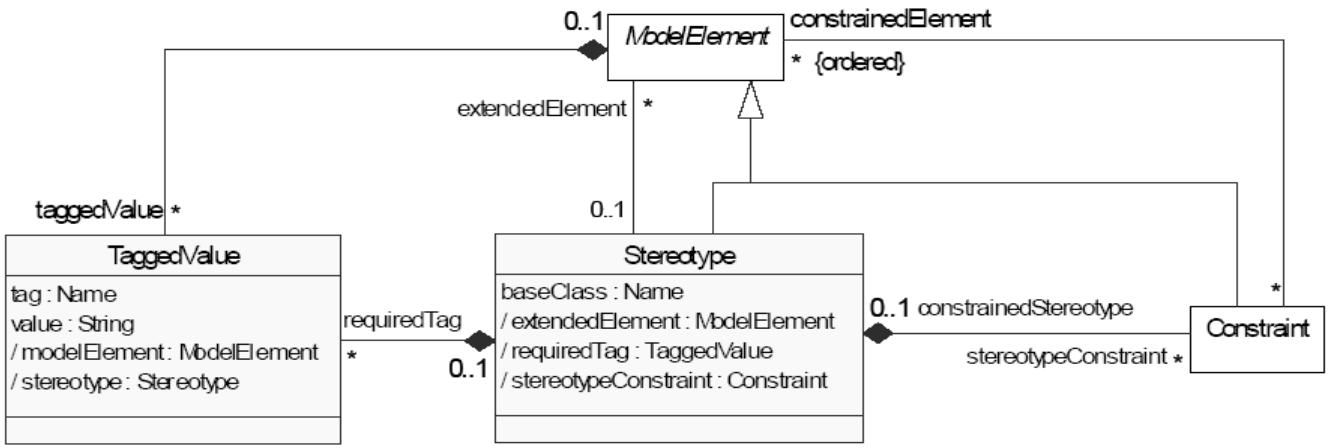


OLAP Metamodel



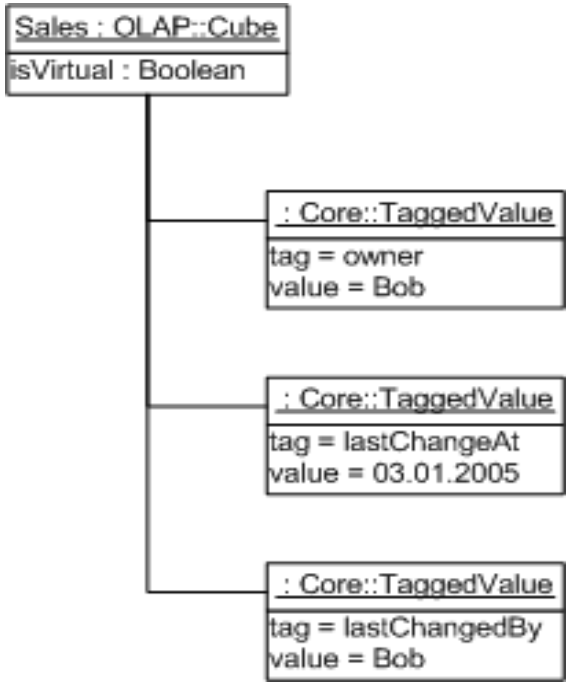
Extending the CWM

- Tagged Values based upon the package *Core*
- Stereotypes also based upon the package *Core*
- Most flexible: Extension classes



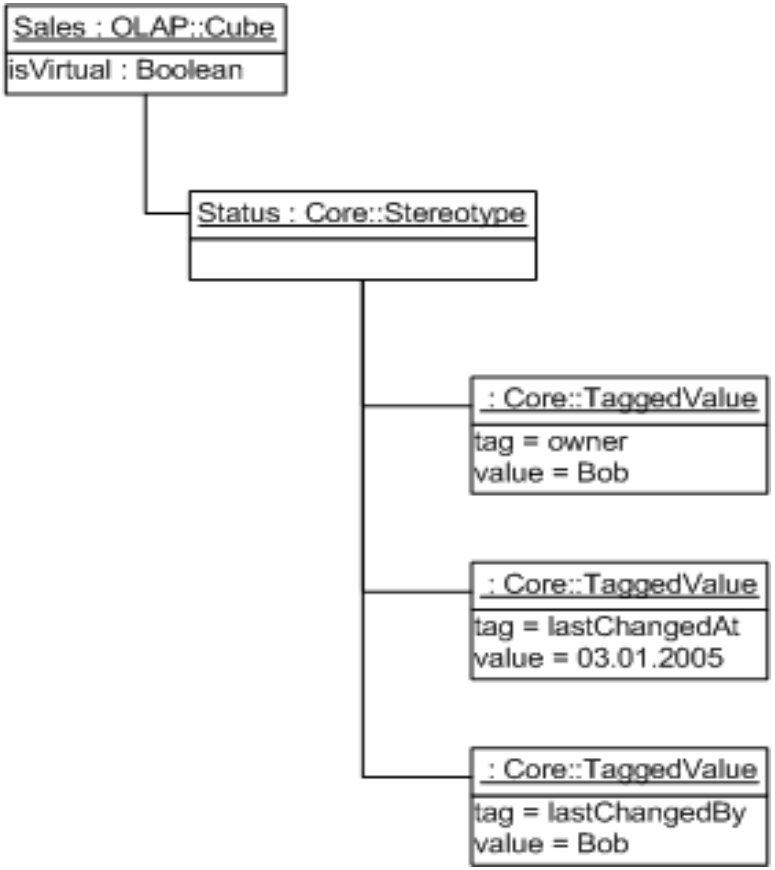
Extending the CWM: Tagged Values

- Tagged Values: Taken from the UML specification (Package *Core*)
- Extension by pairs: Attribute name (tag) = Attribute value (value)
- Independent of release changes in the CWM specification
- Restricted to string values
- No direct access (The user must know about their existence)
- No additional associations in the metamodel



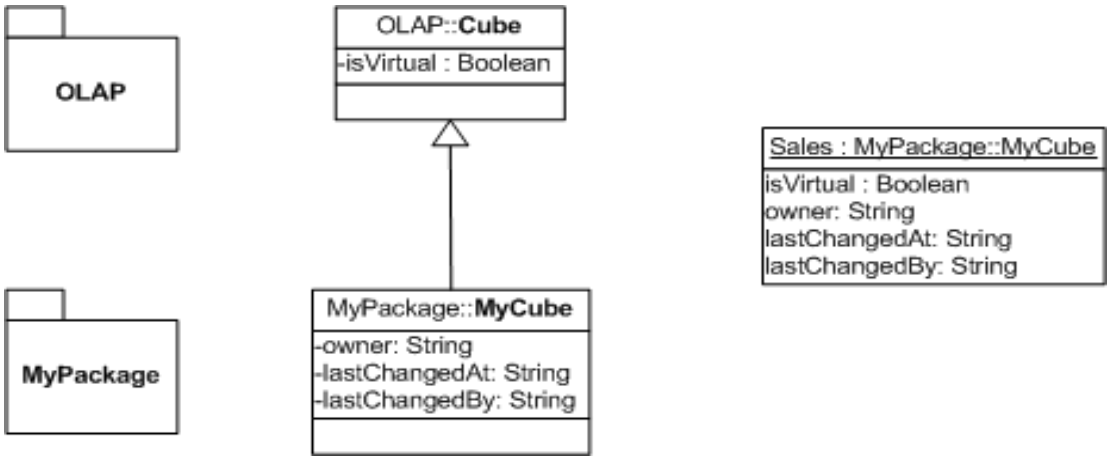
Extending the CWM: Stereotypes

- Stereotype: Taken from the UML specification (Package *Core*)
- Independent of release changes in the CWM specification
- No additional associations in the metamodel
- Stereotypes are used to group multiple Tagged Values



Extending the CWM: Extension classes

- Most flexible way to extend the CWM: User-defined Classes, Associations and Packages
- Degree of freedom for design very high
- Risk of departing from standard
- Inheritence mechanisms should be used
- Dependent from release changes of the CWM
- OMG's Extension Packages are based upon this concept (Essbase, Express)



Need to extend the CWM

- Highly focused on technical metadata
- Missing concepts for describing user rights
  - Heterogenous DW landscapes are frequently characterized by various concepts for user administration and the coordination of user rights and roles
  - Standardization of metadata regarding user rights and authorities enables:
    - Adjustment of user rights
    - Increase transparency and thereby avoiding security leaks

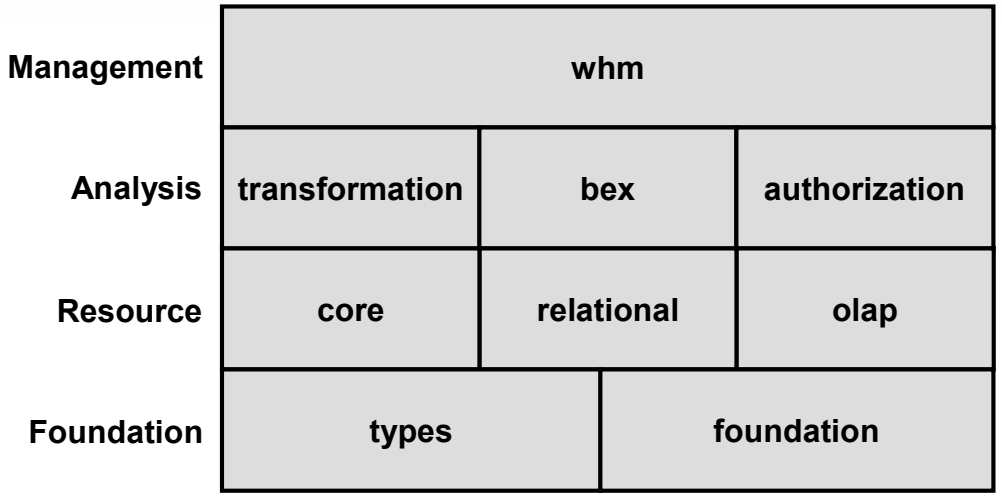
Need to extend the CWM

- Missing concepts for quality management
  - Conceptual, functional, technical
  - Acceptable intervals and alerting mechanisms
- Insufficient consideration of multiple mechanisms of report grouping and distribution via different channels
- The Organizational Data Warehouse process can be incorporated through extension of the CWM

Agenda: Model Management

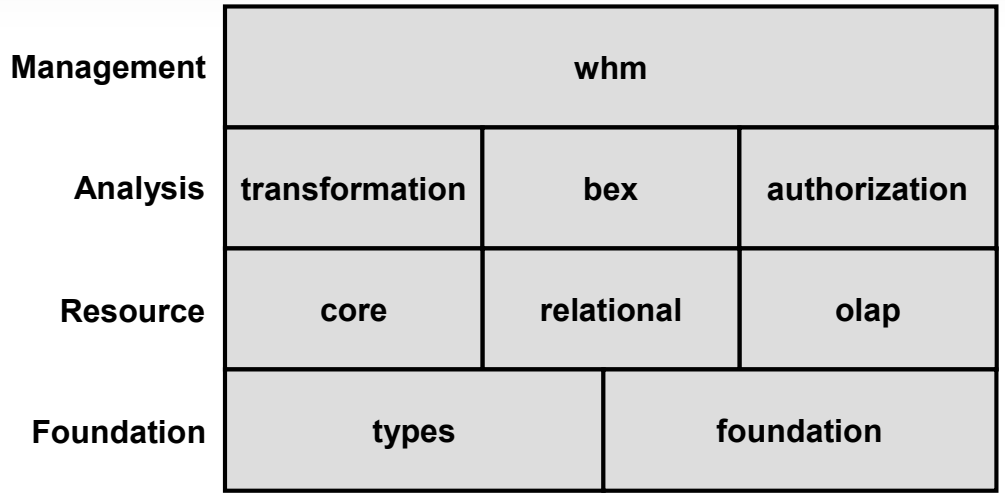
- Graphical model representation with Visio™
- Metadata in Data Warehousing
- Common Warehouse Metamodel (CWM)
- Implementation of the CWM in BW

BW CWM package: *types*



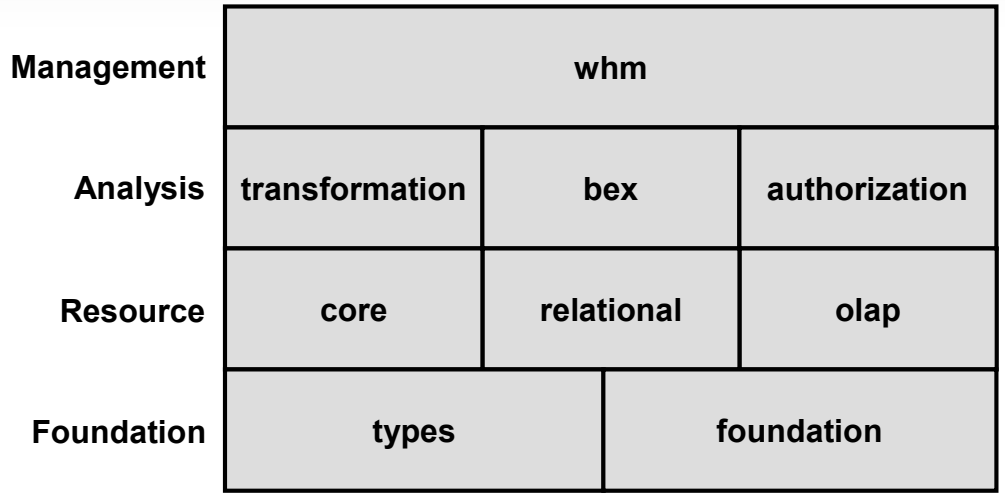
- BW specific data types:
  - Boolean
  - Integer
  - Name
  - String
  - Time

BW CWM package : *foundation*



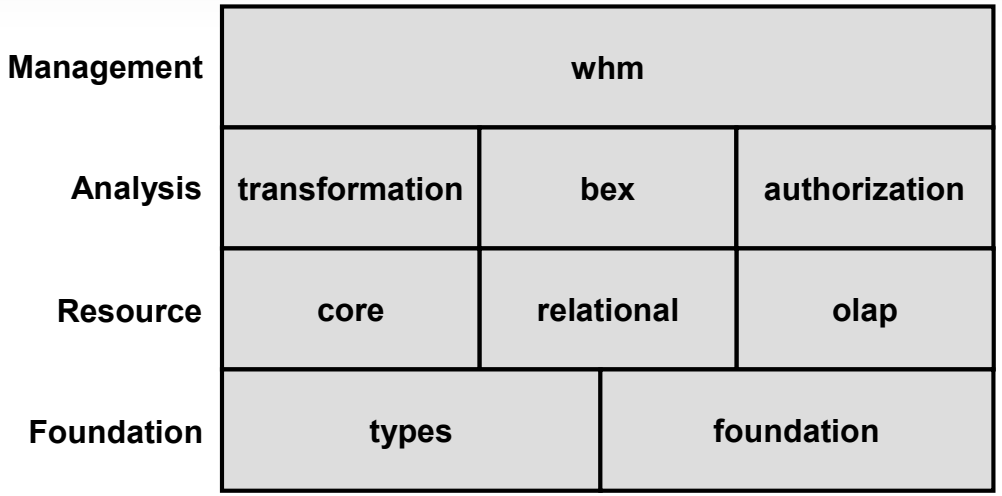
- Associations, e.g. for the connection of objects along the data flow in BW
- Attributes for object properties such as owner, status and caption, for example
- Business Content information

BW CWM package : *core*



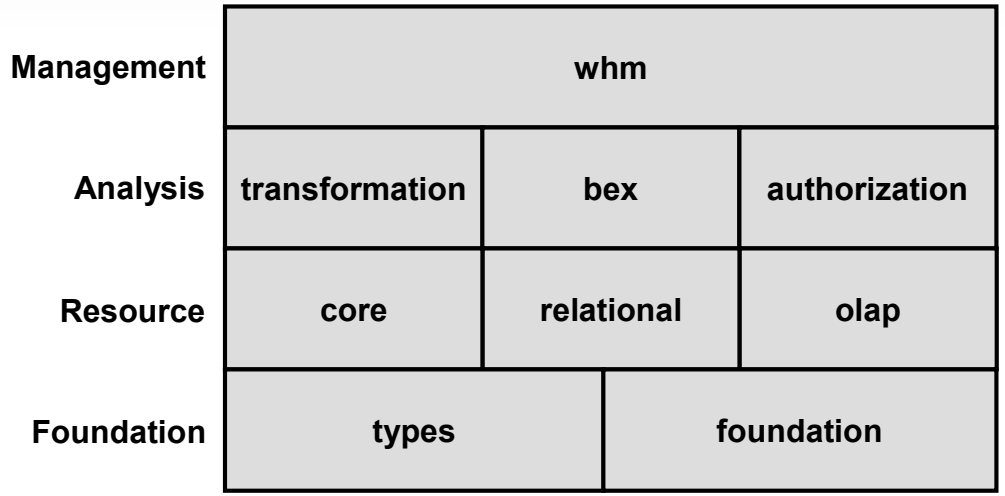
- Class definitions for
  - InfoObjects
  - Currency conversion
  - Persons
  - ABAP code
  - Various source systems

BW CWM package : *relational*



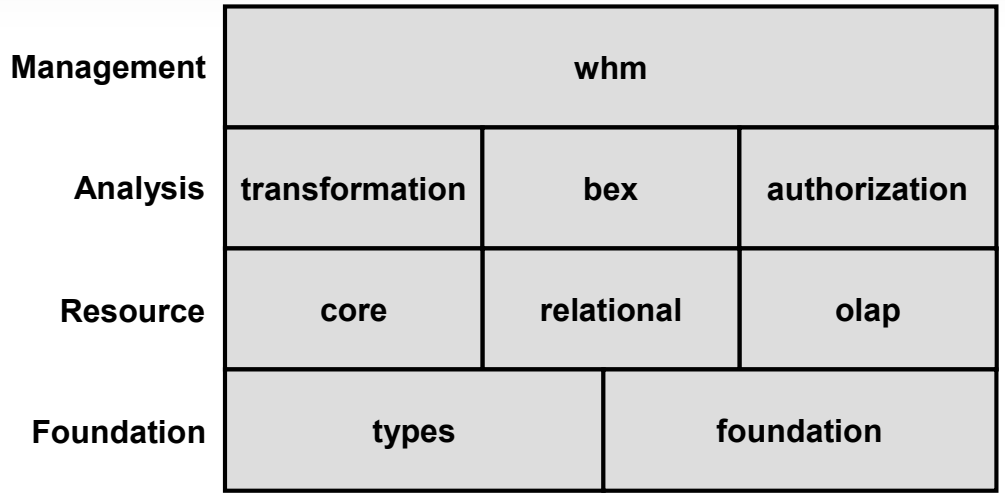
- ODSObjects
- InfoSets
- Indices

BW CWM package : *olap*

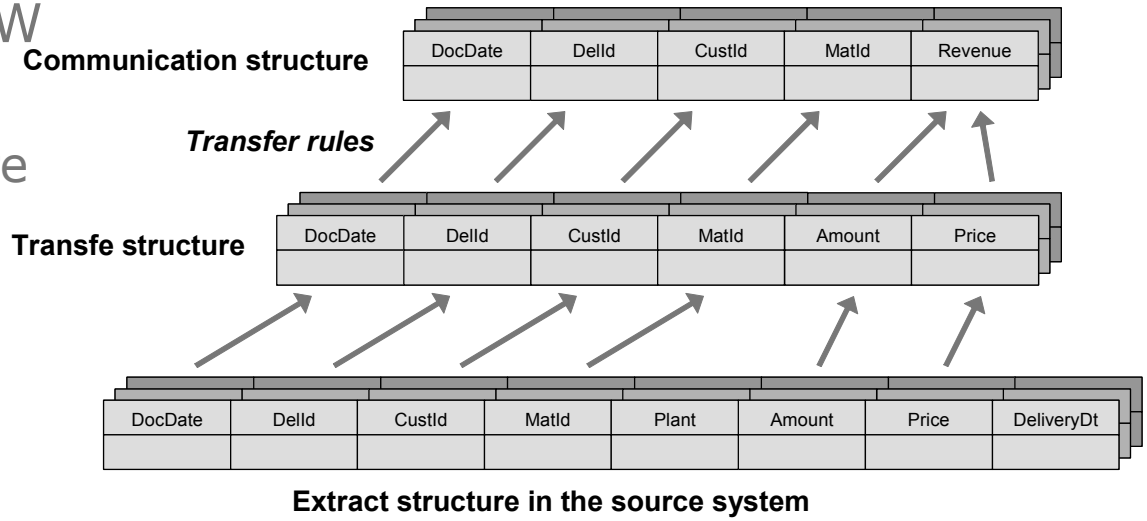


- Classes for InfoProvider and Aggregates
  - –Classes *InfoCube, InfoProvider, MultiProvider, VirtualCube, Aggregate*
- Reporting Agent
  - –Classes *ReportAgentPackage, ReportAgentSetting*

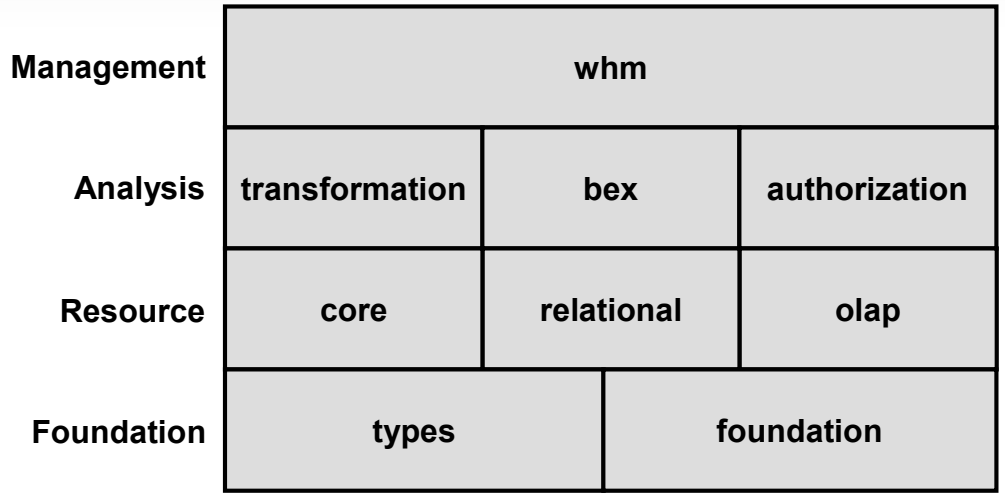
BW CWM package : *transformation*



- Transformations within BW
  - Transfer structure
  - Communikation structure
  - Info-Source
  - Update-Rule
  - etc.

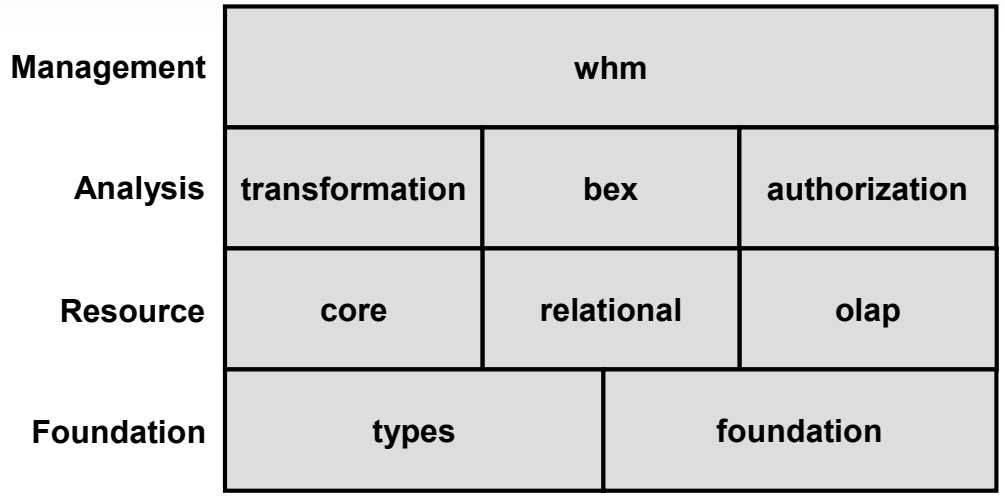


BW CWM package : *bex*



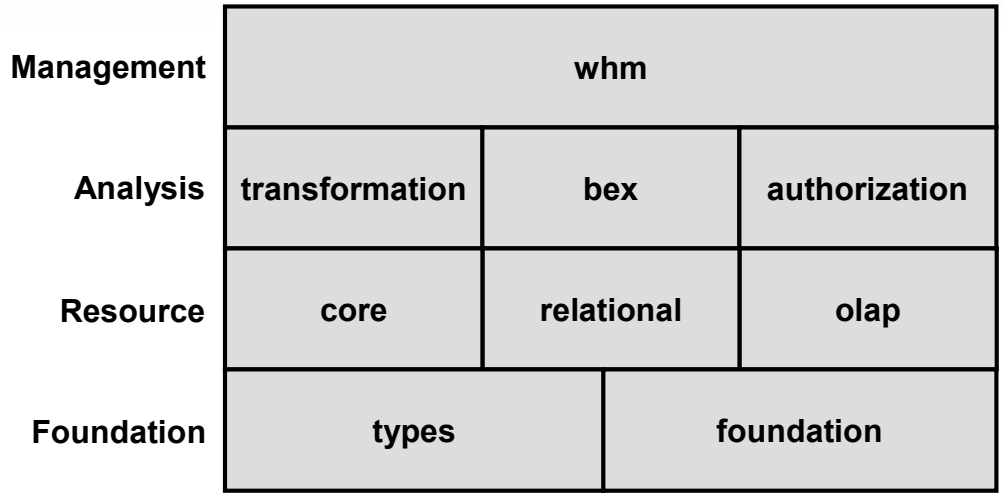
- Classes for data analysis with BEx Analyzer
  - Classes *ExcelWorkbook*, *InfoSetQuery*, *Query*, *QueryCalculatedKeyFigure*, *QueryElement, QueryFormula*, *QuerySelection*, *QueryStructure* and *QueryVariable*

BW CWM package : *authorization*



- Information about user rights according to the SAP role concept
  - Classes *Authorization*, *Role* and *UserGroup*

BW CWM package : *whm*



- Objects needed for operating the Business Information Warehouse
  - Classes *EventChain*, *InfoPackage* and *InfoPackageGroup*

- Classes used for organizing and structuring of objects by means of catalogs and areas
  - Classes *InfoObjectCatalog*, *InfoArea* and *Application*

Agenda: Model Management

- Graphical model representation with Visio™
- Metadata in Data Warehousing
- Common Warehouse Metamodel (CWM)
- Implementation of the CWM in BW

Summary: CWM and BW model management

- The Common Warehouse Metamodel (CWM): industry-wide standard for modelling and exchange of meta data
- The CWM offers a promisin approach of integration of components in a heterogenous data warehouse landscape
- Independant of architecture and platform, no specific tools considered, can be easily and flexible expanded
- Major weak points of the CWM: unsufficient support of
  - User rights
  - Meaning of Data Warehouse content,
  - Quality management
  - Organizational Data Warehouse Process
- SAP BW: Implementation of a CWM compliant extension:
  - Reasonable extension that can be useful as basis for further standardization especially in the area of user rights

Agenda: Enterprise Data Warehousing with BW

- Aspects of physical data model
- Staging architectures in BW
- Multi-Layer EDW architecture
- Information Lifecycle Management

Agenda: Enterprise Data Warehousing with BW

- Aspects of physical data model
- Staging architectures in BW
- Multi-Layer EDW architecture
- Information Lifecycle Management

Compression of Info-Cubes

F fact table without compression

Request	Product	Period	Revenue
1	P1	2003-04	40
1	P2	2003-04	20
1	P3	2003-04	50
1	P4	2003-04	40
1	P5	2003-04	30
2	P1	2003-04	30
2	P2	2003-04	40
2	P3	2003-04	60
2	P4	2003-04	40
2	P5	2003-04	50
3	P1	2003-04	40
3	P2	2003-04	60
3	P3	2003-04	20
3	P4	2003-04	60
3	P5	2003-04	80

compression of all Requests up to Request 2

E fact table

Product	Period	Revenue
P1	2003-04	70
P2	2003-04	60
P3	2003-04	110
P4	2003-04	80
P5	2003-04	80

F fact table

Request	Product	Period	Revenue
3	P1	2003-04	40
3	P2	2003-04	60
3	P3	2003-04	20
3	P4	2003-04	60
3	P5	2003-04	80

Partitioning on database level

fact table

Product	Period	Rev.
P1	2003-04	110
P2	2003-04	120
P3	2003-04	130
P4	2003-04	140
P5	2003-04	150
P1	2003-05	110
P2	2003-05	120
P3	2003-05	130
P5	2003-05	150
P6	2003-05	160

partitioning on database level by time characteristic

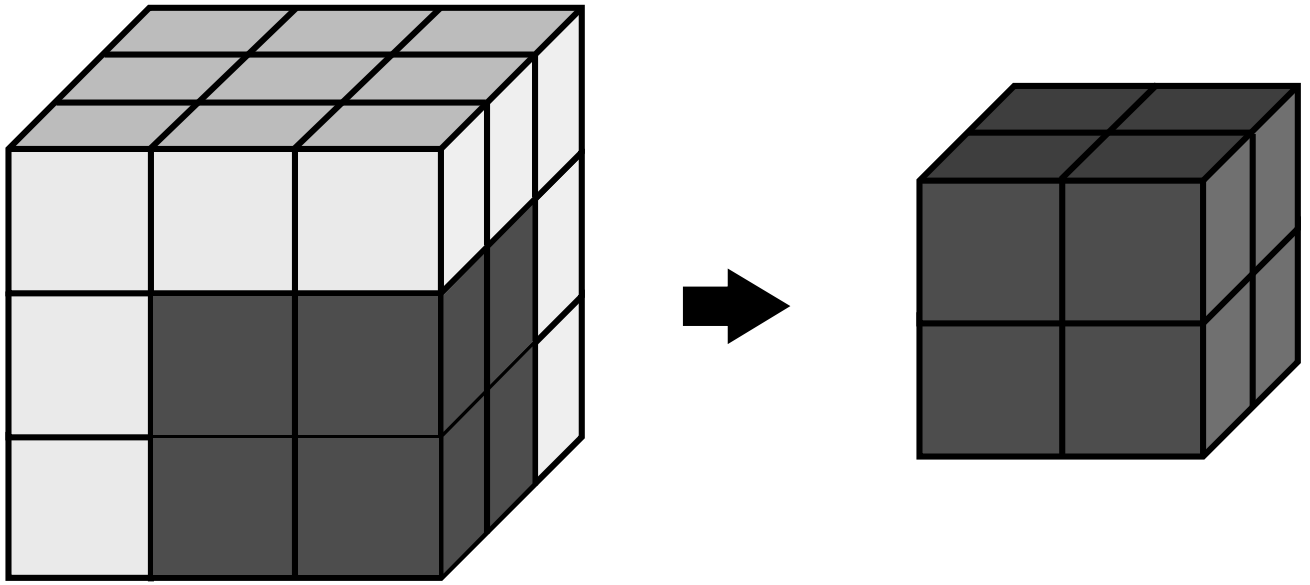
partitioned fact table for 2003-04

Product	Period	Rev.
P1	2003-04	110
P2	2003-04	120
P3	2003-04	130
P4	2003-04	140
P5	2003-04	150

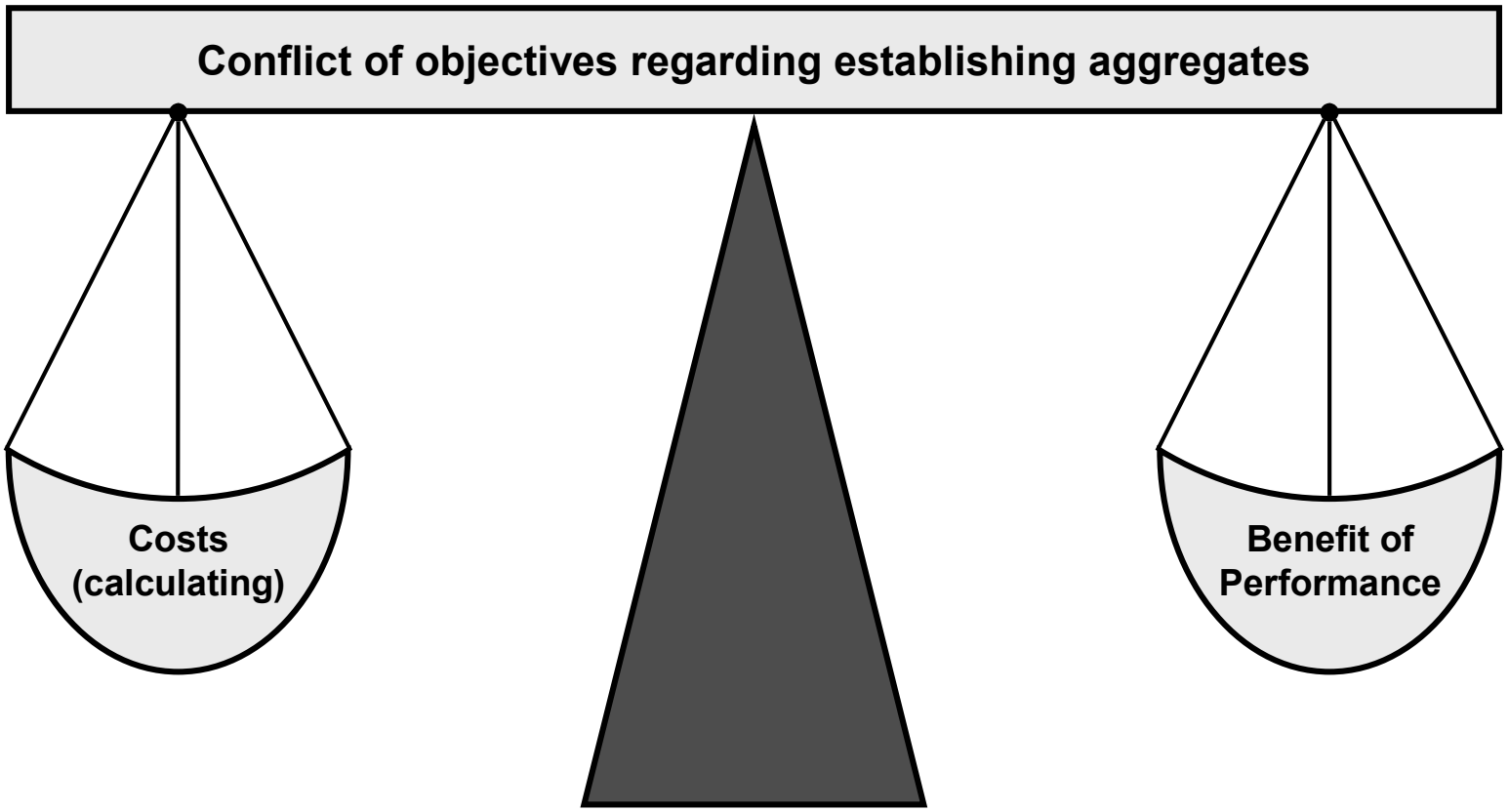
partitioned fact table for 2003-05

Product	Period	Rev.
P1	2003-05	110
P2	2003-05	120
P3	2003-05	130
P5	2003-05	150
P6	2003-05	160

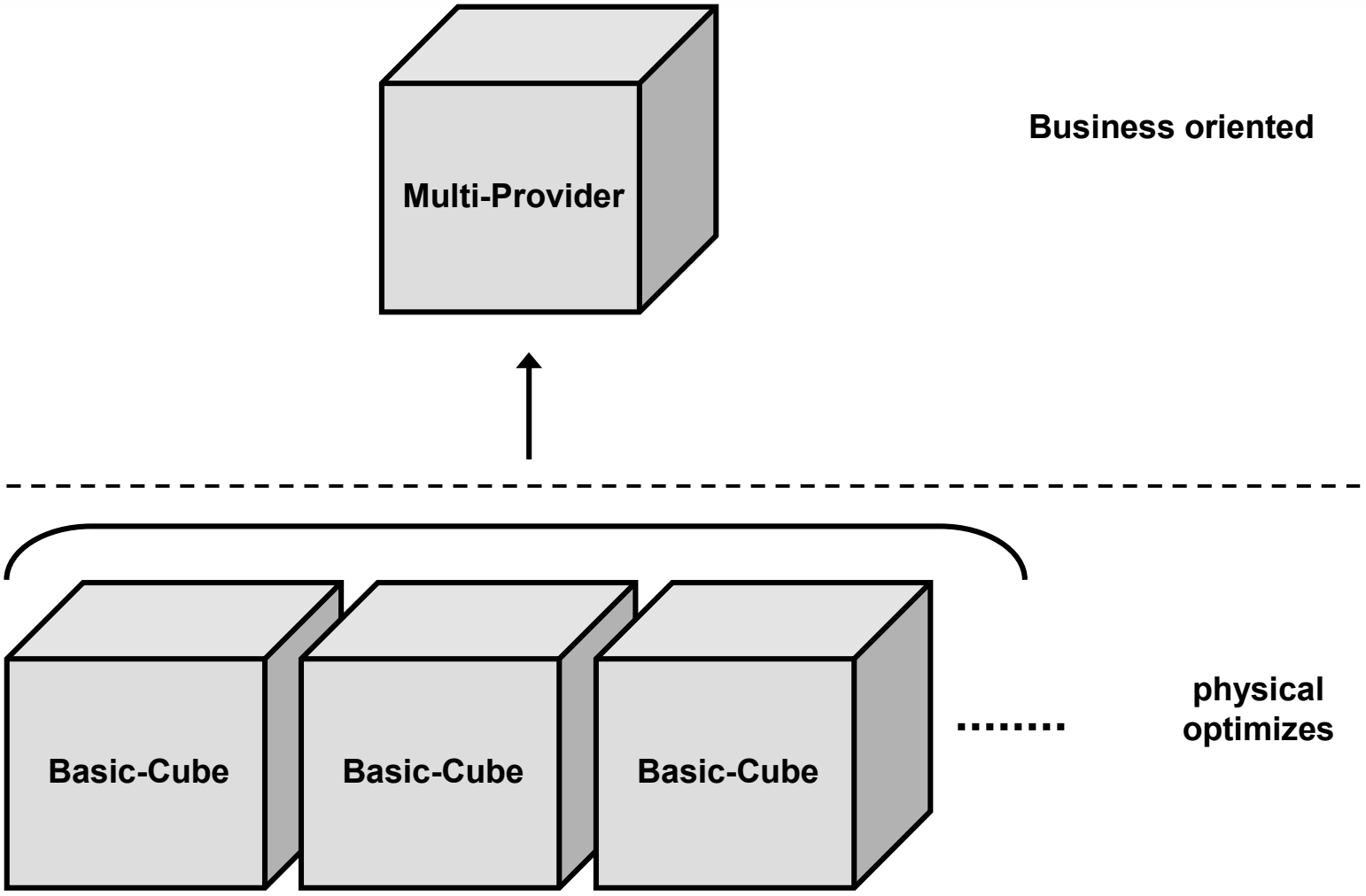
Aggregates



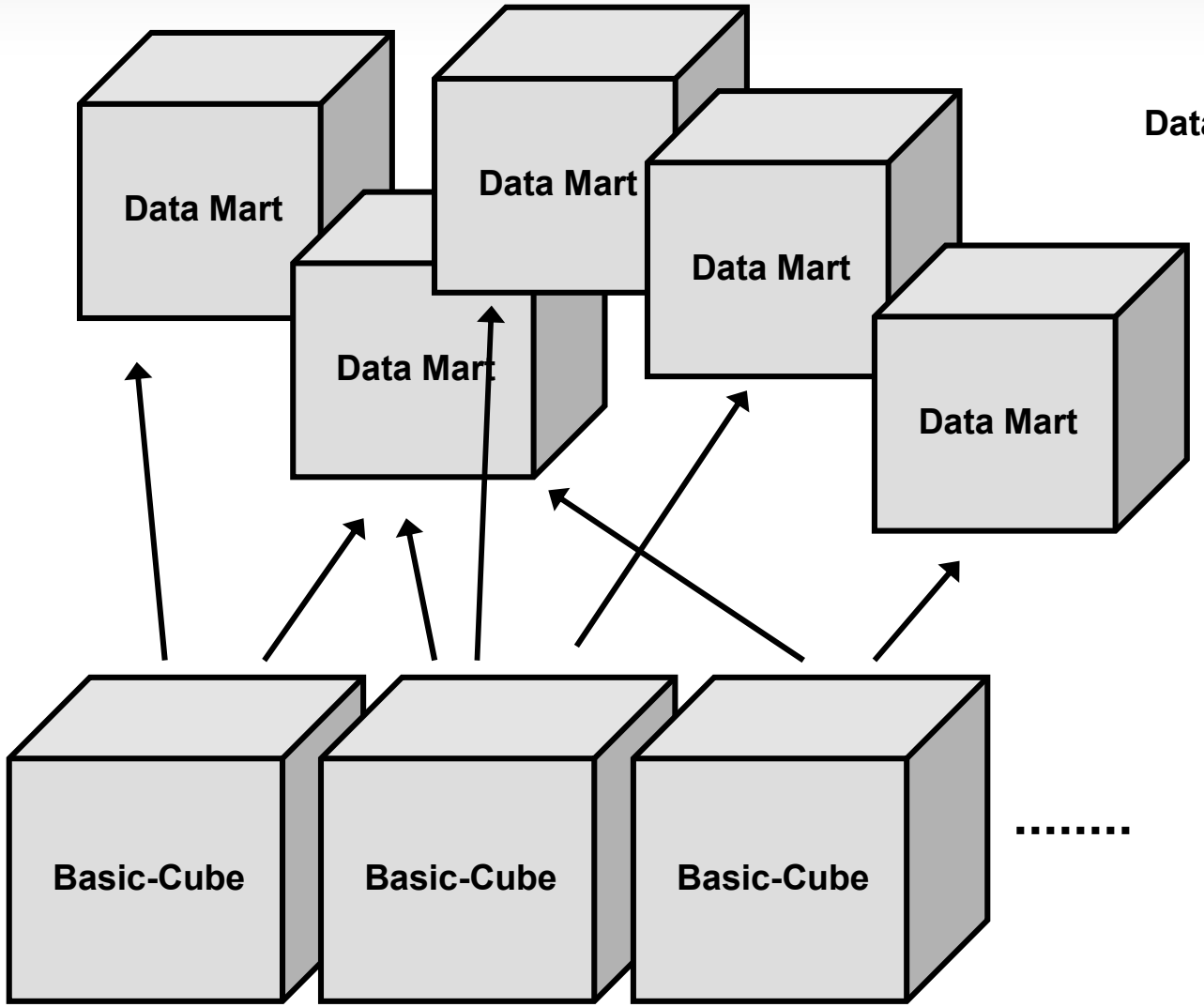
Aggregates: trade-off



2-Layer concept of cube modelling



Data Mart concept



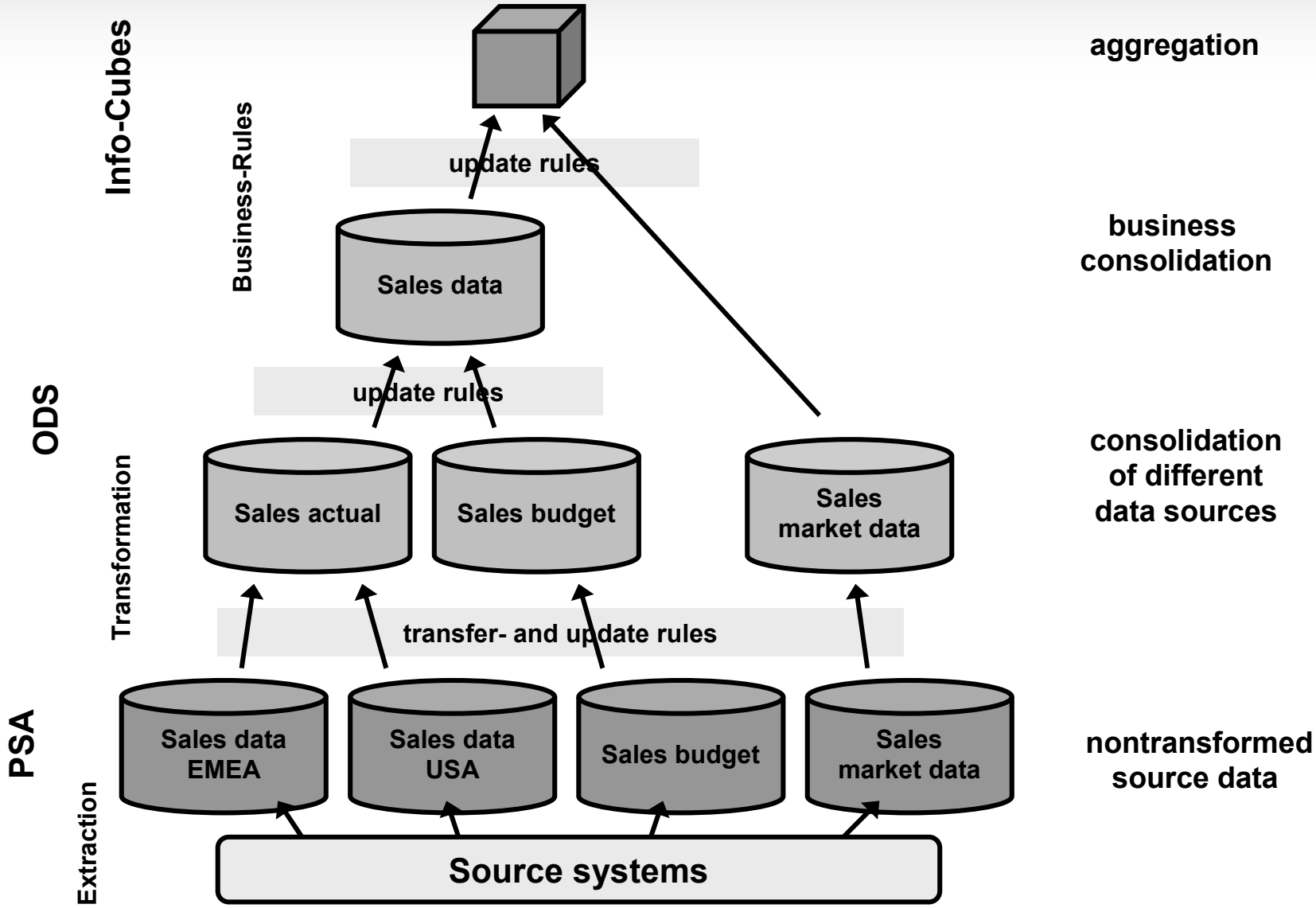
Derived views:
Data Marts based upon Basic-Cubes

Various Info-Cubes as data source

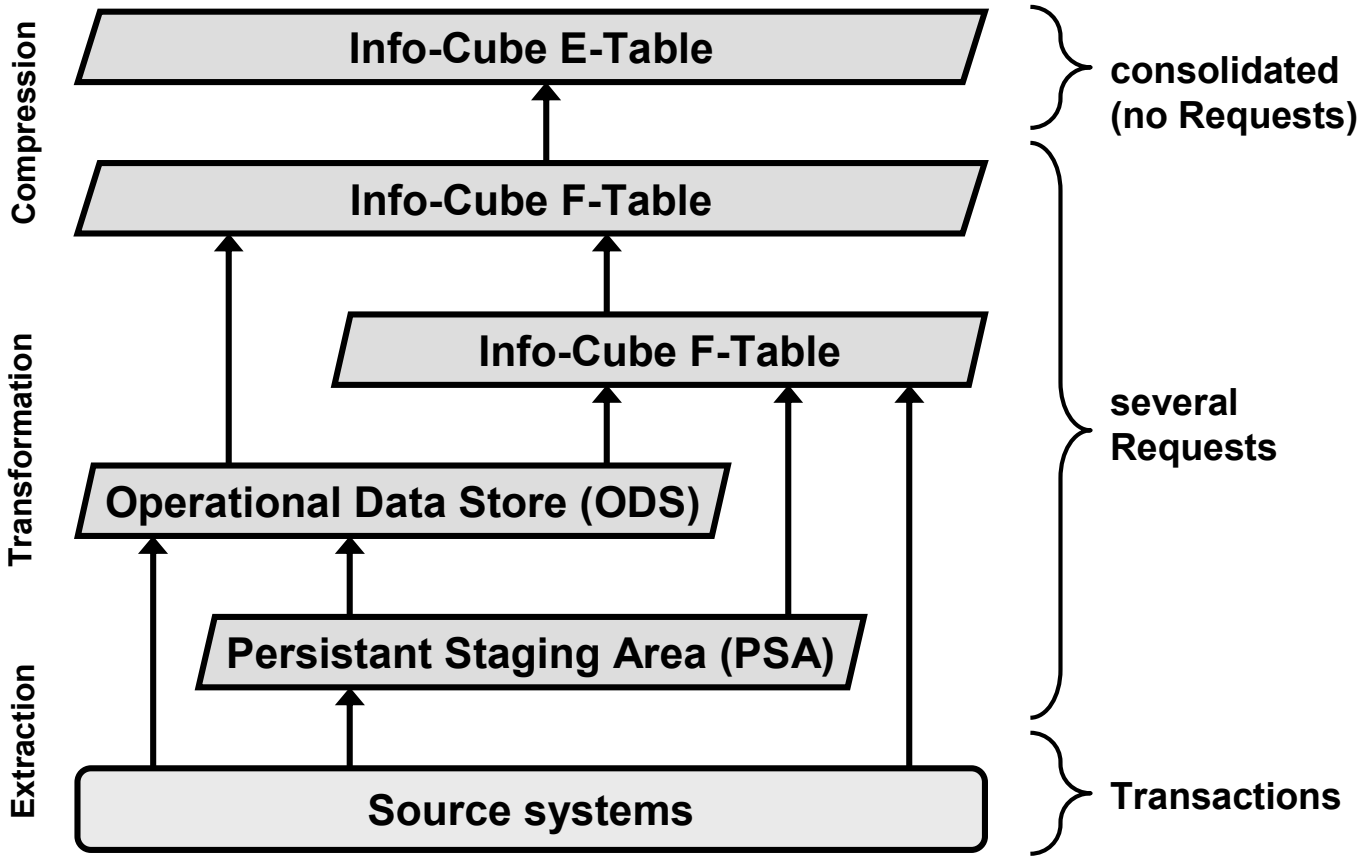
Agenda: Enterprise Data Warehousing with BW

- Aspects of physical data model
- Staging architectures in BW
- Multi-Layer EDW architecture
- Information Lifecycle Management

Data flow and integration architecture



Data flow when using compression of Basic-Cubes

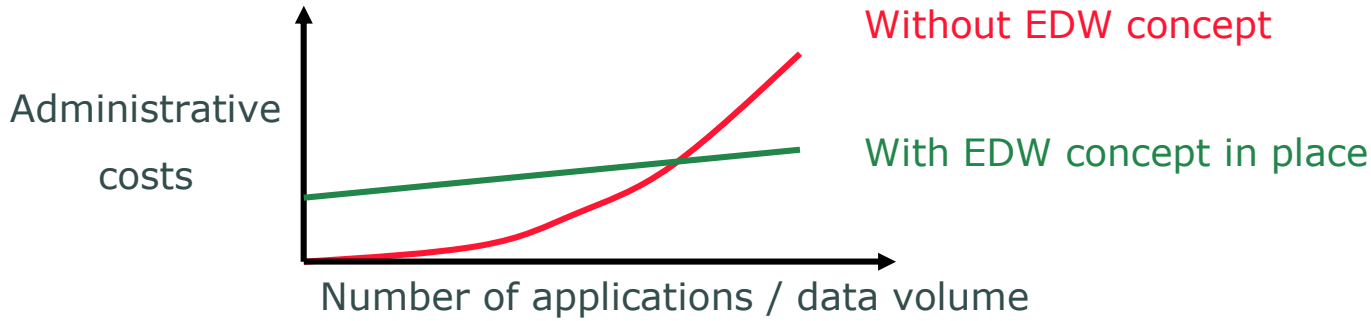


Agenda: Enterprise Data Warehousing with BW

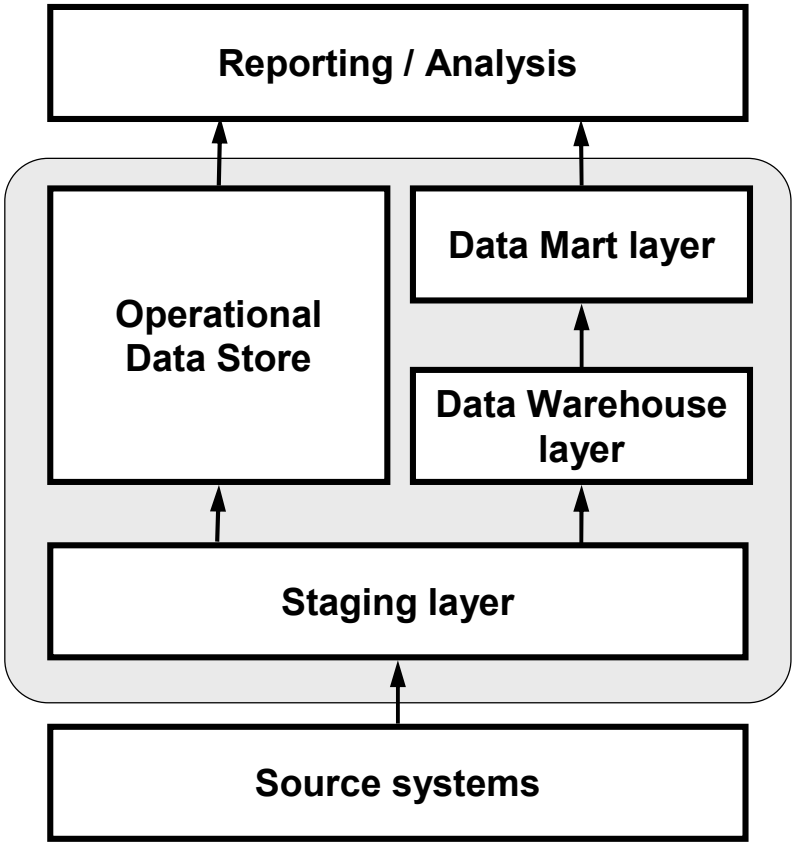
- Aspects of physical data model
- Staging architectures in BW
- Multi-Layer EDW architecture
- Information Lifecycle Management

Motivation of EDW concept

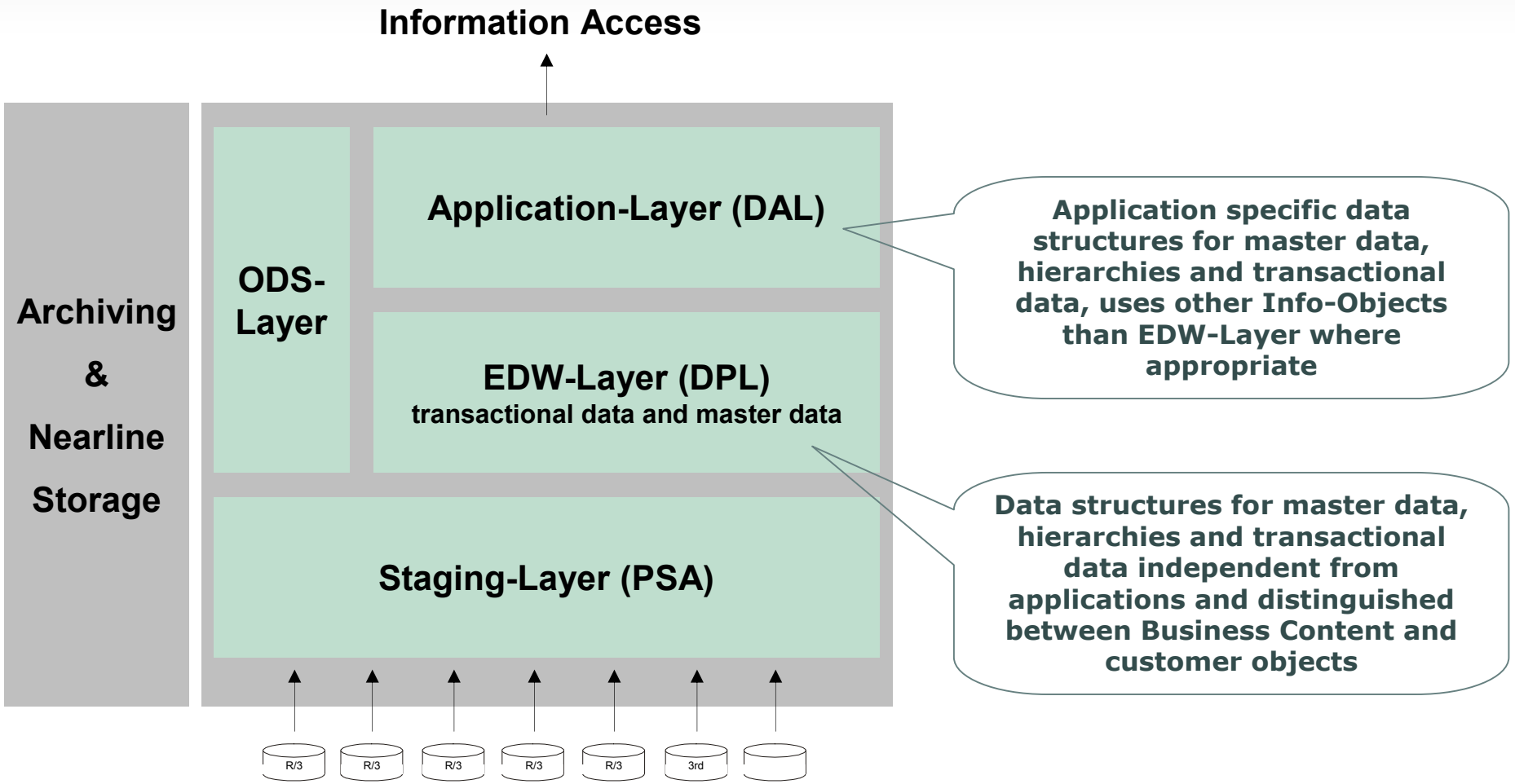
- Data growth
- Increasing number of applications
- Resulting in
  - Increasing administrative costs
  - Higher risk of breakdown of applications
  - Risk of total breakdown



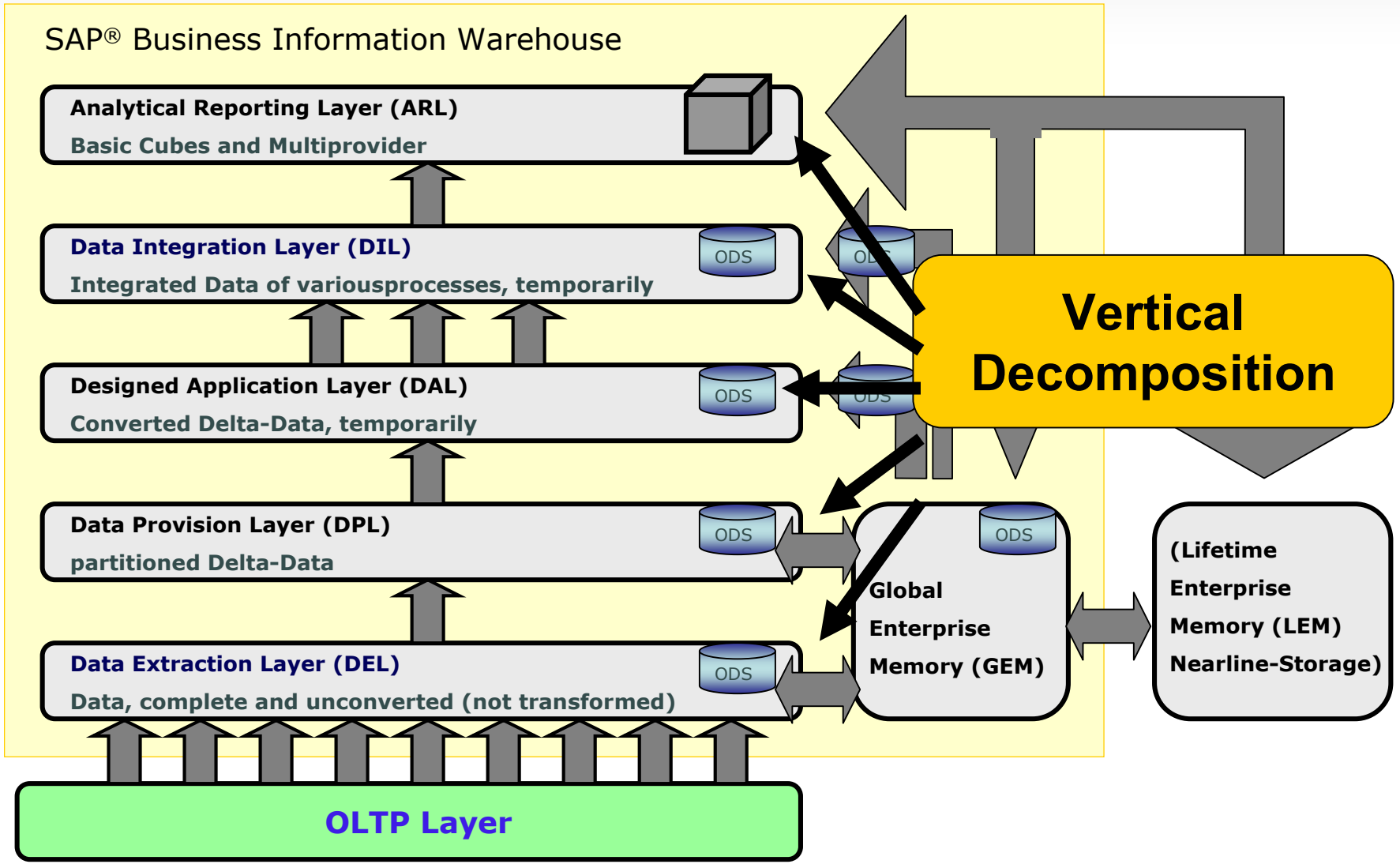
Conceptional Multi-Layer-Architecture



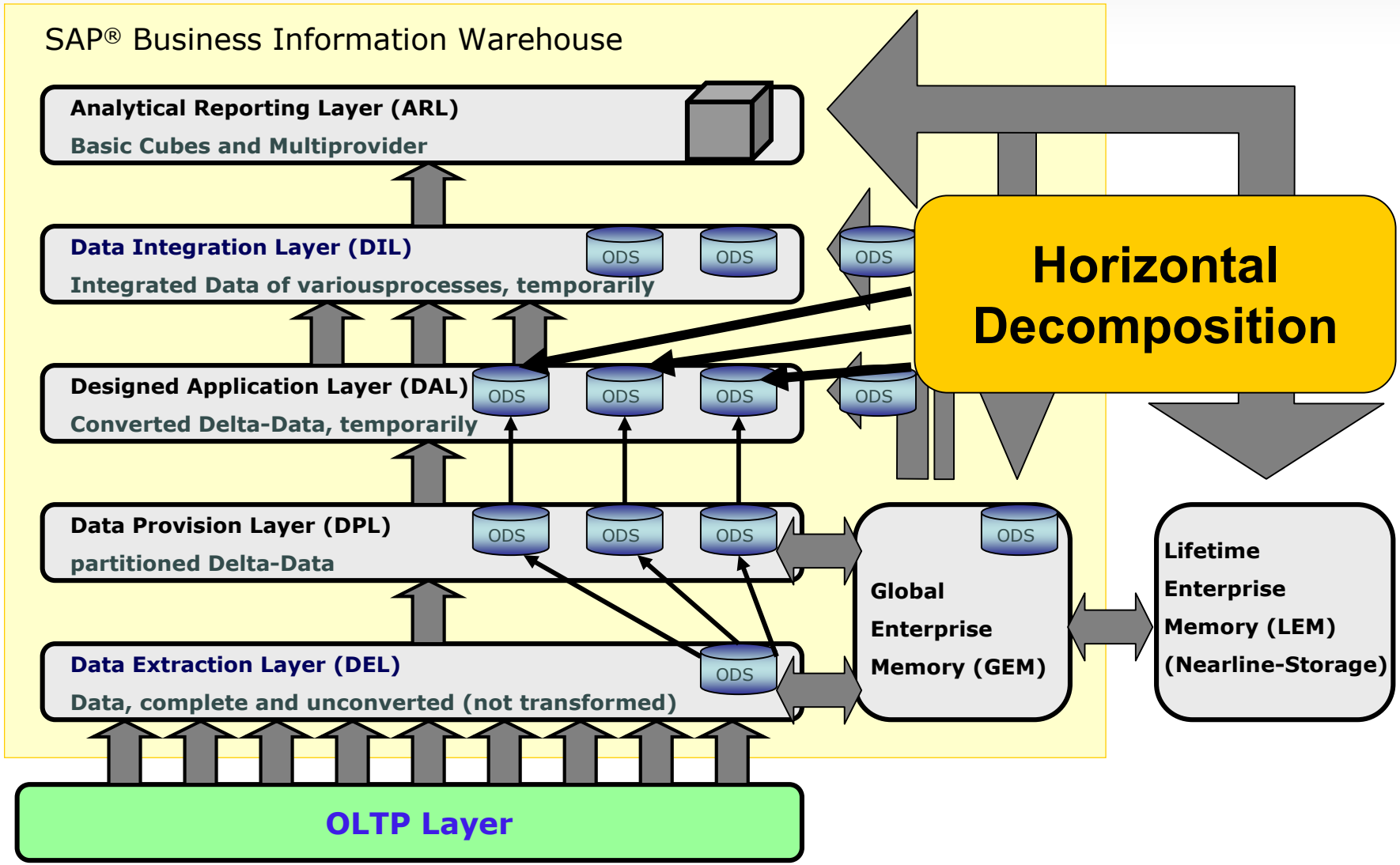
Layer Architecture



Vertical decomposition



Horizontal decomposition

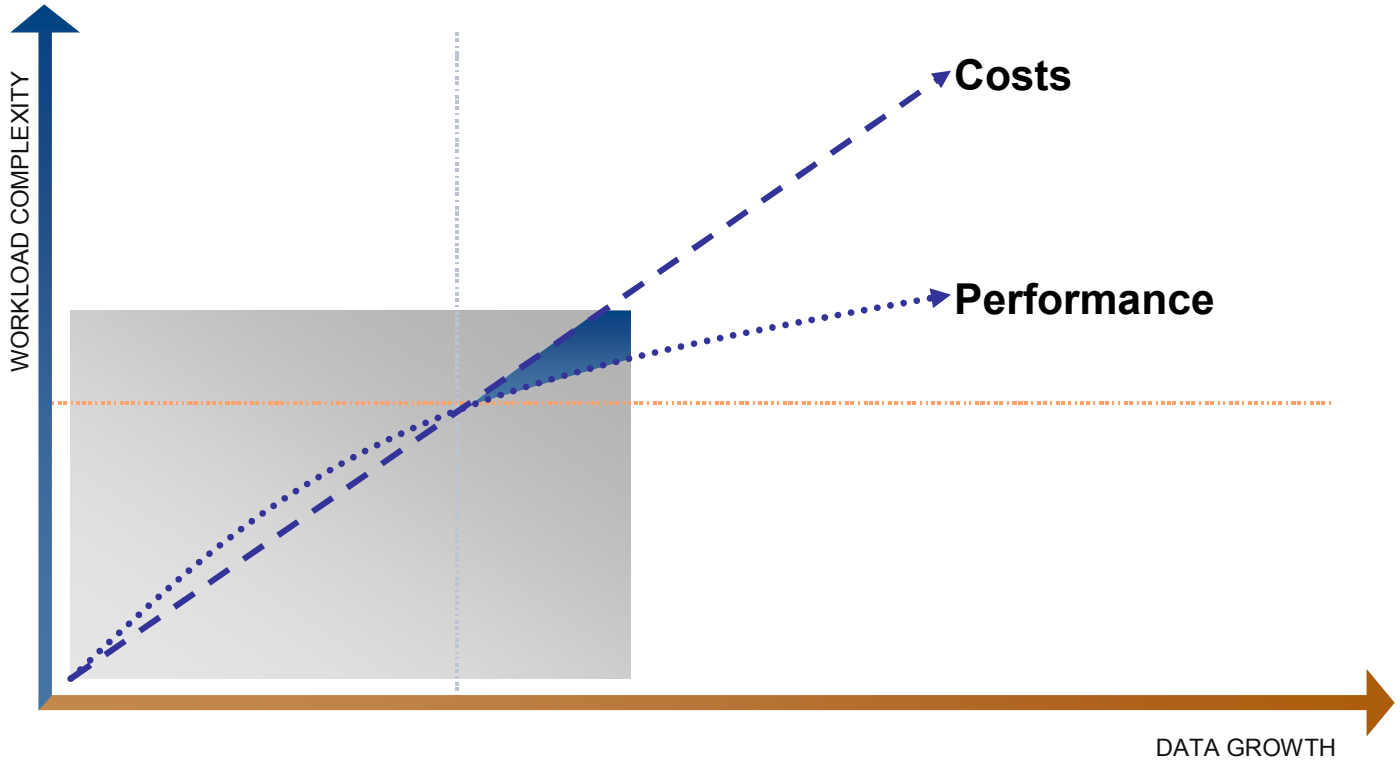


Agenda: Enterprise Data Warehousing with BW

- Aspects of physical data model
- Staging architectures in BW
- Multi-Layer EDW architecture
- Information Lifecycle Management

Existing Warehouse Under Stress

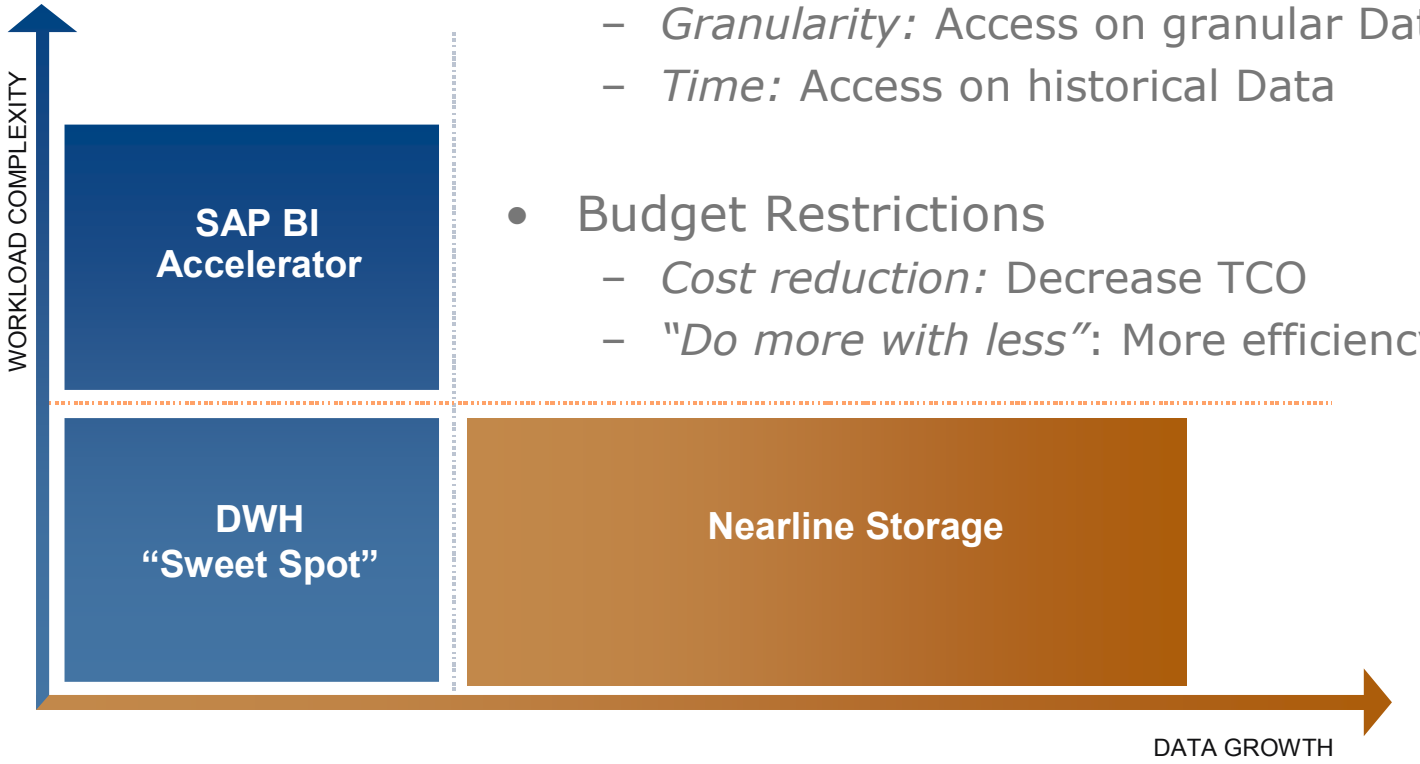
- The Warehouse can cope only to a point
  - Incremental growth already planned
  - More tuning and performance measures taken



SAP BI Accelerator and SAP BI Nearline Concepts

- Ability to compete in the Business
  - *Availability:* worldwide, 24x7, ad hoc
  - *Actuality:* Information JIT
  - *Performance:* fast Access
  - *Application Area:* enterprise-wide
  - *Granularity:* Access on granular Data
  - *Time:* Access on historical Data

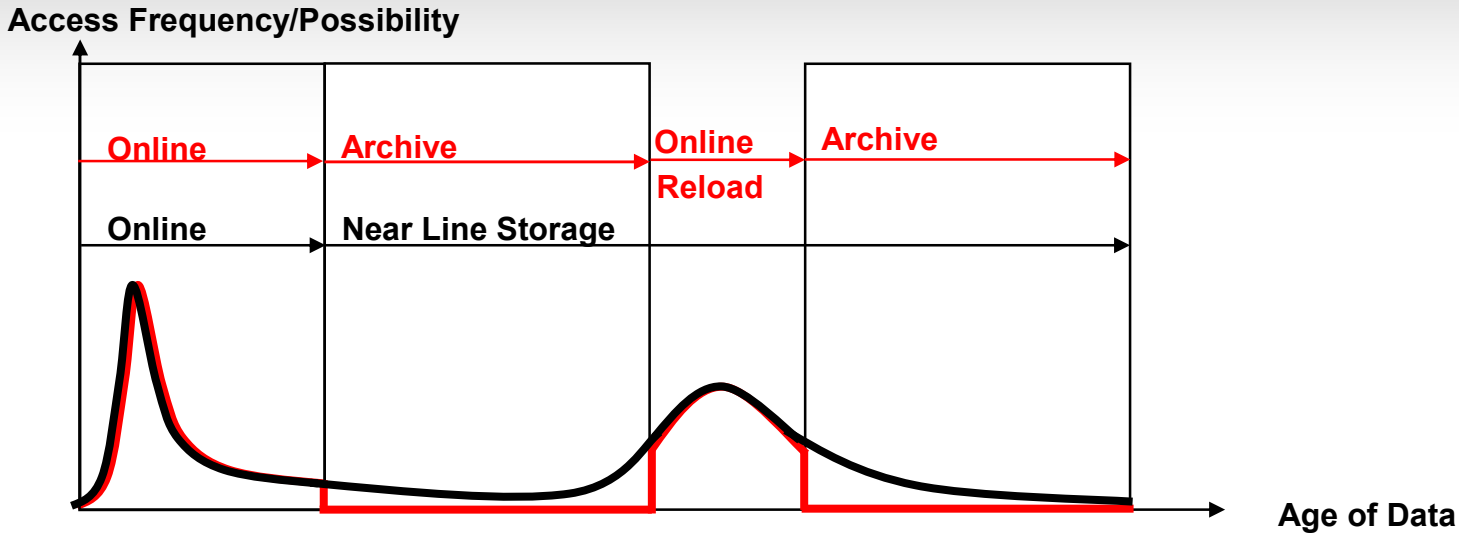
- Budget Restrictions
  - *Cost reduction:* Decrease TCO
  - *"Do more with less"*: More efficiency



Data Aging: Categorizing Information According to Importance

	Online Database	Near-Line Storage	Classic Archive
Frequently read / changed data	✓		
Rarely read data	✓	✓	
Very rarely read data	✓	✓	✓

Classic Archiv vs. Near-Line



Archiving (SAP BW 3.X)

- ADK-based (Archive Development Kit) archiving solution for InfoCubes and ODS objects
- Cost-reduction due to storing data on alternative storage media
- Archived data must be reloaded into the SAP NetWeaver BI database for analysis purposes

NLS (SAP NW 2004s BI)

- SAP NetWeaver BI analyses have direct access to NLS data
- Availability of historic data while reducing costs
- Reloading of data into the InfoCube or DataStore Object only necessary in exceptional cases

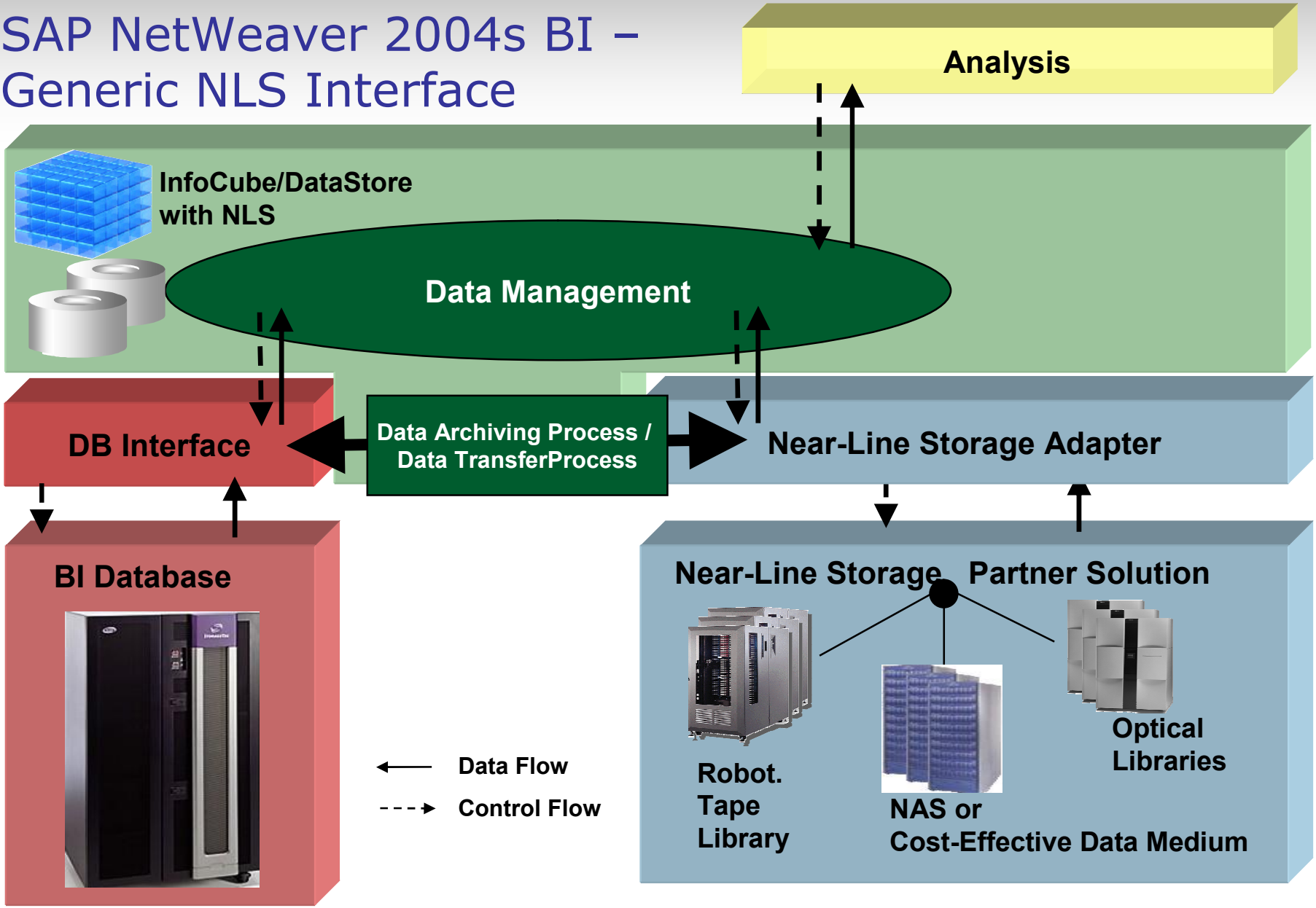
Benefits of a Fundamental ILM Strategy for BI

- Increase Volume
  - Manage and use even larger amounts of information more effectively
  - Information available for any time frame for ad-hoc analyses and rebuilds
- Reduce Resource Consumption
  - Reduction of hardware costs for hard drive hardware on the BW side
  - Main memory and CPU as well as costs for system administration
- Increase Availability
  - Quicker, simpler software- and release management in BW
  - Reduced backup- and recovery times
  - Intelligent data access
- Optimize Performance
  - Speed up loading processes in SAP NetWeaver BI
  - SAP NetWeaver BI query response times in the dialog

The Near-Line Storage Solution for SAP NetWeaver BI

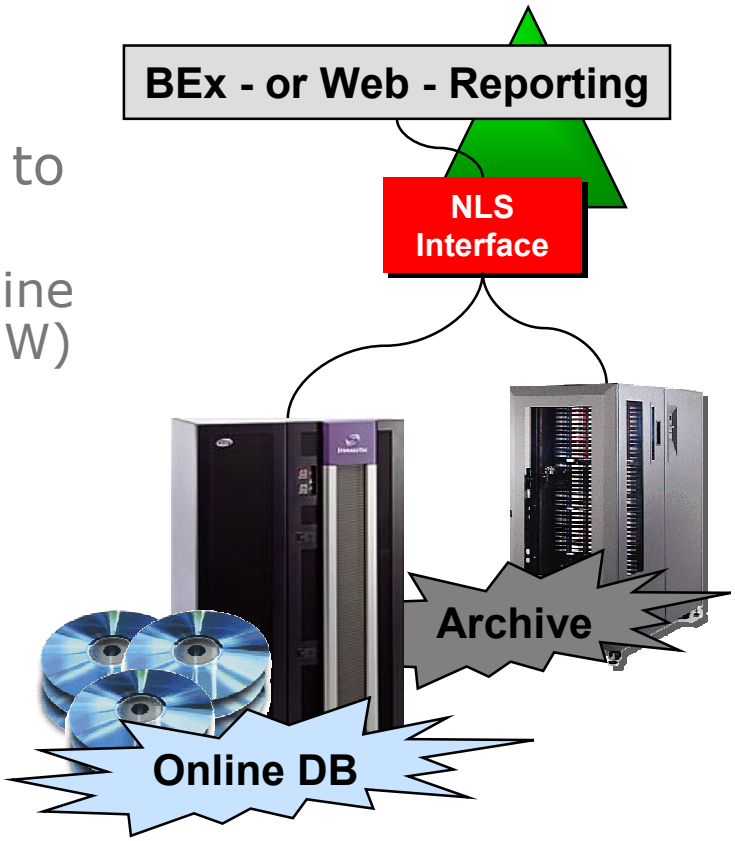
- Transparent access to „non-archived“ and „archived“ data for queries
- NLS support for InfoCubes and DataStore objects
- Open interface for certified partners,
  Present development partners:
  - PBS Software – CBW®
  - FileTek – StorHouse®
  - OuterBay - LiveArchive®
  - SAND Technology - Searchable Archive®

SAP NetWeaver 2004s BI – Generic NLS Interface

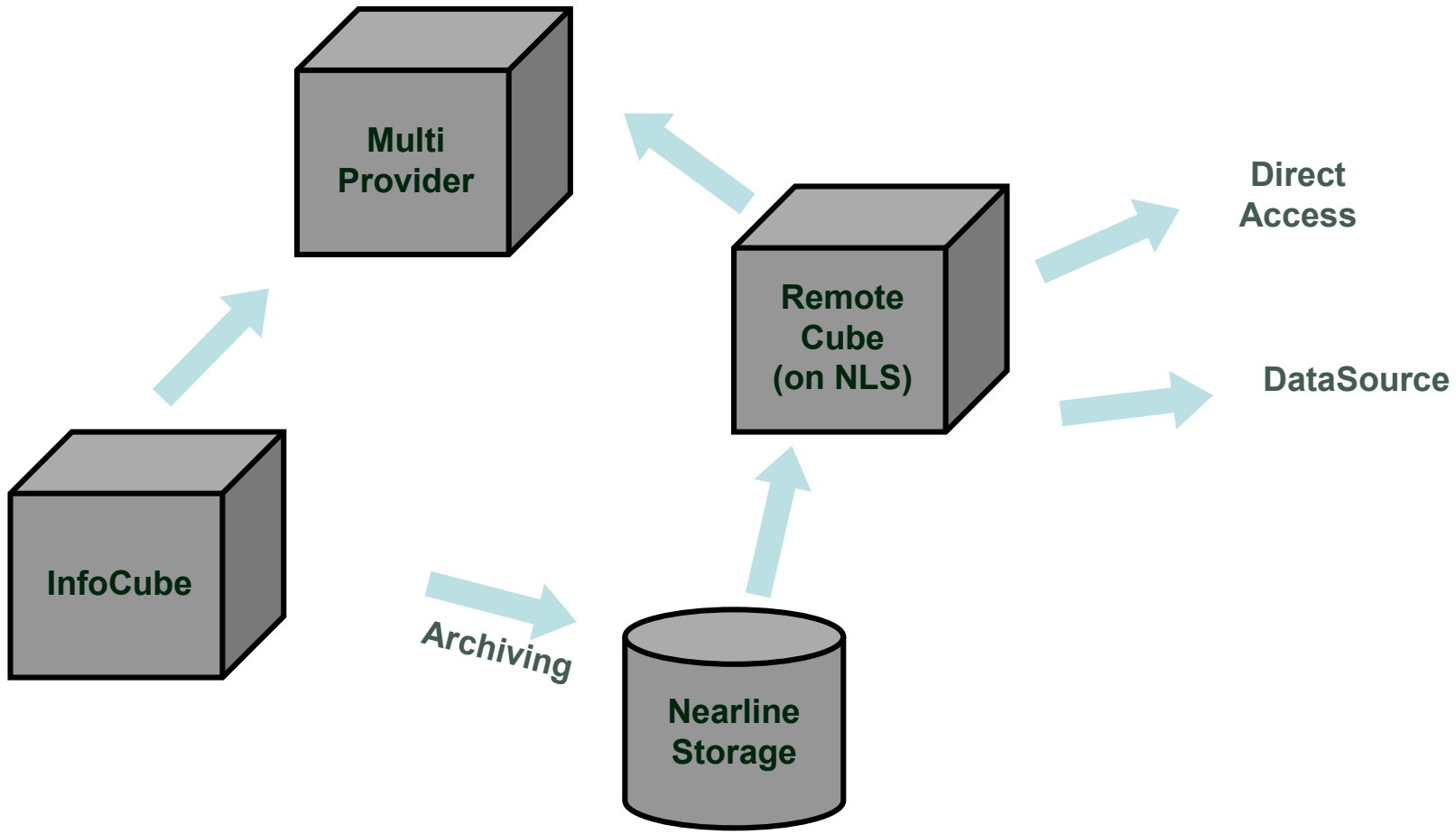


Consistency between nearline and online

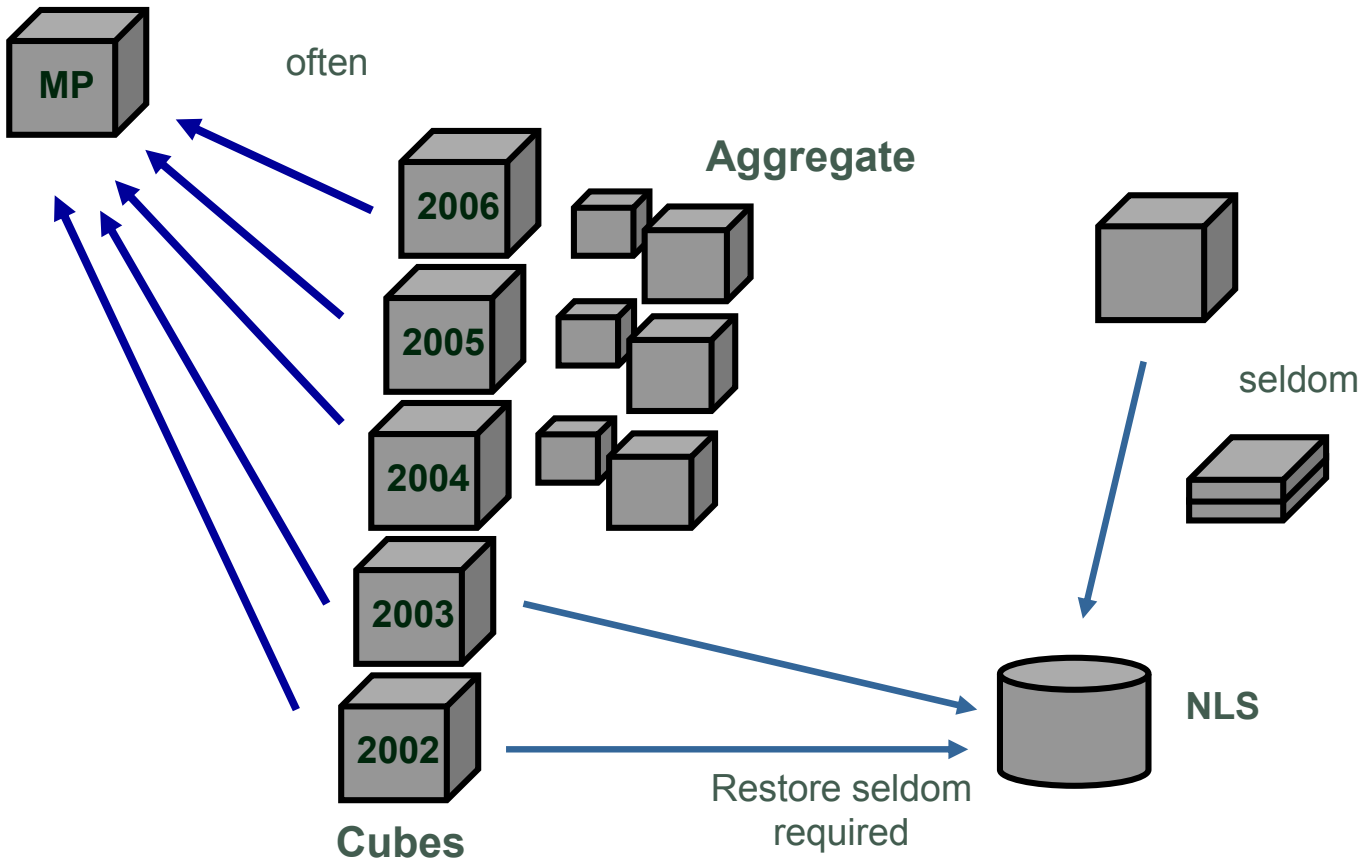
- Analysis and Reporting operate on a combination of online- and near-line datasets. The consistency of the data is an absolute prerequisite.
- Archiving processes into different near-line storage levels have to fulfill transactional requirements with regard to maintaing consistency
  - Archiving and deletion of data in the online database form a logical unit of work (LUW)
  - Rollback mechanisms available for individual archiving steps.
  - The „archive" gets the character of a database.
  - The archive data are usually ‚read only'



NLS with BW 3


Multi Provider
Remote Cube (on NLS)
Direct Access
DataSource
InfoCube
Archiving
Nearline Storage

Logical Partitioning of InfoCubes → Near-Line-Concept



Contact

Dr. Michael Hahne
Freiherr-vom-Stein-Str. 13a
55559 Bretzenheim
Germany

Email:
michael@hahneonline.de
URLs:
www.hahneonline.com
www.hahneonline.de
www.t-adapt.com

fon 0049 (671) 920 36 62
fax 0049 (671) 920 36 61

Dr. Michael Hahne
Productmanager Applications
SAND Technology

Email:
michael.hahne@sand.com
URL:
www.sand.com