Transforming Data
With Intelligence™

TDWI Dimensional Data Modeling Primer

From Requirements to Business Analysis

Previews of TDWI course books offer an opportunity to see the quality of our material and help you to select the courses that best fit your needs. The previews cannot be printed.

TDWI strives to provide course books that are content-rich and that serve as useful reference documents after a class has ended.

This preview shows selected pages that are representative of the entire course book; pages are not consecutive. The page numbers shown at the bottom of each page indicate their actual position in the course book. All table-of-contents pages are included to illustrate all of the topics covered by the course.

TABLE OF CONTENTS

COURSE OBJECTIVES

To learn:

- ✓ ***Concepts of dimensional data modeling***
- ✓ ***The relationship between business metrics and dimensional data***
- ✓ ***Similarities and differences between relational and dimensional data models***
- ✓ ***Requirements gathering techniques for business metrics and dimensional data***
- ✓ ***How to build a logical dimensional model***
- ✓ ***Understand different types of dimensions and where to use each***
- ✓ ***How to translate a logical dimensional model to a star schema design***
- ✓ ***Testing and validating the dimensional model***
- ✓ ***How dimensional data is used to deliver business analytics and OLAP capabilities***

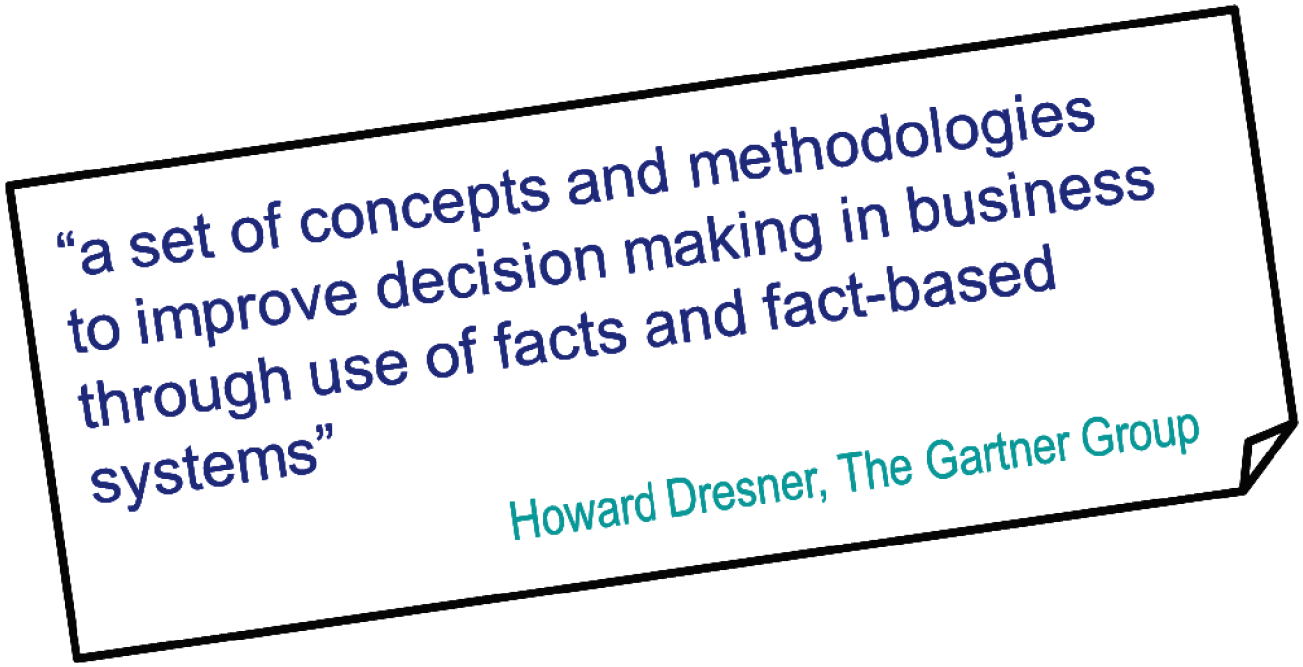
Module 1

Dimensional Modeling Concepts

Dimensional Modeling in Context

Business Intelligence Defined

Business Intelligence Defined



- **The goal is improved decision making.**
- **Concepts and methodologies – not just technology.**
- **Facts and fact-based systems are necessary to implement.**

Dimensional Modeling in Context

Business Intelligence Defined

BUSINESS INTELLIGENCE

Howard Dresner (Gartner Group) first defined business intelligence as shown on the facing page. As BI became a mainstream term, additional definitions have emerged. Among them is the following definition:

"BI is neither a product nor a system. It is an architecture and a collection of integrated operational as well as decision-support applications and databases that provide the business community easy access to business data." (Larissa T. Moss and Shaku Arte, *Business Intelligence Roadmap,* Pearson Education, 2003)

David Loshin defines business intelligence as: "The processes, technologies, and tools needed to turn data into information, information into knowledge, and knowledge into plans that drive profitable business actions. Business intelligence encompasses data warehousing, business analytic tools, and content knowledge management." (*Business Intelligence: The Savvy Manager's Guide*, Addison Wesley, 2003)

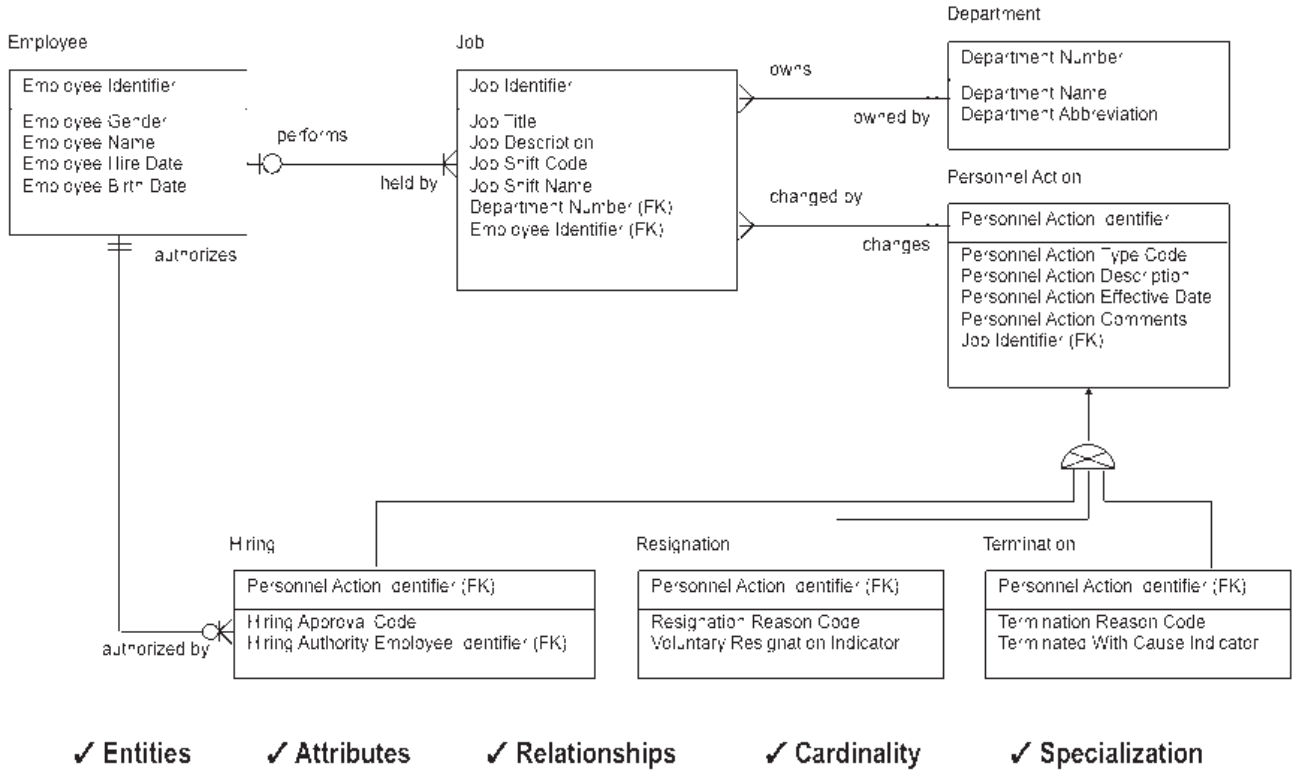
Business intelligence provides the ability to transform data into usable, actionable information for business purposes. BI requires:

- Collections of quality data and metadata important to the business
- The application of analytic tools, techniques, and processes
- The knowledge and skills to use business analysis to identify/create business information
- The organizational skills and motivation to develop a BI program and apply the results back into the business

The foundation that enables BI is the enterprise architecture—business, data, and technical. A well-implemented data warehousing program provides much of that foundation.

Comparing E-R and Dimensional Models

A Quick Review of Entity-Relationship Modeling



✓ Entities ✓ Attributes ✓ Relationships ✓ Cardinality ✓ Specialization

Comparing E-R and Dimensional Models

A Quick Review of Entity-Relationship Modeling

ENTITY-RELATIONSHIP MODELS

This diagram illustrates a simple entity-relationship (ER) model. The primary components of an ER model are:

- **Entity**—An entity is a subject about which the business has the need, will, and means to collect data—a person, place, thing, concept, or event that is of business interest. Entities are represented as labeled boxes in the ER model. Examples in the diagram include EMPLOYEE, JOB, and DEPARTMENT.

- **Attribute**—An attribute is a property or characteristic of an entity that can be collected as data. Attributes are listed inside the box of the entities that they describe. *Job_shift_code* and *job_title*, for example, are attributes of the entity JOB.

- **Relationship**—A relationship is an important association between pairs of entities that is of business interest and may be collected as data. Examples of relationships in this diagram include EMPLOYEE PERFORMS JOB and DEPARTMENT OWNS JOB.

Other important ER concepts include:

- **Cardinality,** which describes the number of occurrences of each entity type that may participate in an occurrence of a relationship. Cardinality options include *zero or one*, *exactly one*, *one or more*, and *zero or more*. In this diagram some of the relationship cardinalities are: a PERSONNEL ACTION changes *exactly one* JOB; a JOB is changed by *one or more* PERSONNEL ACTIONs; a JOB is held by *zero or one* EMPLOYEEs. Sometimes cardinality as defined above is divided with "cardinality" indicating the maximum number and "optionality" indicating the minimum number.

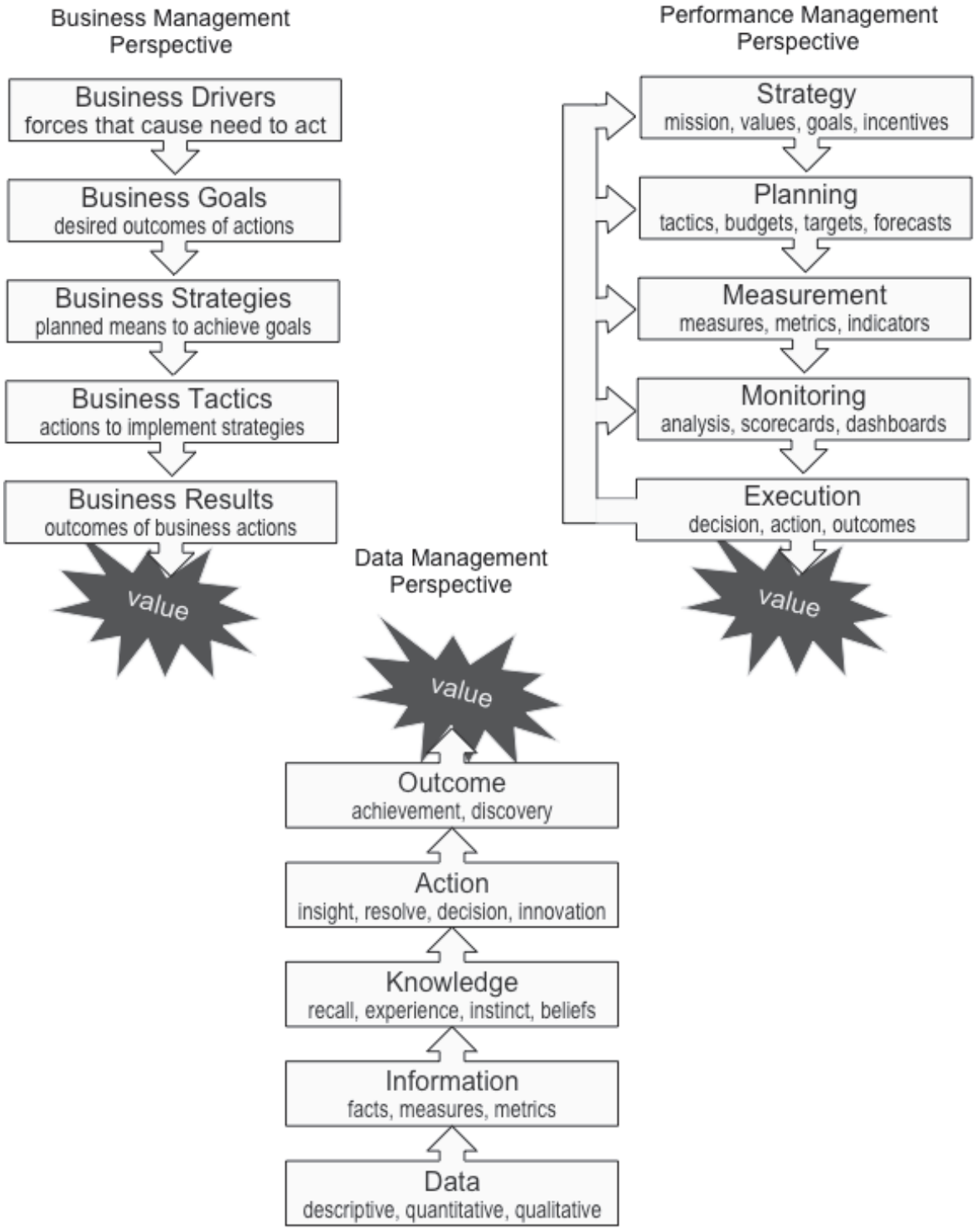
- **Specialization** (also called subtyping) creates a hierarchy or parent/child relationship between an ENTITY super-type and its sub-types. Specialization makes sense when the entity sub-types have unique attributes or participate in relationships not common to all sub-types. In this diagram, PERSONNEL ACTION is an entity super-type with three sub-types.

Module 2

Requirements Gathering for Dimensional Modeling

Business Context for Data Modeling

Business Value



Business Context for Data Modeling

Business Value

IMPORTANCE OF BUSINESS CONTEXT

Business context is essential to delivering business value. The context determines the nature of BI program—the people who will use the information, the kinds of information they need, the business processes to be affected, and the information services to be provided. Business context provides the means to align business intelligence results with business goals.

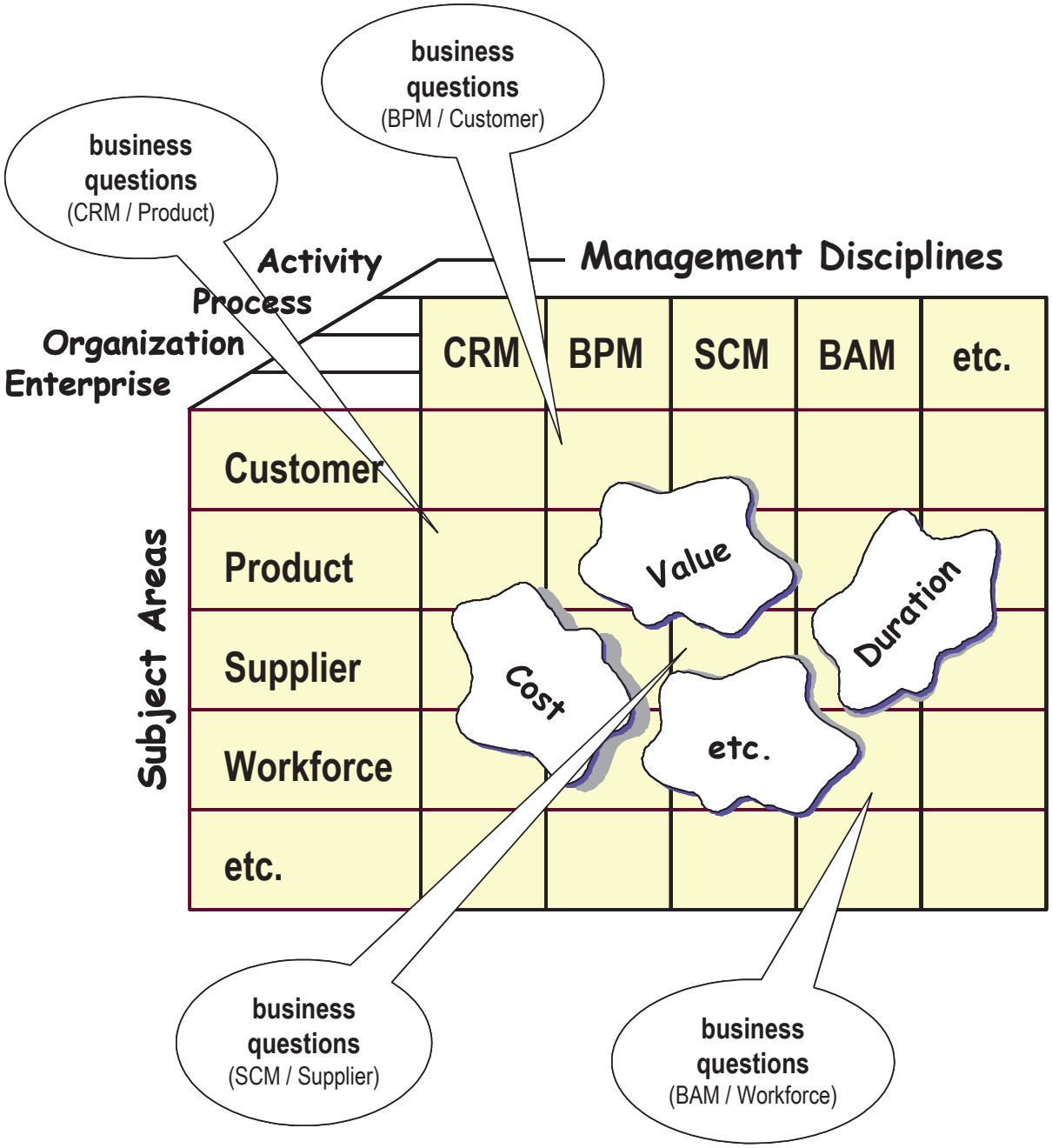
BUSINESS VALUE PERSPECTIVES

As previously described, dimensional modeling is really a business modeling process—modeling of business processes and activities. The "big picture" view of modeling context begins by understanding how we create business value. The facing page illustrates three perspectives of value creation:

- Business management views value as positive results of business activities that align with the tactics, strategies, goals, and drivers of the business.
- Data management views value as business outcomes that are enabled through useful data that enhances knowledge to inform business decisions and actions.
- Performance management views value as successful execution of continuously evolving strategy and planning where data and information are critical elements of a feedback system for continuous growth and improvement.

Business Questions as Requirements Models

A Framework for Business Questions



Business Questions as Requirements Models

A Framework for Business Questions

LIST OF BUSINESS QUESTIONS

Requirements modeling begins with business questions—the foundation for all subsequent dimensional modeling activities. The first context-level model is a robust and representative list of business questions that are within the scope of the project. It is not possible to develop a complete and exhaustive list of all questions that might be asked. The objective is to create a list of questions that is fertile enough to drive development of a highly adaptable data structure capable of answering questions that are not yet known. A major strength of the dimensional model is its ability to provide such a data structure.

A DIMENSIONAL VIEW OF METRICS

The diagram on the facing page illustrates a three-dimensional view that is useful in managing a comprehensive set of business metrics. The dimensions are:

- Levels of metrics—enterprise, organization, process, and activity—that were previously discussed.
- Subject areas of your business, typically found in a subject area model and the basis for the subject-oriented data warehouse. Subjects vary depending on the industry and the things that make your business unique. They include such things as customer, supplier, product, service, finance, etc.
- Performance management methods used by your business. These include disciplines such as business performance management (BPM), customer relationship management (CRM), supply chain management (SCM), etc.

INSIDE THE FRAMEWORK

Each intersection of the three dimensions is an opportunity to find business questions and to manage business metrics. At the intersection of *enterprise, product,* and *CRM,* you might ask:

- What needs to be known about products at the executive level to make CRM effective?
- What are the KPIs for products that are important to CRM?
- How are product measures for CRM related to those for BPM?
- How are product and customer measures for CRM related?

METRIC CLASS WORDS

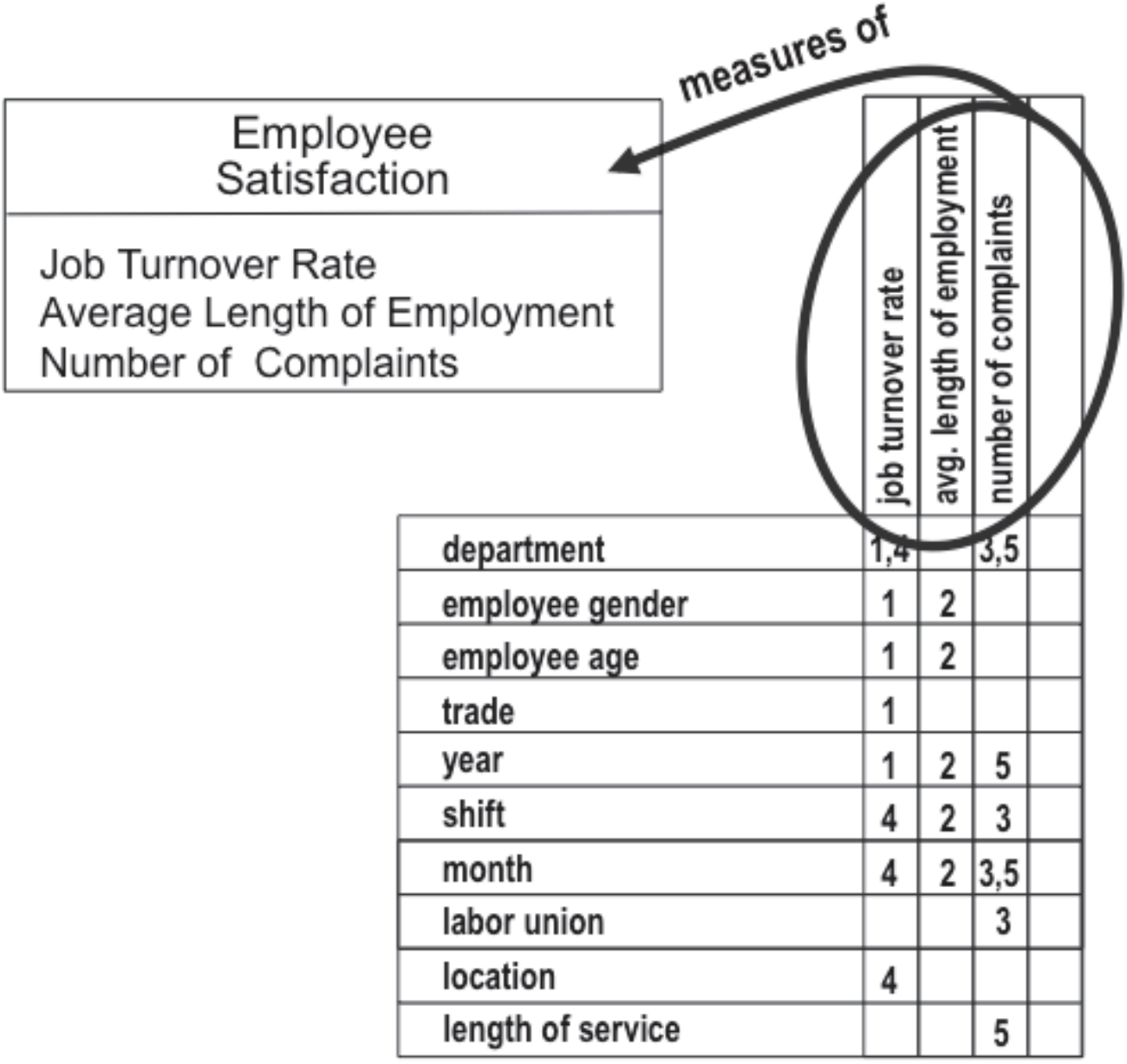
Similar to the class words of common data elements (date, code, etc.) metrics have class words (e.g., cost, value, duration, count) that describe the kind of quantity that they express.

Module 3

Logical Dimensional Modeling

Modeling Meters and Measures

A Group of Related Business Measures



Modeling Meters and Measures

A Group of Related Business Measures

FINDING METERS

The first step of developing a logical dimensional model is identifying meters and measures. A meter is a group of related measures that can collectively be given a business name. The set of facts in the fact/qualifier matrix are the source of measures. Group those facts that are measures of the same business performance concept—EMPLOYEE SATISFACTION, CUSTOMER LOYALTY, PRODUCT PROFITABILITY, WORKFORCE PRODUCTIVITY, etc. Each business performance concept becomes a meter and the set of related facts the measures contained by the meter.

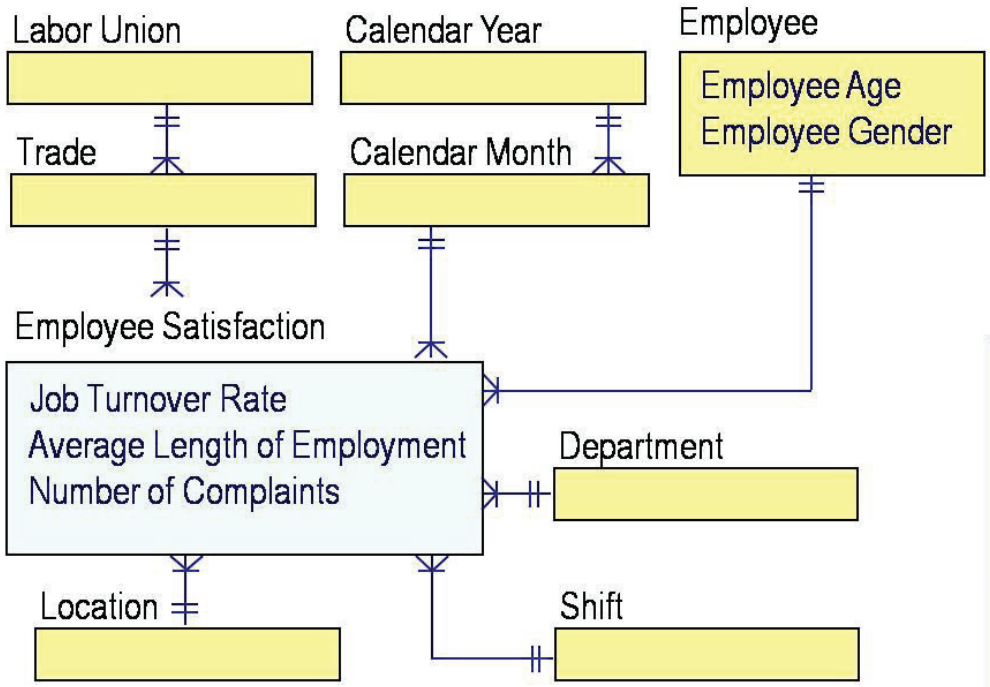
The example domain has been limited to EMPLOYEE SATISFACTION; thus, all of the facts are related by that domain. Sometimes business questions are listed and fact/qualifier analysis performed for a broader or more complex domain. You may, for example, have a set of questions that yield facts about both CUSTOMER VALUE and PRODUCT PROFITABILITY—a combination of facts that aren't readily combined in a single meter.

A further examination of questions 6 and 7 reveals that these are not measures of EMPLOYEE SATISFACTION. Question 8 is already addressed by question 3. These are therefore excluded from the scope of this model and would be addressed by a separate model.

This relatively simple example yields one EMPLOYEE SATISFACTION meter containing three measures: *job_turnover_rate*, *avg_length_of_employment*, and *number_of_complaints*.

Modeling Dimensions

Adding Dimensions from the Qualifiers



	job turnover rate	avg. length of employment	number of complaints	
department	1,4		3	
employee gender	1	2		
employee age	1	2		
trade	1			
year	1	2		
shift	4	2	3	
month	4	2	3	
labor union			3	
location	4			

Modeling Dimensions

Adding Dimensions from the Qualifiers

DIMENSIONS IN THE LOGICAL MODEL

Once dimension hierarchies are known, the logical dimensional model is extended by adding dimensions and associating them with the meter. For every measure contained in the meter, all associated qualifiers must be represented by a dimension. To add dimensions to the model:

1. Include each dimension hierarchy previously identified. Associate only the lowest level of each hierarchy with the meter as a one-to-many relationship.

2. Examine the remaining qualifiers to find those that may be dimension attributes instead of dimension levels. In this example, *employee gender* and *employee age* are attributes that describe *employee*. Thus *employee* becomes the dimension level, with age and gender modeled as attributes. The dimension level *employee* is associated with the meter as a one-to-many relationship.

3. All remaining qualifiers are mapped as flat (single-level, non-hierarchical) dimensions, and each is associated with the meter using a one-to-many relationship.

tdwi

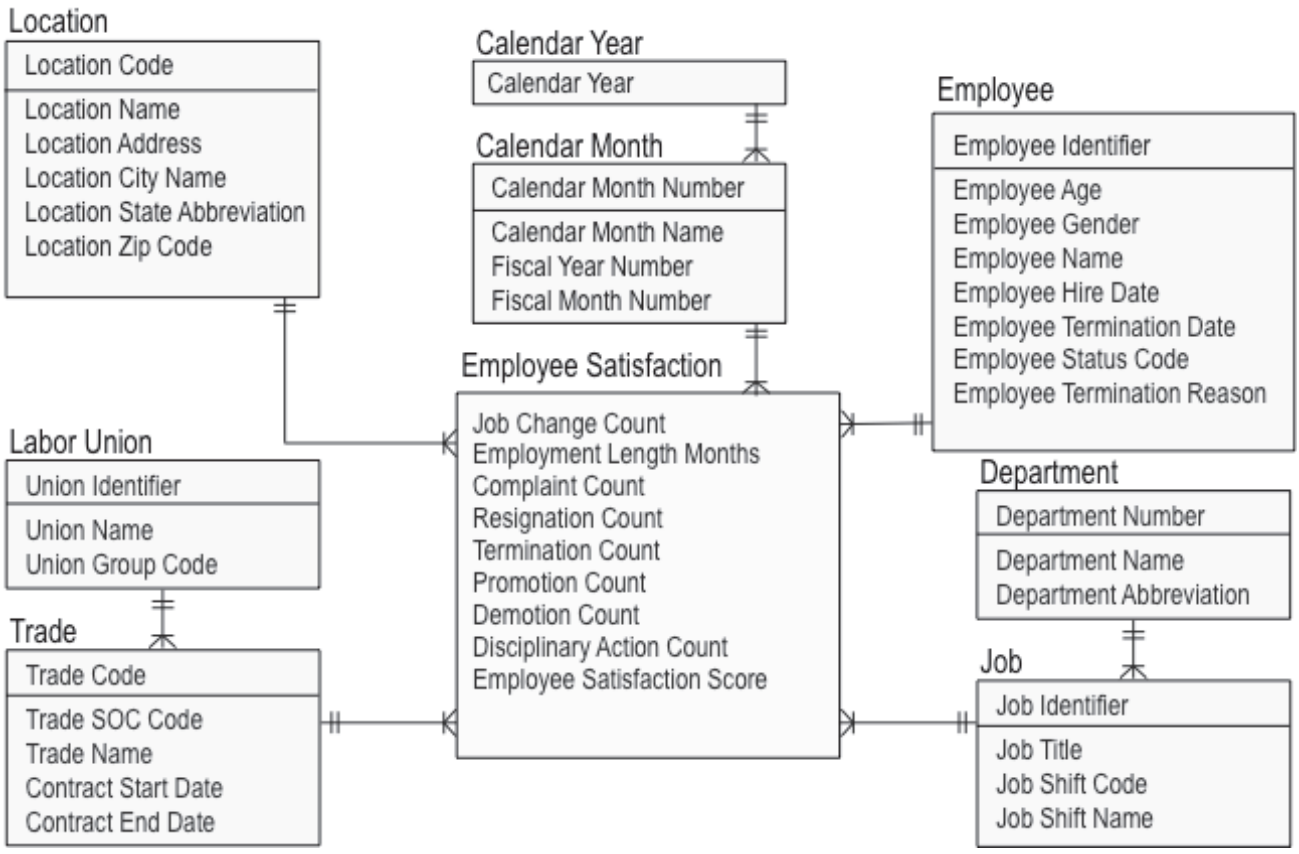
Transforming Data With Intelligence™

Module 4

From Logical Model to Star Schema

Star Schema Dimensions

Naming the Dimensions



Star Schema Dimensions

Naming the Dimensions

FROM LOGICAL TO PHYSICAL

Logical dimensional models are typically implemented in either of two forms:

- Multi-dimensional databases, also called cubes, are used with multi-dimensional OLAP (MOLAP) tools. For these implementations, the MOLAP tool includes a cube generation capability. The logical dimensional model captures the information needed to describe a desired cube to the tool.

- Star schema databases are used with relational OLAP (ROLAP) tools. For these implementations, physical modeling activities design the star schema.

NAMED DIMENSIONS

With either implementation, each dimension needs to be named. Dimension names are key navigation components for business analysts and others using OLAP tools; thus, it is important that the names have business meaning.

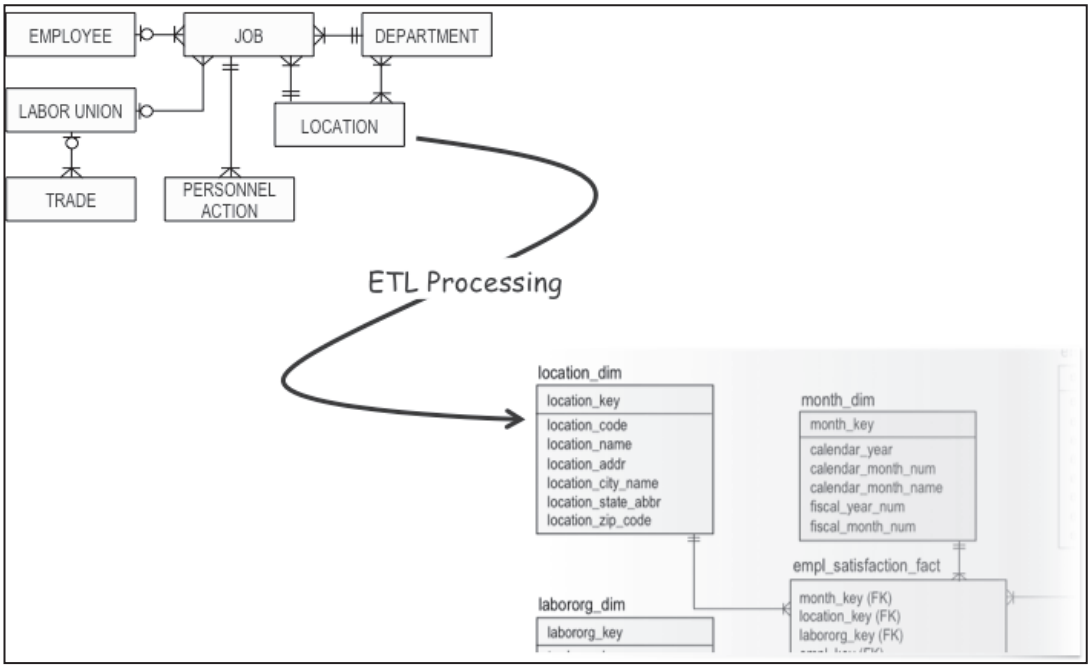
In flat (single-level, non-hierarchical) dimensions, it is common practice for the dimension to have the same name as the dimension level. For multi-level dimensions, a collective name that encompasses all dimension levels is established.

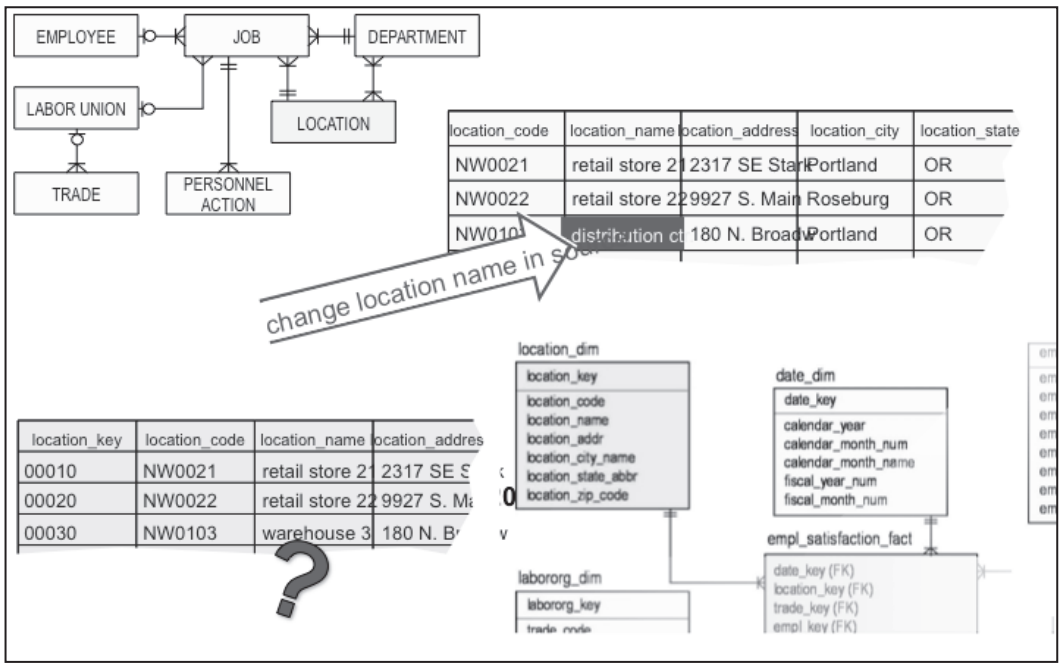
In the EMPLOYEE SATISFACTION example:

- *Employee* is a flat dimension and retains the name *employee*.
- *Location* is a flat dimension that retains the name *location*.
- *Labor union* contains *trade, and* is a multi-level dimension with the collective name *labor organization*.
- *Year* contains *month,* and is a multi-level dimension collectively named *month*.
- *Department* contains *job,* and is a hierarchical dimension named *employment organization*.

Star Schema Design Challenges

Slowly Changing Dimensions





Star Schema Design Challenges

Slowly Changing Dimensions

DESIGNING FOR DIMENSION VOLATILITY

One of the challenges of OLAP analysis is analyzing data over time when the dimensions don't remain constant across the time period of interest. Changing data values are common in dimensional data—a problem known as "slowly changing dimensions." The terminology doesn't mean to imply that some dimensions change more quickly than others; rather, that dimension values change less frequently than the values of measures in a fact table.

When the data values in a dimension change, three options are possible to implement the dimension. Using Ralph Kimball's widely accepted terminology, these options are:

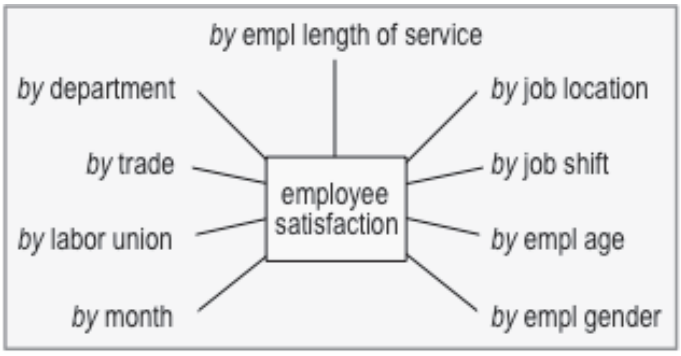
- Type 1 dimensions that overwrite dimension rows without retaining a history of dimension changes.
- Type 2 dimensions that preserve dimension history by inserting new rows into the dimension table.
- Type 3 dimensions that keep versions of dimension values by creating new columns for changed data.

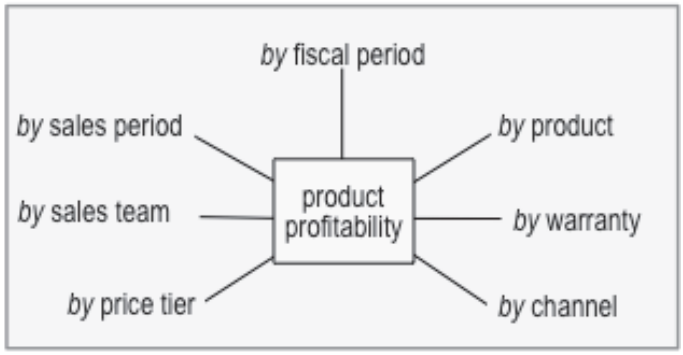
Module 5

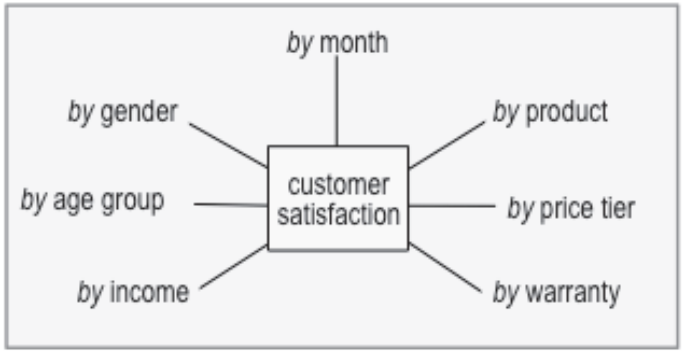
Dimensional Data and Business Analytics

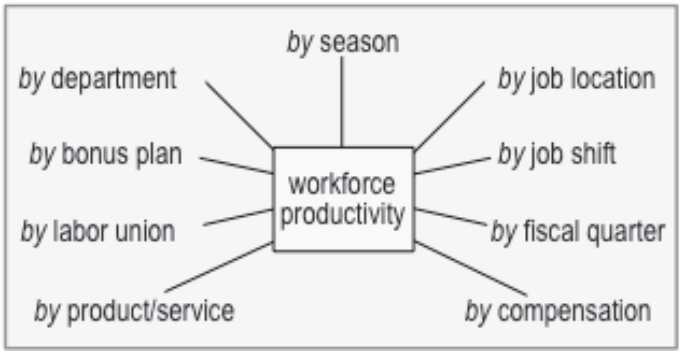
Delivering Business Value

Data Enabled Business Analysis









Delivering Business Value

Data Enabled Business Analysis

MANY DATA MARTS

A typical business intelligence environment encompasses many different data marts at different levels of detail, with different focus and purpose, yet with substantial overlap of dimensions and some overlap of measures. The example that we've followed throughout the course—employee satisfaction—is but a single data model in a more complex environment. The employee satisfaction model is a summary, and it is likely that a data mart exists to contain more detailed, atomic level data. It is also likely that many other data marts exist with similar measures and similar dimensions. Consistency and conformity across data marts is essential to support cross-functional business analysis.

Effective Dimensional Modeling

Critical Success Factors

- ✓ Active business involvement
- ✓ Commitment to adopting conformed dimensions
- ✓ Data modeling tools
- ✓ Skilled data modeling staff
- ✓ Time-boxed approach
- ✓ Business focus
- ✓ Understanding of business analytics concepts

Effective Dimensional Modeling

Critical Success Factors

KEYS TO SUCCESS The facing page lists several best practices and success factors for high-impact dimensional modeling.

