Previews of TDWI course books are provided as an opportunity to see the quality of our material and help you to select the courses that best fit your needs. The previews can not 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. The pages shown are not consecutive. The page numbers as they appear in the actual course material are shown at the bottom of each page. All table-of-contents pages are included to illustrate all of the topics covered by a course.

TDWI Dimsional Data Modeling Primer

From Requirements to Business Analytics

TABLE OF CONTENTS

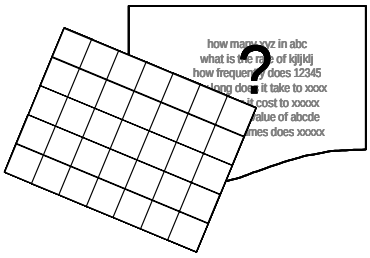
Module 1

Dimensional Modeling Concepts

Dimensional Modeling Basics

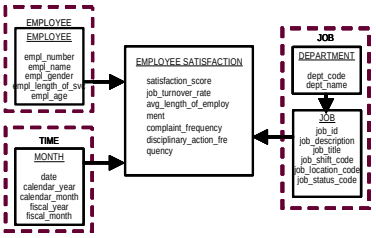
Dimensional Data Models

Conceptual Data Model



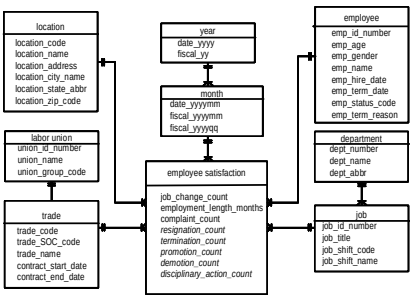
- an analysis process
- model of business requirements
- starting from vague and uncertain
- evolving to specific and certain
- business questions list & fact/qualifier matrix

Logical Data Model



- a business (non-technical) design process
- model of business solution
- starting from specific business requirements
- evolving to product specification
- logical dimensional data model

Physical Data Model



- an implementation (technical) design process
- model of technology solution
- starting from product specification
- evolving to database specification
- star schema

Dimensional Modeling Basics

Dimensional Data Models

LEVELS OF MODELING

Dimensional data design, like any other design process, involves a transition from abstract to highly specific. Abstract models are a means to understand requirements, while implementation models must specify the solution precisely. A three-level approach to data modeling works well where:

- Conceptual modeling describes the business needs.
- Logical modeling describes the business solution.
- Physical modeling specifies the technology solution.

CONCEPTUAL MODELS

Conceptual models are produced through analysis of business needs, and are intended to structure business requirements such that they can be verified and used as input to logical data modeling. When designing dimensional data, conceptual models include a representative list of business questions and analysis of those questions to understand them as a collection of data facts and data qualifiers.

LOGICAL MODELS

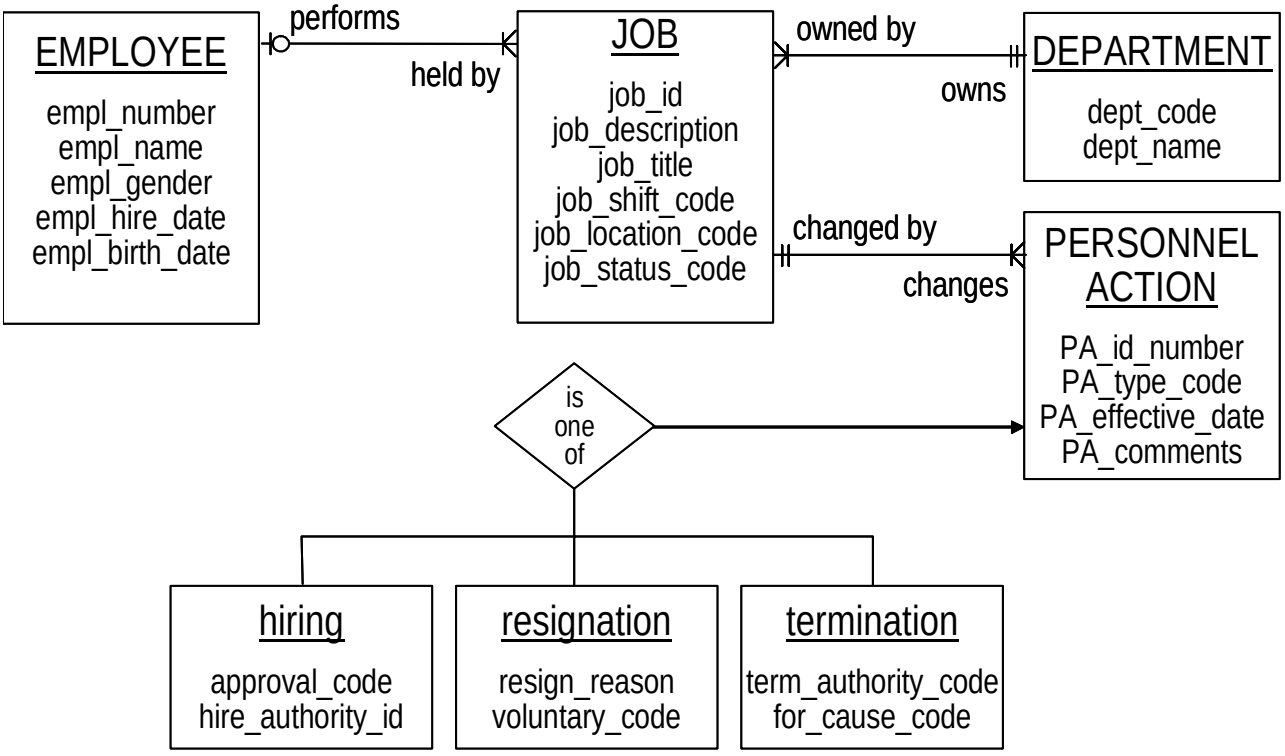
Logical modeling is a point of transition. The modeling activities change from analysis (understanding the needs) to design (understanding the solutions needed to satisfy those needs). A logical model is analogous with a product specification, describing *what* is to be produced but not detailing *how* it is to be assembled. A logical dimensional model meets this objective for design of dimensional data.

PHYSICAL MODELS

Physical models are produced by technical design processes. They describe and specify technology solutions. The physical design process transforms a logical model into a specification for implementation. Physical modeling adds the details necessary to describe how a product is structured and assembled. When designing dimensional data, the most common and widely accepted physical model is a star-schema.

Comparing Relational and Dimensional Models

A Quick Review of Relational Modeling



Comparing Relational and Dimensional Models

A Quick Review of Relational 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_descriptio*n and *job_title,* for example, are attributes of the entity JOB.

- **Relationship**—A relationship is an important association among 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.

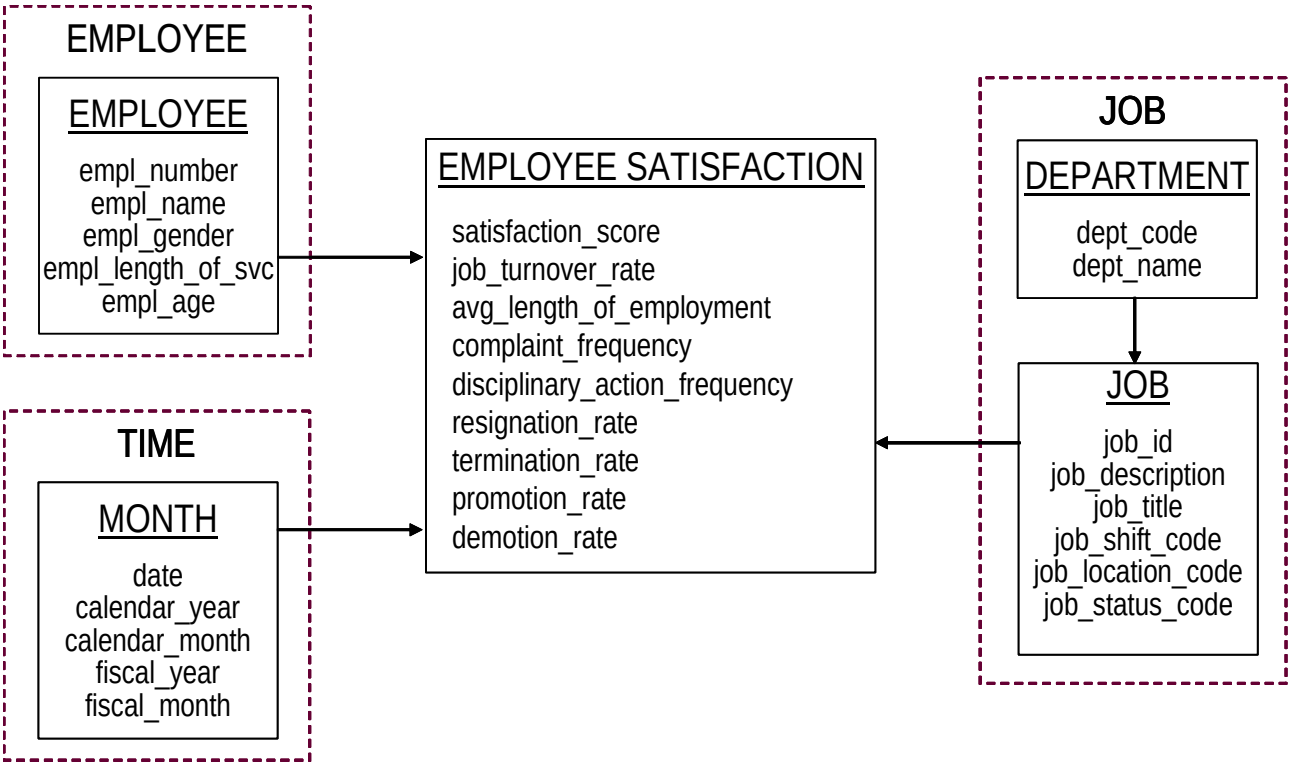
Other important ER concepts include:

- **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.

- **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.

Comparing Relational and Dimensional Models

Introduction to Logical Dimensional Modeling



Comparing Relational and Dimensional Models

Introduction to Logical Dimensional Modeling

COMPONENTS AND STRUCTURE OF THE DIMENSIONAL MODEL

This diagram illustrates a logical dimensional model. The components of the model include:

- A **meter** that contains related **measures** of business interest. Each logical dimensional model has only one meter. In this example, the meter is EMPLOYEE SATISFACTION. Related measures include *satisfaction_score*, *job_turnover_rate*, *avg_length_of_employment*, *complaint_frequency*, etc. Visually, meters in a logical dimensional model look much like entities in the ER model, and measures look much like attributes in the ER model.
- **Dimensions** that provide the means to select, sort, filter, and summarize business measures. In this diagram, the dimensions are EMPLOYEE, TIME, and JOB.
- **Dimension Levels** describe hierarchies that exist within dimensions. This diagram contains one multi-level dimension—JOB—which contains dimensions levels of DEPARTMENT and JOB. The EMPLOYEE dimension contains dimension level EMPLOYEE, and the TIME dimension contains dimension level MONTH. Visually, dimension levels in a logical dimensional model look much like entities in an ER model.
- **Dimension Attributes** include identifiers and descriptive data about dimension levels. Some of the dimension attributes of JOB are *job_id*, *job_description*, and *job_title*. Dimension attributes are diagrammed in the same way as attributes of entities in an ER model.
- **Associations within a dimension** describe the dimension hierarchy and are represented as one-to-many relationships from one dimension level to another—for example, one DEPARTMENT contains many JOBS.
- **Dimension to meter associations** are represented as one-to-many relationships from the lowest level of each dimension to the meter. In this diagram, the examples are MONTH to EMPLOYEE SATISFACTION, JOB to EMPLOYEE SATISFACTION, and EMPLOYEE to EMPLOYEE SATISFACTION. What this means is that each unique set of EMPLOYEE SATISFACTION measures is for one MONTH, one JOB, and one EMPLOYEE.

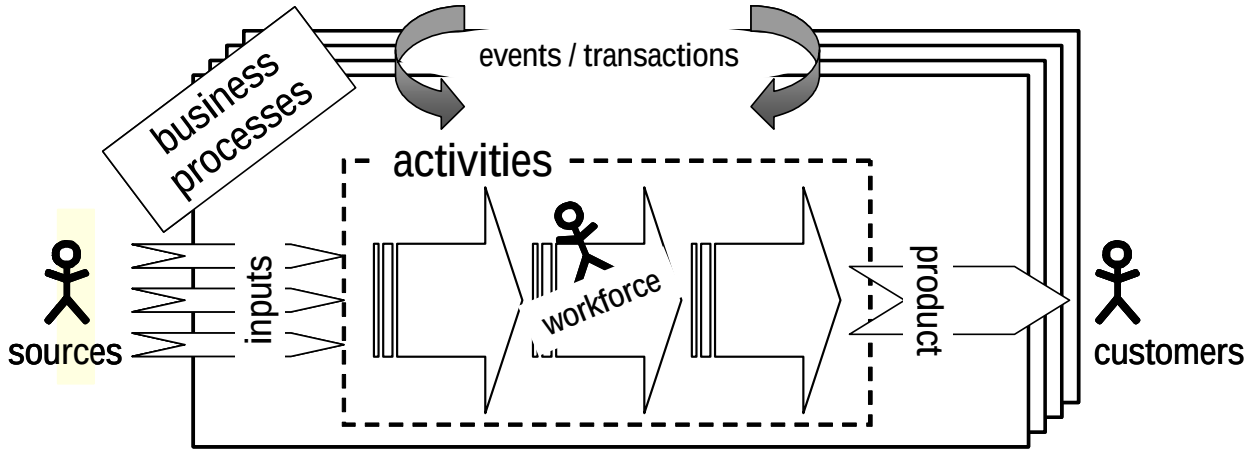
Module 2

Requirements Gathering for Dimensional Modeling

This page intentionally left blank.

Business Context for Data Modeling

Business Processes and Business Metrics



Which business processes are in the scope of modeling?

What can be measured about sources and inputs?

What can be measured about products and customers?

What can be measured about activities and workforce?

What can be measured about events and transactions?

Business Context for Data Modeling

Business Processes and Business Metrics

REQUIREMENTS AND BUSINESS PROCESSES

As discussed in Module One, business processes are the foundation to identify business metrics. The components of business processes—customers and products, sources and inputs, activities and workforce, events and transactions—are the things that can be measured in a business. Business metrics are typically associated with one or a group of business processes. The diagram on the facing page illustrates the components of a business process:

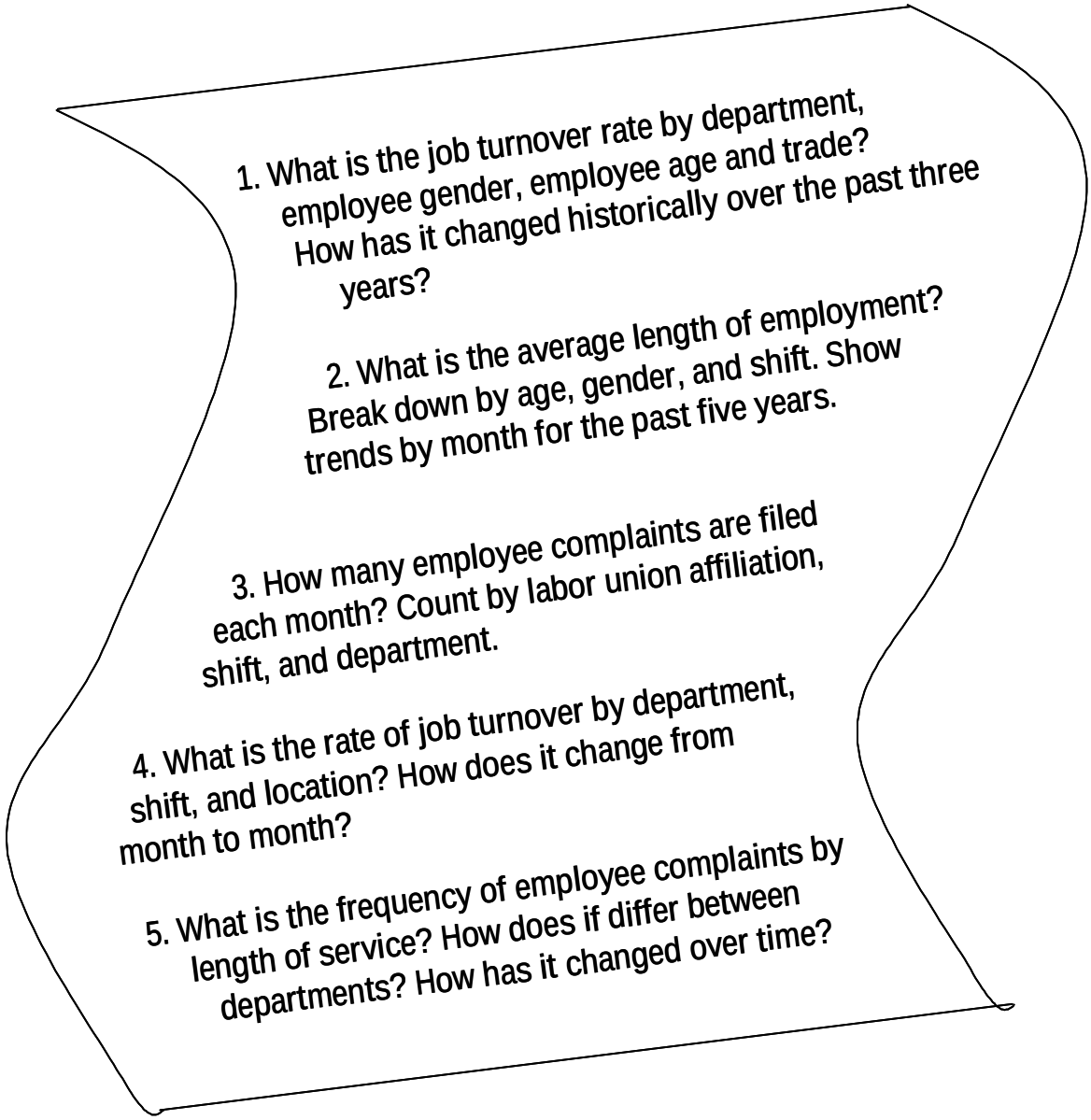
- Customers
- Products
- Activities
- Workforce
- Inputs
- Sources
- Events

Each business process component may be a measurement subject—something that can be measured in the business. Routinely measuring business process components and comparing those measures to targets, thresholds, and past performance is the foundation of business metrics.

Business drivers, goals, and strategies collectively establish context for the program. The drivers describe reasons; they help to define the business case. Goals describe measurable business results; they help to determine metrics and information needs. Strategies describe kinds of actions; they help to identify targeted business processes. Together, from a metrics perspective, they describe why to measure, what to measure, and where to measure.

Business Questions as Requirements Models

Examples



Business Questions as Requirements Models

Examples

SAMPLE QUESTIONS

Although we don't normally think of a list as a model, a list of business questions is a model of requirements for business information. It is not practical to develop an exhaustive list of all questions that might ever be asked in a specific domain. Fortunately an exhaustive list isn't needed. A representative and robust list serves as a model of the kinds of questions that need to be answered. The most significant measures and dimensions of a business domain are readily identified from the business questions. The analysis and design processes of dimensional modeling extend, expand, and refine the collection of measures and dimensions such that the resulting data structure has the ability to answer many questions not found on the original list.

The list of questions on the facing page are a set of examples that we will build upon throughout this course. By the end of this course, you'll see how extension, expansion, and refinement work to create a rich and adaptable data structure from these questions.

Fact/Qualifier Analysis

Mapping Business Questions

2. What is the average length of employment? Break down by age, gender, and shift. Show trends by month for the past five years.

3. How many employee complaints are filed each month? Count by labor union affiliation, shift, and department.

4. What is the rate of job turnover by department, shift, and location? How does it change from month to month?

	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											

Fact/Qualifier Analysis

Mapping Business Questions

EXAMPLE CONTINUED

This example extends the matrix to map three additional business questions from the EMPLOYEE SATISFACTION domain.

Notice that a single fact—*job turnover rate*—appears in both question 1 and question 4. The fact appears only once in the matrix; however, new qualifiers and new fact/qualifier associations are added as a result of question 4.

Also observe that many of the qualifiers apply to multiple business questions. Each qualifier appears only one time in the matrix, but participates in many associations with the set of facts.

Further, note that a single association of fact and qualifier, such as j*ob turnover rate* with *department*, may occur in more than one business question. When this occurs, *all* questions are recorded in the association cell of the matrix.

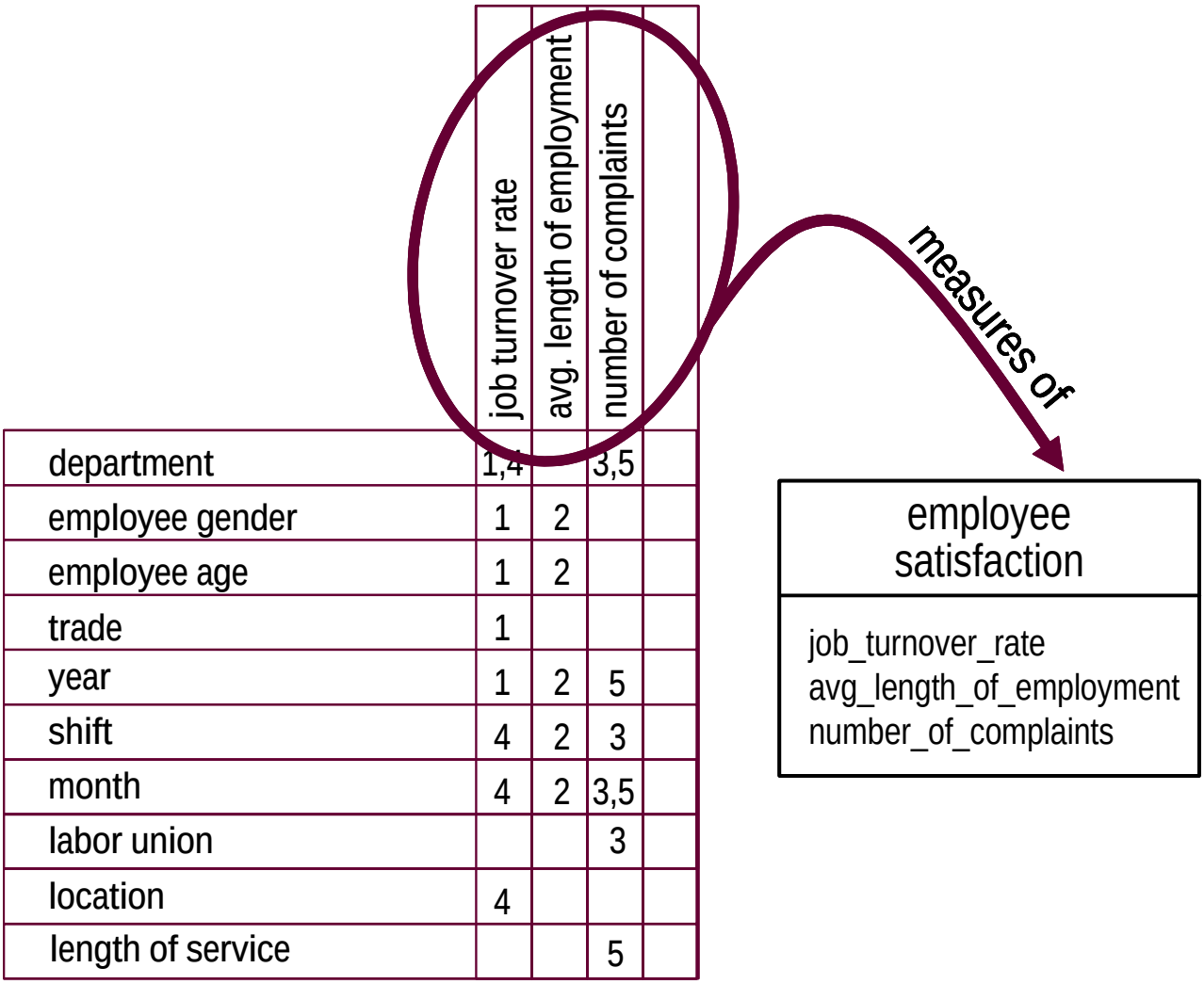
Module 3

Logical Dimensional Modeling

This page intentionally left blank.

Modeling Meters and Measures

A Group of Related Business Measures



Modeling Meters and Measures

A Group of Related Business Measures

FINDING METERS

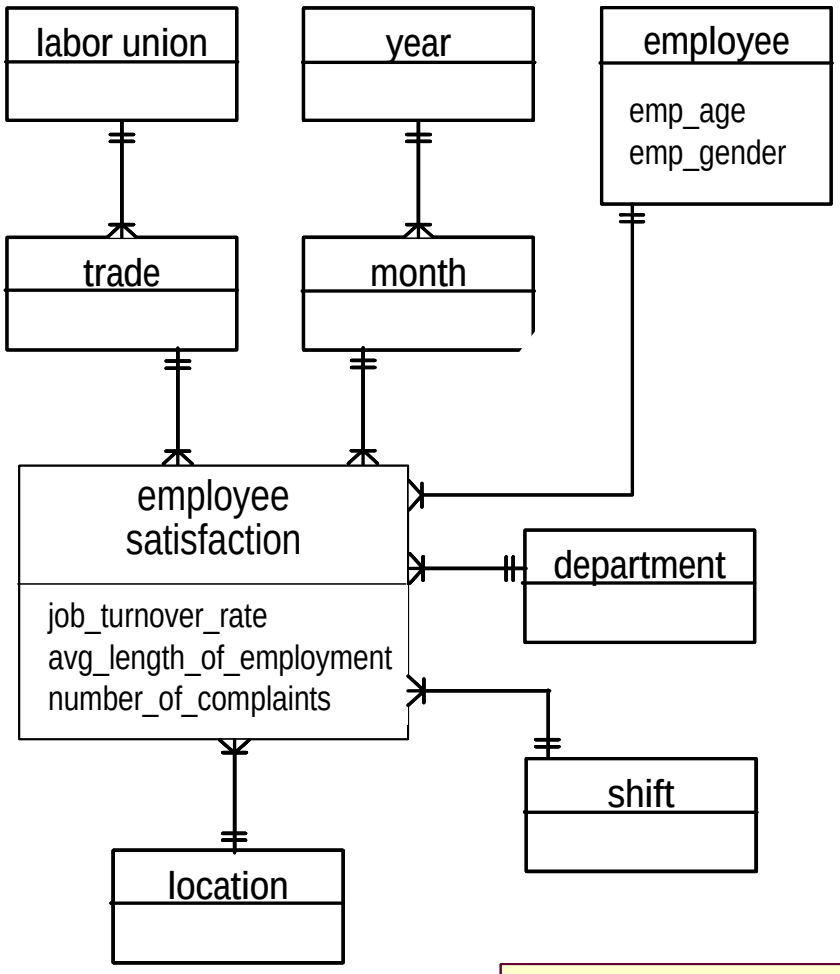
The first step of developing a logical dimensional model is identification of meters and measures. A meter is a group of related measures that can collectively be given a business name. The set of facts in the fact/qualifier matrix are the source of measures. Group those facts that are measures of the 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.

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:

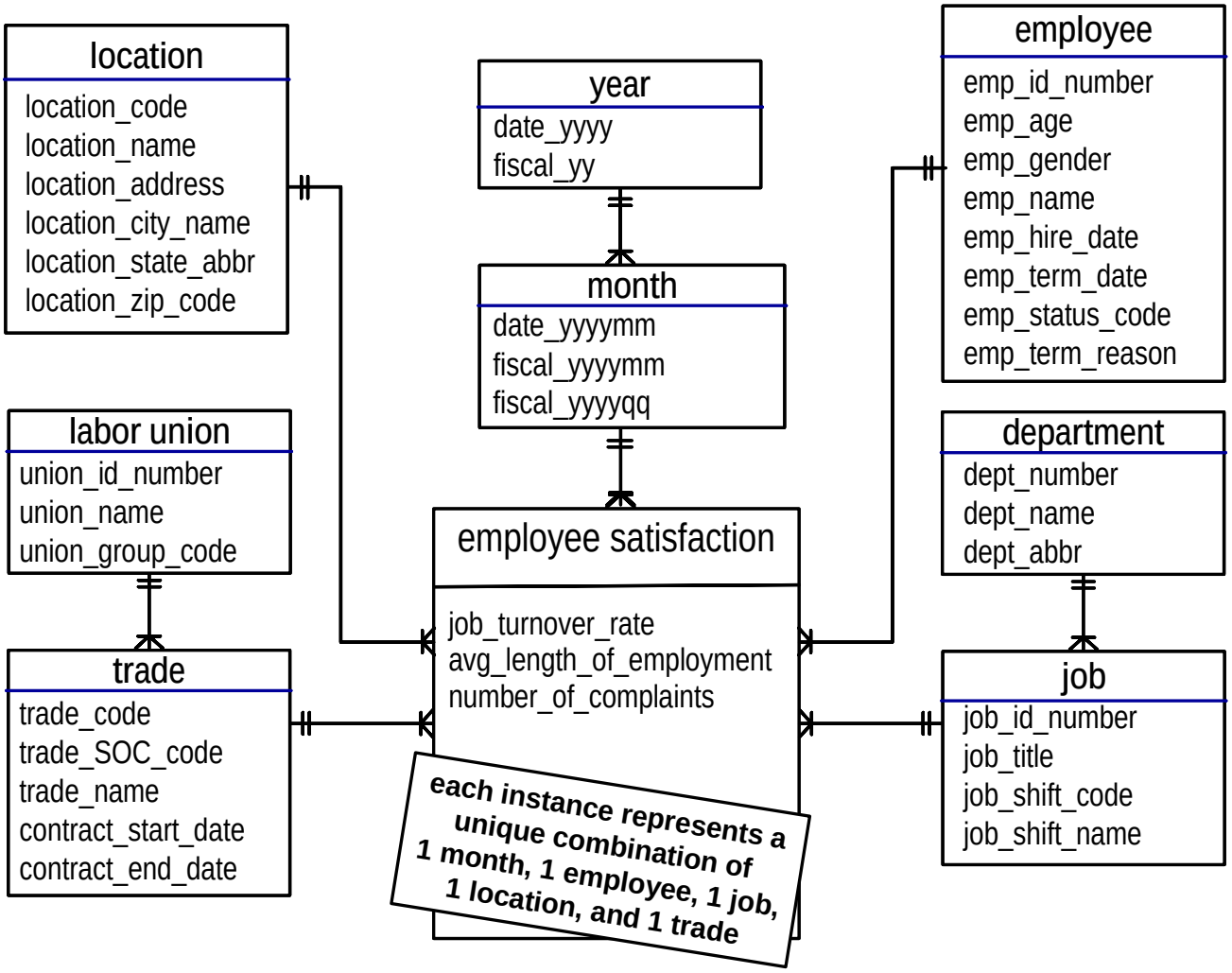
1. Include each dimension hierarchy previously identified. Associate only the lowest level of each hierarchy with the meter as a one-to-many relationship.

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.

More about Meters and Measures

Granularity and the Meter



More about Meters and Measures

Granularity and the Meter

BUSINESS METRICS LEVEL OF DETAIL

Granularity describes the level of detail contained in the meter. Every measure in a meter must be at the same level of detail. Declaring the grain of the meter is an important verification step and yields an essential piece of information for the next steps of dimensional modeling. The grain is determined by the set of dimensions associated with the meter.

By reading the cardinality of meter-to-dimension relationships, it is clear that every instance of EMPLOYEE SATISFACTION contains measures for a unique combination of:

- One month
- One employee
- One job
- One location, *and*
- One trade.

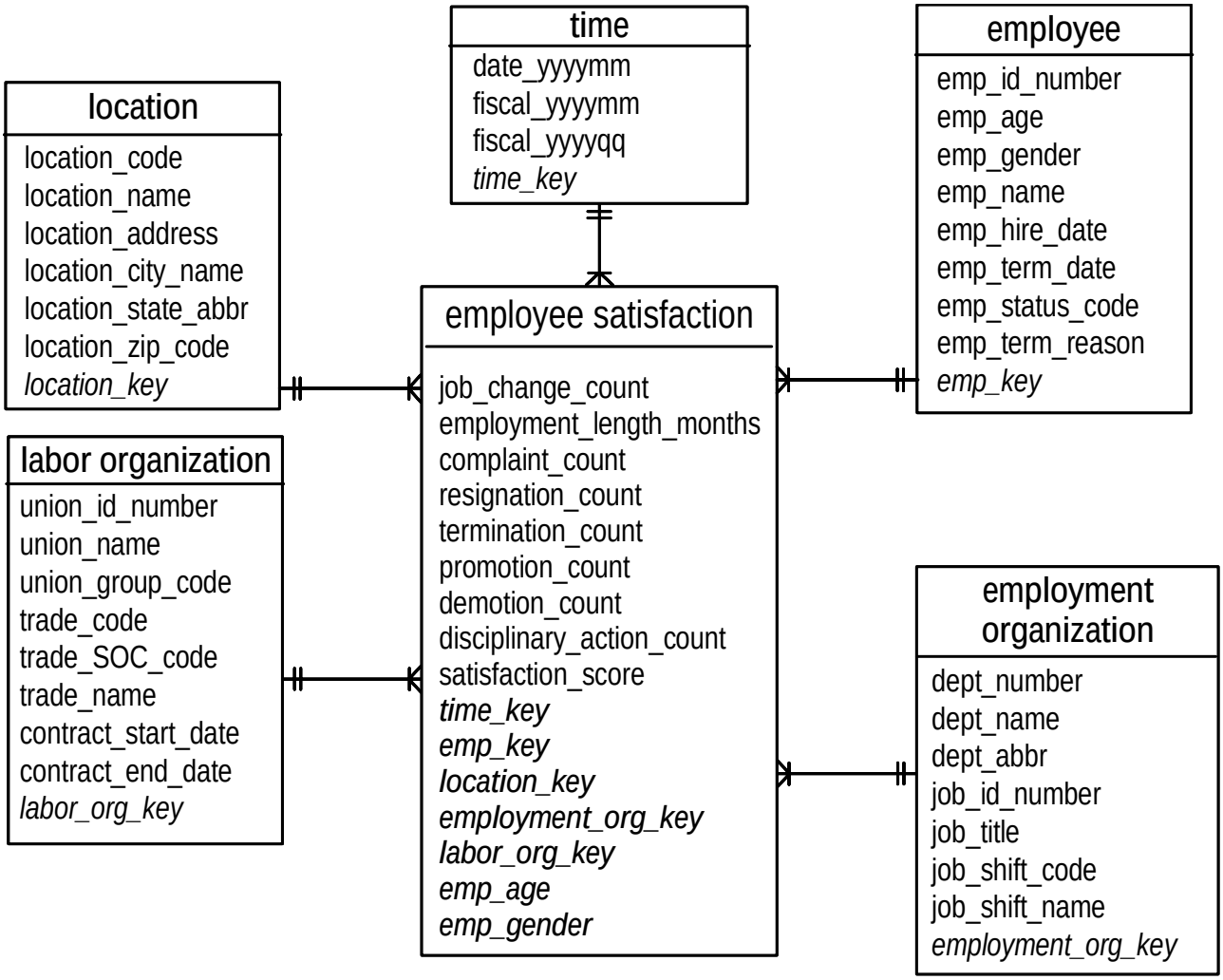
Module 4

From Logical Model to Star Schema

This page intentionally left blank.

Star Schema Fact Tables

Defining the Fact Table Key



Star Schema Fact Tables

Defining the Fact Table Key

DIMENSION TO FACT NAVIGATION

The foundation of OLAP processing is navigation from a set of dimensions to the facts associated with those dimensions. For example, the business question:

> 4. What is the rate of job turnover by department, shift, and location? How does it change from month to month?

asks that a specific measure (rate of job turnover) be reported by a group of dimensions (department, shift, location, and month). Each unique value of *job_turnover_rate* is determined from a unique combination of key values for all of the participating dimensions.

Note that *job_turnover_rate* does not exist in the star schema on the facing page. It is a measure that must be calculated based on the values of other data in the fact table. Derived measures are discussed on the next set of pages.

A COMPOSITE KEY

The fact table key is simply the composite of all dimension table keys—a concatenation of dimension surrogate keys. This works well because the keys are used only for navigation by the OLAP tool and are never exposed to a business user of the tool.

TABLELESS DIMENSIONS

Note that the dimensions tables for employee age and gender have been removed. Once the keys *emp_age* and *emp_gender* are migrated into the fact table, there is no need to maintain them redundantly as single column tables. It is also acceptable, but not essential, to remove those elements from the *employee* dimension table.

Star Schema Design Challenges

Slowly Changing Dimensions—Type 2 Example

Location Table as of January 1

location_key	location_code	location_name	location_address	location_city	location_state
00010	NWR001	retail store 1	2317 SE Stark	Portland	OR
00020	NWR002	retial store 2	9927 S. Main	Roseburg	OR
00030	NWWHSE	warehouse	180 N. Broadw	Portland	OR

Location Table as of September 1

location_key	location_code	location_name	location_address	location_city	location_state
00010	NWR001	retail store 1	2317 SE Stark	Portland	OR
00020	NWR002	retial store 2	9927 S. Main	Roseburg	OR
00030	NWWHSE	warehouse	180 N. Broadw	Portland	OR
00040	NWD001	dist. center 1	180 N. Broadw	Portland	OR
00050	NWD002	dist. center 2	440 Coburg Rd	Eugene	OR

Keep Current & History Rows in Dimension Table

Star Schema Design Challenges

Slowly Changing Dimensions—Type 2 Example

KEEPING ALL OF THE HISTORY

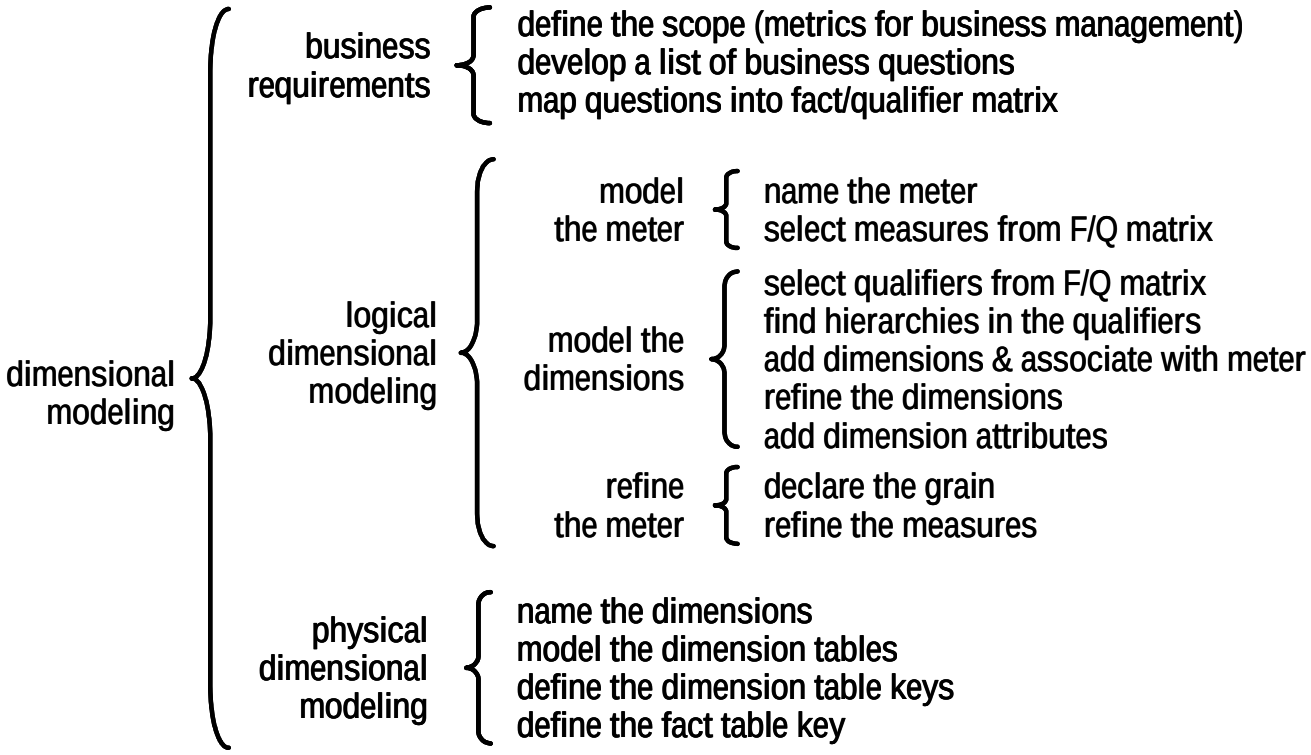
Type 2 dimensions insert a new row, with a new surrogate key value, into the dimension table whenever a change occurs to the value of a dimension element. This is the most common design choice for slowly changing dimensions.

The obvious advantage is that a type 2 dimension preserves all of the history of dimension data changes. For a structure that is also dimensioned by time, there is no need to time-stamp dimension records. Each row of the fact table is associated with only one row of each dimension table, so the time dimension serves as the time stamp.

The most significant disadvantage of type 2 dimensions is rapid growth. Type 2 dimensions may result in very large dimension tables, contributing to both database size and query performance concerns.

Modeling Process Summary

From Business Requirements to Star Schema



Modeling Process Summary

From Business Requirements to Star Schema

PROCESS OVERVIEW

The facing page illustrates dimensional data design activities from requirements through physical design. This diagram provides a simple, easy-to-reference summary of the dimensional modeling process from beginning to end.

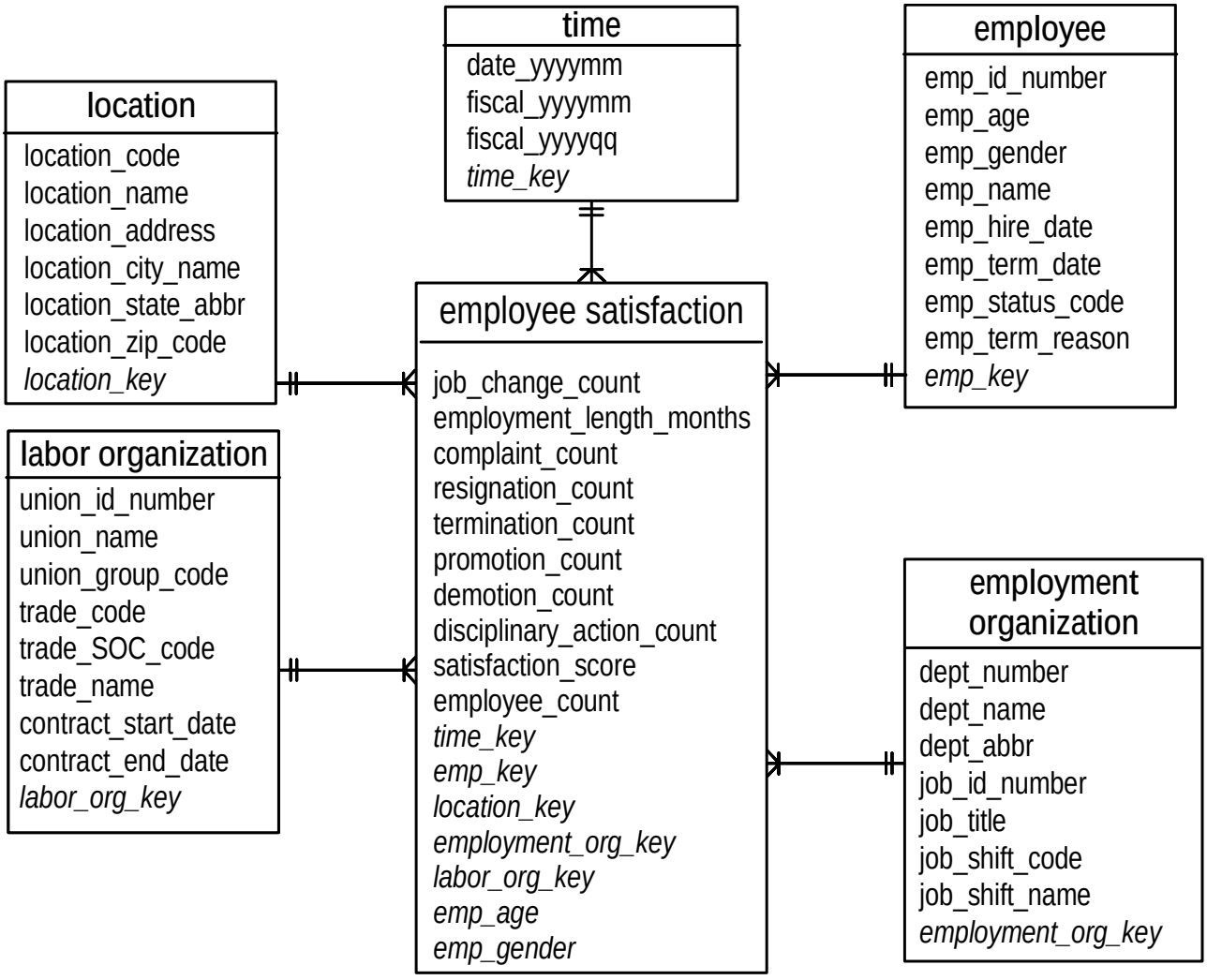
Module 5

Dimensional Data and Business Analytics

This page intentionally left blank.

An OLAP Demonstration

Business Metrics in Action



An OLAP Demonstration

Business Metrics in Action

OLAP DEMO

This OLAP demonstration completes the chain of activity from requirements to business analytics.

NOTES

Exercises

Exercise Instructions and Worksheets

Exercise 2: Business Questions

Instructions

STARTING TO UNDERSTAND REQUIREMENTS

Using the following statement as a basis, develop a list of questions to meet the need.

> Your instructor wants to track her/his performance as a teacher. The goal is continuous improvement as an instructor by understanding what is and is not effective and successful in the classroom. Items of interest include: instructor rating by student evaluations across time, which courses receive the highest ratings, etc.

The instructor is the business subject matter expert and the source of information for this exercise. You may ask questions or conduct an informal interview to gather the information that you need. Remember to address all the following aspects:

- **People**—Who are the stakeholders for monitoring and evaluating instructor performance? What metrics do they use today? What unanswered questions do they have? What levels of metrics do they need?

- **Performance**—Which classes of metrics are within the scope of modeling—Financial? Process? Customer? Growth? How do stakeholder questions relate to these four categories of performance management?

- **Process**—Which business processes are within the scope of modeling? What can be measured about those processes? What metrics are needed to effectively manage the processes? What metrics are needed for knowledge workers to perform the process activities? Which process components make good metrics subjects?