CHAPTER 7

Modeling the Data for your Business

Regardless of what technology will be used, the heart and soul of the data warehouse is the data itself. How the data is organized can have a significant impact on how well the environment will work. A great deal of thought and care must go into the design of the data. This chapter presents the concepts covering how the data should be organized to support reporting and analysis. The business team will learn the following:

- Purpose of dimensional models
- Basic components of a dimensional model
- How to contribute to the modeling process

A case study of the dimensional model for call center operations is used to reinforce the concepts presented. This chapter also includes content that is geared toward the more technical members of the project team, including the following:

- A technique to document the dimensional model, specifically geared toward the business community
- A process for developing the dimensional model with participation from the business community
- An introduction to several advanced modeling concepts
- How to take the model forward

This chapter is not intended to provide an in-depth guide to dimensional modeling. It does share the basic concepts that can be used to communicate more effectively between the business and technical team members. Before diving into the modeling concepts, the first section reviews the rationale for dimensional modeling and why it is of interest to business people.

The Purpose of Dimensional Models

A specific modeling technique has evolved in order to support the types of queries and analyses that businesses require. This technique is called *dimensional modeling*. This approach has been applied to data warehousing for nearly thirty years and is supported by a wide variety of database platforms and data access or business intelligence tools. Dimensional models support the business perspective of the data, and today's technology ensures that they can be effectively implemented.

Dimensional modeling is a formal data modeling technique that is used to organize and present data for analytical and reporting use. The focus is on the business perspective and representation of data. The goal is to free the data that has been captured and stored by the operational systems and make it available to the business community. Regardless of how data is structured, business people will ask questions based upon their frame of reference. This perspective is driven by the basic characteristics of the industry and how the company is organized, so why not organize the data to reflect this business perspective? The two primary goals for dimensional modeling are *ease of use* and *query performance*. These are the principles that guide the entire dimensional modeling process.

There are other data modeling techniques that play an important part in overall systems development. They help ensure that the data itself and the relationships between different data elements are clearly defined. For operational systems, it is important that the data be organized to facilitate transaction processing. This includes ensuring transaction integrity and speed. The type of modeling used for operational system design is called *entity-relationship (E-R) modeling*. This may also be referred to as *normalized modeling*. One specific form of E-R modeling represents the data in *third normal form (3NF)*. There is a complete discipline surrounding this approach to data modeling. This is mentioned to acknowledge the value and purpose of E-R modeling for operational system design. The following two sections examine the main objectives of dimensional modeling.

Ease of Use

In order to ensure that people will use a data warehouse, the data must be presented in a manner that makes sense to them. If it is too confusing or does not mirror the way the business runs, then people are not likely to use it. Therefore, the dimensional model must cleanly represent the basic components of the business. In addition, the model must be presented in terms that are used by the business.

A well-designed dimensional model should be obvious to the business community and be met with a confirmation that it is indeed correct. If it causes

a great deal of confusion within the business community, then the model has not yet captured the business perspective.

TIP **Don't do what is easier to implement, do what is easier for the business users.**

Query Performance

The second, and equally important, goal of a dimensional model is to ensure good query performance. If requests do not run in a timely manner, the data warehouse will not be used and will not be helpful to the business. Dimensional modeling takes the need for this query performance into account as part of the inherent design approach. The data is organized in order to provide consistent performance both for queries that are requested up front and for those that crop up later. All possible queries cannot be defined in advance, so the technique takes this into account to provide support for unpredictable access patterns.

Understanding Your Data

The requirements gathering process detailed in Chapter 6 identified the different kinds of data that are needed for the DW. The prioritization process, also discussed in Chapter 6, helps ensure that the project charter and scope align with the business requirements, including analyses and the data sources needed. Now, detailed data analysis is needed to really understand what is stored in the operational databases that have been identified. A robust operational database is likely to have hundreds of tables and thousands of individual data elements.

A lot of legwork should be done by the core project team to determine which data elements may be useful. The systems staff members responsible for maintaining that application can be invaluable in this research. The objective is to narrow down the sheer number of data elements that need to be studied. This is also called *source data analysis*. This can range from running simple queries to much more sophisticated analysis. At a minimum, simple queries should be run against the application system's data structures. The types of queries include the following:

- Has this data element ever been populated?
- When was the last time this table was updated?
- What percentage of the rows in this table contain data?
- What are the possible values for this data element?
- What is the frequency with which these values are used?

The results of these queries can be used to determine which elements would be useful for reporting and analytical purposes. If the data element has never been populated, then this does not need further study. If a data element has not been changed in years, this may indicate core reference data that is still valid or it may indicate that this data is no longer being captured and/or maintained by the operational system. Obsolete data elements do not need to be addressed in the modeling process.

It is also helpful to have business analysts participate in this analysis. They can often help the team to understand the data content. Over time, the original name and meaning of a column can be lost or changed. It is helpful to know how the data is used today. The objective is not to isolate data elements that have been requested for reporting, but to identify all data elements that contain business data. This is often done through an elimination process. For example, data elements that are obsolete or those that are used to drive security for the operational system do not contain business data that would need to be included in the data warehouse.

Data profiling is a more comprehensive set of activities intended to look at data across all systems, operational and analytical. This is described in more detail in Chapter 8.

TIP Look at all the data elements, not just those specifically identified for reporting.

After there is a clear understanding of what the business is trying to accomplish and a familiarity with the data that is available, it is time to look at how that data should be organized.

What Is a Dimensional Model?

A *dimensional model* is a data model organized for the purpose of user understandability and high performance. There are two basic parts of a dimensional model: the *dimensions* and the *facts*. These are the building blocks that comprise all dimensional models, simple or complex.

Dimensions

Dimensions are groupings of data elements in major business categories. Common dimensions include the following:

- Customers
- Products
- Dates

- Suppliers
- Vendors
- Accounts

The individual data elements are called *dimensional attributes*, or *reference data*. The dimensional attributes are used as row and column headings for reports. They are used to create lists of options to determine what to include or exclude on a report. The relationship between these dimensional attributes creates drill paths or the ability to navigate up and down a hierarchy.

The need for dimensional data is often recognized while gathering business requirements. It may not be directly communicated, but realized when someone needs a report by region, by week, and by product category. Each of the terms following the word ''by'' is a dimensional attribute. These should be included in dimensions to support that type of reporting.

An example of a customer dimension is shown in Figure 7-1. This is a highly simplified example that only shows the customer's address and date of birth attributes. Some of these attributes relate to each other in a hierarchy, while others are simply additional characteristics of the customer. Any of these attributes could be used to constrain a query or for use in a report.

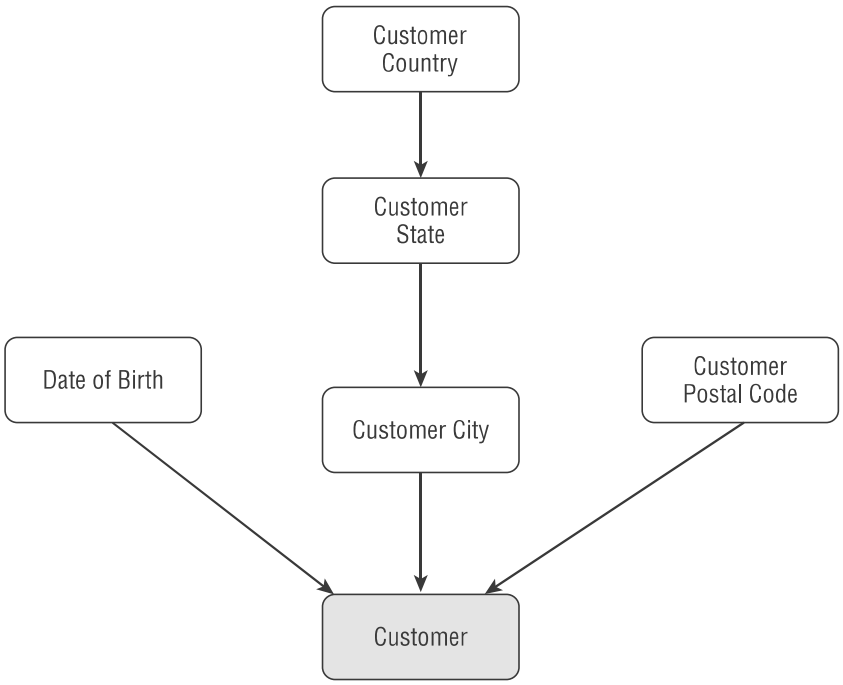

Figure 7-1 Sample customer dimension

TIP **A dimension can include attributes that are descriptive and that relate to each other, creating hierarchies.**

The arrows represent one-to-many or one-to-one relationships. This notation is used because it is more meaningful to business users to visualize the drill paths, rather than the more traditional representation of ''crow's feet.'' In fact, if looking at the arrows causes technical staff members to furrow their brows, this is the same reaction that crow's feet have to business staff members. The intent of the business dimensional model is to communicate with the business.

Facts

Facts are the measurement of business events. These are captured as specific information about a business event or transaction. These are measured, monitored, and tracked over time. Facts are typically the amounts and counts that show up as the body of reports. Facts are used as the basis for all calculations. Examples of facts include units ordered, retail price, amount paid, claim payment amount, gross margin, budgeted dollars, revenue forecast, and loan balance, among others.

The facts are only interesting within the proper context, and the context comes from the dimensions. For example, the fact that a company had $10,000 in sales is not useful unless you know that it was from red shoes, in the Chicago market, the week before Valentine's Day.

Using Both Parts of the Model

Dimensional models can represent very complex businesses. Although there can be many different aspects of a business, most people deal with only a few variables at a time. Most of us can easily draw a three-dimensional cube, but it is much more difficult to draw a four dimensional cube. Similarly, while the overall model may contain over ten dimensions, often only three are reflected on a single report at a time.

The way that people think about data is often defined by the layout of a spreadsheet: rows and columns with perhaps a separate worksheet to represent another variable. For example, Figure 7-2 is a common sales performance report showing the monthly profit results for the current year, reported by product category. Each region of the organization is represented as a separate page. This report is constrained to the current year of data, and uses product, date, and sales organization dimensions. The Product Category attribute is from the Product dimension. The constraint for the current year is on the year attribute of the Date dimension. The months are also from the Date dimension. The Sales Regions (per sheet) are from the Sales Organization dimension. The fact itself is the Sales Profit, listed in thousands of dollars.

Sales Profit for Northeast Region
For Current Year to Date
(*Profit in thousands of Dollars*)

	Months						
Product Category	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**
Camping Accessories	19	22	37	52	65	84	83
Women's Clothing	63	68	87	65	62	74	69
Men's Clothing	72	81	80	94	87	103	78
Athletic Shoes	201	214	194	183	191	192	199
Fishing Accessories	6	3	8	11	21	17	18
Backpacks	3	4	4	12	16	19	15
Tents	88	83	91	137	139	189	120
Total	452	475	501	554	581	678	582

Figure 7-2 Sample Sales Results Report

The dimensions and facts are used together to create basic reports such as this one, or more sophisticated analyses. More detail about using dimensional models can be found in the section ''A Call Center Case Study.''

Implementing a Dimensional Model

Many different database technologies are available today to store data. Many of these have been developed specifically to support data warehouses. It is useful to have a basic understanding of these in order to be able to put the dimensional model into the appropriate context. A dimensional model is not inherently tied to a specific technology, and it can be implemented in a variety of different ways. Different types of databases include the following:

- **Relational databases** are one of the most commonly used databases for both operational and data warehouse systems today. When a dimensional model is stored in a relational database it is called a *star schema*. This is due to the visual appearance of the dimension tables surrounding the fact table.
- **Multi-dimensional databases** are specifically designed to support a dimensional view of the data. The data is stored in proprietary array structures. When a dimensional model is stored in a multi-dimensional database, it is called a *cube* (even though it may actually contain more than three dimensions).
- **Proprietary databases** are also available on the market. Many of these are designed specifically to support reporting and analytical use. There is a wide range of different methods used to physically organize and access the data in these environments. Once the dimensional model is completed, there is often no additional design work required to determine how the data will be structured in these environments. This is all handled by the proprietary system.

When database software and hardware are fine-tuned and bundled together, they are called a *data warehouse appliance*. These concepts are included here because they are commonly used and most readers have likely heard several of these terms.

Diagramming Your Dimensional Model

There are different ways to document and present dimensional models. One of the most common ways that dimensional models are depicted are as tables to be stored in a relational database. The dimensional model can be documented using the same modeling tool that is used to develop any other data models for the relational database. Each of the dimensional attributes is included and represented using logical names that should be meaningful to the business. This type of table diagram is easily understood by systems professionals, but it is not as clear to business professionals.

Another method to document your dimensional model is to present business diagrams. The intent is to visually present the model in terms that more closely reflect the interface that will ultimately be presented for access. This is called *business dimensional modeling*, and it can be documented using any visual diagramming tool.

Careful analysis must be performed and dimensional modeling principles must be followed in either case. The primary difference is how the model is presented to the business. The diagram notation of the business dimensional model is reviewed next. The process to develop a dimensional model is discussed later in this chapter.

The Business Dimensional Model

With the increasing variety of options for building the data warehouse, it is important to split the business perspective from the technical perspective regarding the data. The *Business Dimensional Model (BDM)* is a data model that is specifically geared toward working with the business community. It serves as an abstraction layer that insulates the model from technical implementation details. The model also serves as a communication vehicle between the business and systems groups. The model shows diagrams of the dimensions and the facts so that the details can be reviewed and discussed in business terms. This also separates the business discussion from any technical discussions.

The business side of the problem can be addressed in enough detail so that the systems team can implement the model in any technology. Limited technical tips are shared at the end of this chapter to guide the project team when implementing the dimensional model in a relational database management

system. Other technical implementation details are not covered in detail in this book. The project team should work directly with the technology vendors for guidelines and recommendations specific to whatever tool is being used.

A dimensional model can reflect a wide range of data from multiple data sources. The focus is on understanding the dimensionality of the business itself and the facts that are needed to measure that part of the business. The business dimensional model will reflect all of the data to be included in the data mart. A summary of all of the notation used for the business dimensional model is provided at the end of the chapter.

Business Dimensions

Each business dimension of the model will be designed and diagrammed separately. The business attributes are included to fully describe the dimension. Each business attribute of the dimension is depicted with a rectangle, as shown in the sample dimensions in Figure 7-3. For each dimension, you need to identify the lowest level of detail that exists. This is also called the *grain* of the dimension. This lowest-level attribute is shaded. The relationship between each of the business attributes is noted with an arrow. The direction of the arrow shows the direction to drill down to see more detail. These are also commonly referred to as *hierarchies*. Note that this organization's fiscal year does not align with the calendar year. The grain for this dimension, or the lowest level of detail, is the Day attribute. Each individual business attribute is included in the dimension diagram.

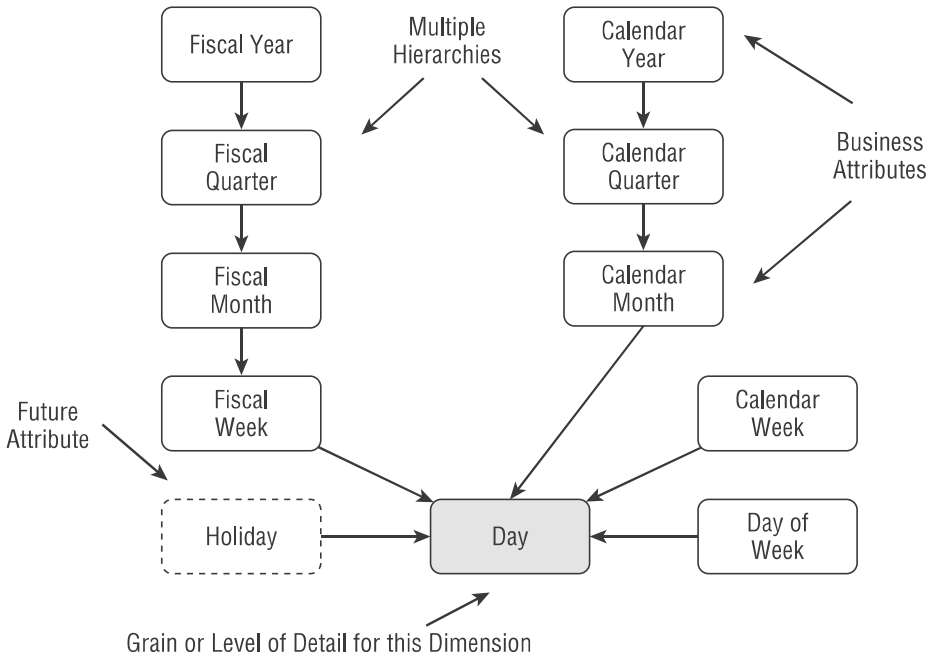

Figure 7-3 Date dimension example

Sometimes a specific business attribute is requested but is not captured in any current source systems. To facilitate communication, this can be included in the dimension diagram, but the notation is different. The dotted rectangle indicates that the data element is either not captured or is not to be included in the initial implementation. In this example, the Holiday attribute is planned for the future. This ensures that the model reflects true business needs, but it also helps set and maintain expectations that the element is not going to be available at this time.

The dimension diagram itself is only part of the documentation that is needed for the dimension. The diagram shows each attribute with a useful business label. In Figure 7-3, note that each of the date hierarchies is uniquely named. This helps ensure that the correct attribute is easily selected for reporting. It is also important to have a clear definition of each attribute, and several sample values. The sample values are often what will spark recognition of what an attribute represents. This does not need to be a complete list of all the possible values for this attribute. Table 7-1 shows the additional documentation that is needed to support the dimension diagram.

DATA NAMES AND DEFINITIONS

The business community must take a lead role in developing clear, meaningful names for each of these attributes. The business must also be responsible for creating the definition of each attribute. These are critical to ensure that the model is easily understandable and accurately documented. Meaningful data definitions are one part of overall data governance. This topic is important enough that Chapter 8 is devoted to discussing data ownership and governance issues.

Fact Groups

The second part of the model contains the facts, which is where the business measurements are stored. Modeling the facts is much more than simply creating a list of the business measures that are needed. Each of these facts must be reflected within the proper context. This can be understood by looking at how the data is captured and how the business uses each fact. The dimensions that are relevant to these facts are shown. The grain, or lowest level of detail that applies, is also identified for each applicable dimension. Often, several facts will have the same dimensionality and identical grain. These facts can be put together into a *fact group*.

Table 7-1 Date Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Day	Calendar day information was reported	6/31/2003, 9/1/2003
Holiday	This is a future attribute to indicate the name of the federal holiday recognized by the U.S. government.	Memorial Day, President's Day
Fiscal Week	The week of the fiscal year that this date rolls up to; used for reporting purposes	FY 2003 Week 16
Fiscal Month	The month of the fiscal year that this date rolls up to; used for reporting purposes	FY 2003-10, FY 2003-11
Fiscal Quarter	The quarter of the fiscal year that this date rolls up to; used for reporting purposes	FY 2003-Q1, FY 2003-Q2
Fiscal Year	The year that this date rolls up to based upon the company's fiscal calendar	FY 2003
Calendar Month	The calendar month used for reporting purposes	March 2003, October 2003
Calendar Quarter	The calendar quarter used for reporting purposes	CY 2003-Q1, CY 2003-Q2
Calendar Year	The calendar year used for reporting purposes	CY 2003
Calendar Week	The calendar week ending Saturday and used for reporting purposes	Week Ending 11/21/1998
Day of Week	Attribute to indicate the calendar day of the week	Wednesday, Saturday

Figure 7-4 shows a fact group for retail sales. The name of the fact group is in the hexagon at the center of the diagram. Each dimension that applies to this fact group is diagrammed as a bubble around the hexagon. The specific grain is also included in each bubble. This is important because facts can have similar dimensionality, but be at different levels of detail. If you have two facts that have the same dimensions but one is daily and the other is monthly, these must be split into two separate fact groups. You must design only single-grain fact groups. Mixing the grain of a fact group can cause query results to be incorrect.

TIP **Never mix the grain of facts in a fact group.**

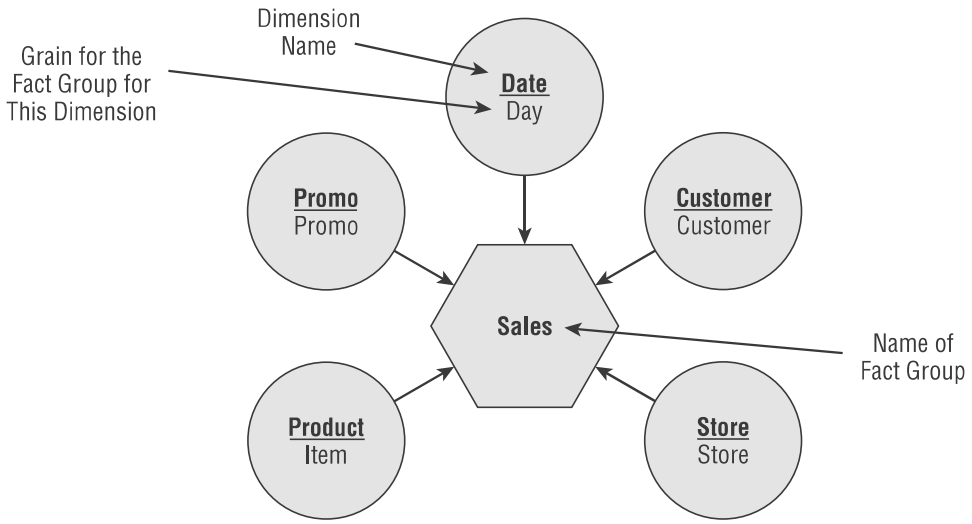

Figure 7-4 Retail Sales fact group

The name and definition for each fact must be documented. Table 7-2 shows this information for a Retail Sales Fact Group. The table also includes a column that indicates what to do with the facts as you drill up and down through hierarchies and group the facts by other attributes from the dimensions. This is the default *aggregation rule* for each fact. The aggregation rule is also needed if a query excludes a dimension entirely. For example, if you want to know the number of units sold by Calendar Month, then the other dimensions are not included in this request. The Units Sold can be summed and then grouped by the Calendar Months to fulfill this request. The most common default aggregation rules are to sum, or count, the fact.

Table 7-2 Retail Sales Fact Definitions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Units Sold	Number of consumer units sold	SUM
Sales Dollars	Dollar amount for the units sold	SUM

Using the diagrams for the dimensions and fact groups, you can see what types of queries can be supported. A more complete case study is shown next to better illustrate this capability.

A Call Center Case Study

Most large organizations have some sort of call center. These centers are sometimes run internally or may be outsourced to third-party organizations.

The call center in this example is run internally to handle customer calls. These calls may be to place an order, check the status of an order, request information, file a complaint, or share a compliment. This case study shows the business dimensional model for this organization. While this shows a mostly complete picture of a dimensional model, keep in mind that it has been simplified to facilitate an introduction to the modeling approach, rather than to represent a complete model for a real call center. Clearly, there are many nuances and details that are specific to each organization that cannot be addressed here.

Call Center Dimensions

The following sections describe each of the dimensions needed. Both the dimensional diagram and the supporting table provide a complete description of each of these dimensions.

Date Dimension

The Date dimension for this call center case study is the same dimension shown previously in Figure 7-3. This supports analysis by fiscal and calendar years. Definitions of each of the attributes are provided in Table 7-1.

Time Dimension

Because call patterns can vary drastically throughout the day, it is important to be able to track the calls at a fine level of detail. Keeping the time of day attributes separate from the date attributes helps simplify the model, reducing the size of the dimension, which provides implementation benefits. The basic hierarchy of the Time dimension shows a straightforward rollup of time. The Time dimension can also be used to track the work shift of the organization. Although the work shifts currently change on the hour, the company is discussing changing the shifts around. In order to accommodate any possible future definition of the work shift, it is shown to the minute. The Time dimension diagram is shown in Figure 7-5.

Time dimensions are helpful to facilitate the representation of different parts of the day. If the only need for time or timestamp data is to calculate the elapsed time between two events, this would not be needed. If specific attributes are needed for grouping and reporting, these can easily be stored in a Time dimension. Also needed are special considerations to accommodate the shift between standard and daylight saving time. Separate rows can be included in the table to represent the second hour between 1:00 A.M. and 2:00 A.M. that occurs when the clocks are set back in the fall.

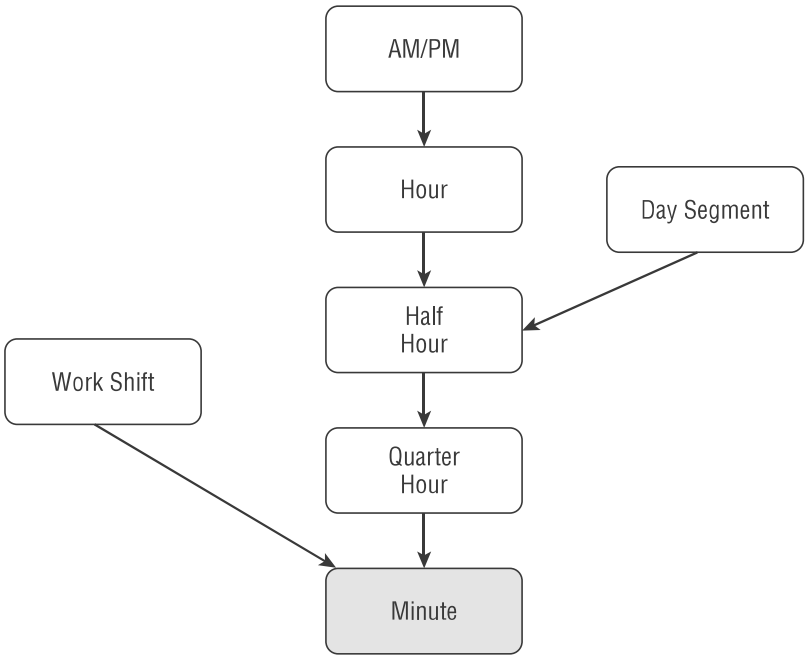

Figure 7-5 Call Center Time dimension

The data definitions for the Call Center Time dimension are included in Table 7-3.

Table 7-3 Call Center Time Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Minute	The specific minute of the day. These begin at midnight and finish at 11:59 P.M.	1:01 A.M.; 6:32 P.M.
Work Shift	Three shifts are worked on a daily basis. This attribute describes to which work shift this minute of the day belongs. These shifts are stored and noted based upon the local time zone. For example, the morning shift runs from 7:00 A.M. to 4:00 P.M. local time. In other words, people are working the morning shift in Maine before the headquarters is even open in Oregon.	Morning, Afternoon

(*continued*)

Table 7-3 (*continued*)

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Quarter Hour	This attribute contains the minutes grouped into 15-minute segments. This is needed to track employee time at the quarter-hour level of detail. This is represented by the beginning minute of the 15-minute range.	4:30 P.M. quarter hour, 4:45 P.M. quarter hour
Half Hour	This attribute contains the minutes grouped into 30-minute segments. This could also be described as two quarter hours. This is represented by the beginning minute of the 30-minute range.	4:00 P.M. half hour, 4:30 P.M. half hour, 5:00 P.M. half hour
Hour	This attribute contains the specific hour of the day. This is represented by the beginning minute of the 60-minute range. Note that these could be stored in 24-hour, or military, format (e.g., 14:00) if this is meaningful and useful to the business community.	1:00 A.M. hour; 2:00 P.M. hour
AM/PM	This is a quick representation of the notation for the minutes between midnight and noon, and noon to midnight.	AM, PM
Day Segment	There are divisions of time that are not cleanly aligned with the clock notations of AM and PM or the employee shifts. These are defined by the patterns when customers call our hotline.	Before work (6:00 A.M. to 8:00 A.M.); Morning (8:00 A.M. to 11:30 A.M.); Lunchtime (11:30 A.M. to 1:30 P.M.); Overnight (11:00 P.M. to 6:00 A.M.),

Customer Dimension

The Customer dimension provides information about the individuals who place calls to the company. While some attributes are known through doing business with this customer, the organization may also purchase external demographic data. Figure 7-6 shows the Call Center Customer dimension.

The data definitions for the Call Center Customer dimension are included in Table 7-4.

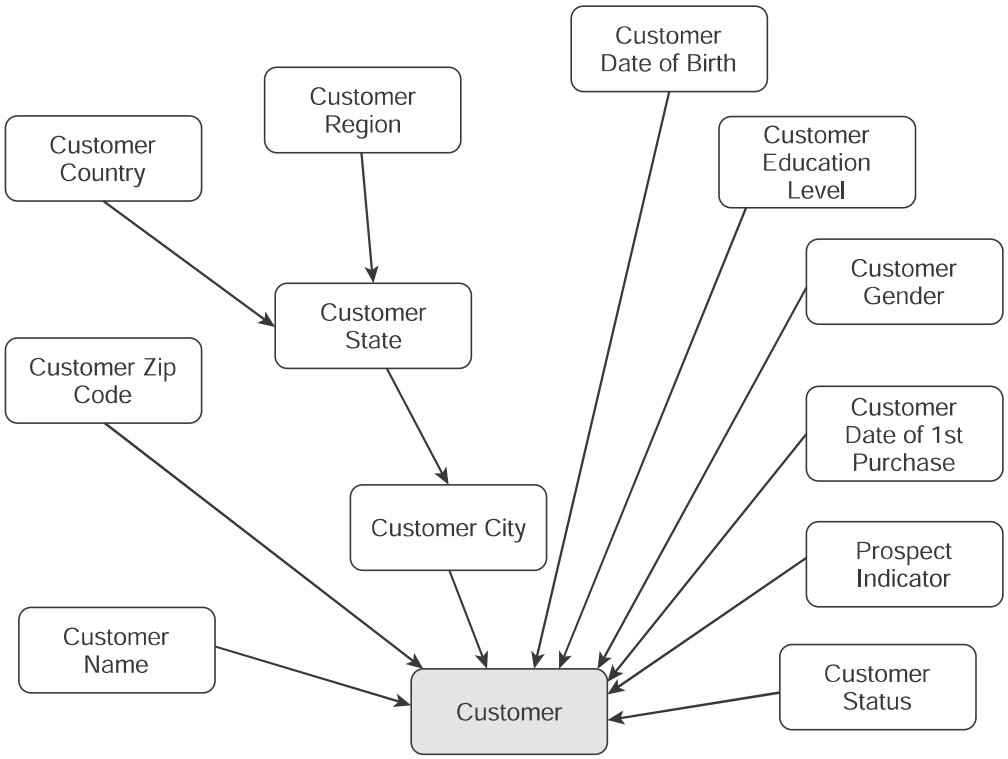

Figure 7-6 Call Center Customer dimension

Table 7-4 Call Center Customer Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Customer	A unique identifier for the individual consumer	100, 101
Customer Name	Name of the customer	Kate Twaddle, Sarah Mogensen
Customer Zip Code	Zip code for the primary residence for this customer	48801, 60663
Customer City	Name of the city of the primary residence for this customer	Harbor Beach, Peoria
Customer State	Name of the state of the primary residence for this customer	Illinois, Texas
Customer Country	Name of the country of the primary residence for this customer	US
Customer Region	Sales region to which the customer belongs	East, Central, Southeast

(*continued*)

Table 7-4 (*continued*)

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Customer Date of Birth	Date the customer was born. This is used to determine the customer's age.	04/24/1947, 08/29/1972
Customer Education Level	Highest level of education that this customer completed	High School, Associates, Bachelors, Masters
Customer Gender	Gender of the customer	Male, Female
Customer Date of First Purchase	Date when the customer made his or her first product purchase	03/14/1998, 07/03/2004
Prospect Indicator	Denotes whether this person has even been a customer, or whether this person is known through a marketing campaign or has requested information	Customer, Prospect
Customer Status	Indicates whether this is customer is still an active customer. The customer status is set to not applicable if this is a prospect.	Active, Inactive, Not Applicable

Employee Dimension

Managing a call center is highly dependent upon the customer service representatives, the people who answer the phones. There is a high turnover rate for call center employees. Productivity of these employees is critical. Figure 7-7 shows the Cell Center Employee dimension.

The data definitions for the Call Center Employee dimension are included in Table 7-5.

Call Dimension

Each specific call must be tracked by the call dimension. There is very little that is known about the call itself that is not already included in other dimensions. This is what is used to link the participation of multiple different employees for a variety of activities. This can happen if a call is transferred. The data warehousing term for this is a *degenerate dimension*. Figure 7-8 shows the Call dimension.

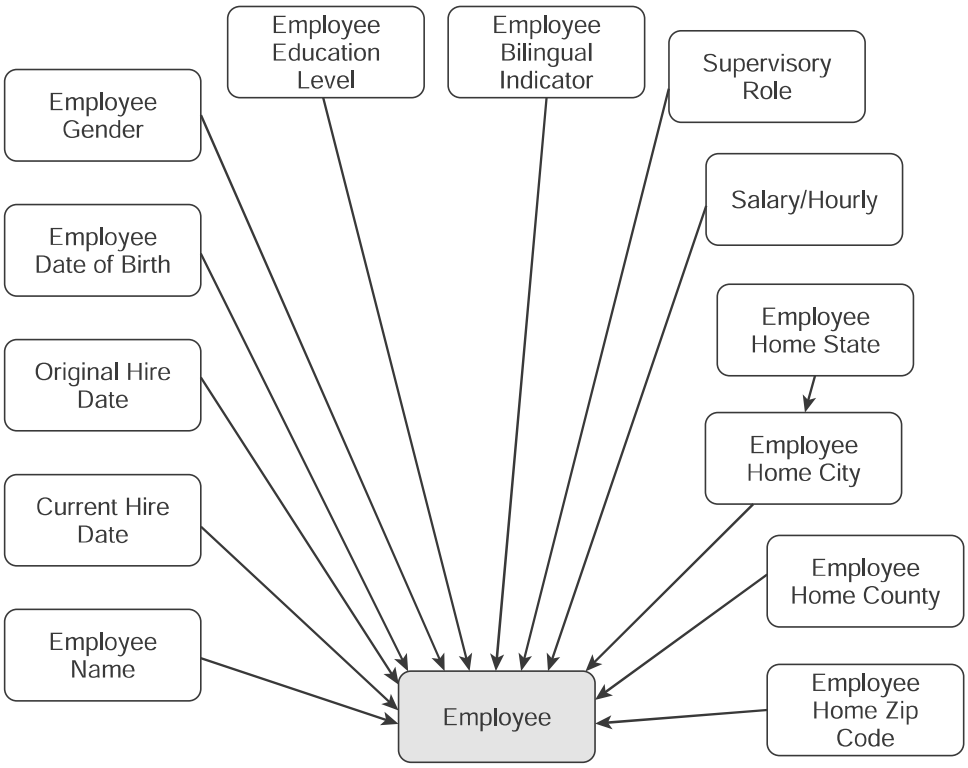

Figure 7-7 Call Center Employee dimension

Table 7-5 Call Center Employee Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Employee	Uniquely identifies each employee	2000, 3874
Employee Name	Name of the employee	Debbie Smith, Paul Jones
Current Hire Date	Most recent date that this person was hired.	09/01/2005
Original Hire Date	Date that this person was hired for the first time	09/01/2005
Employee Date of Birth	Employee's birth date	12/17/1975
Employee Gender	Gender of the employee	Male, Female
Employee Education Level	Indicates the highest level of education that this employee has completed	High School, Associates, Bachelors, Masters

(*continued*)

Table 7-5 (*continued*)

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Bilingual Employee Indicator	Indicates whether the employee speaks more than one language fluently	Bilingual, Not Bilingual
Supervisory Role	Indicates whether this person has any managerial responsibilities. This includes having both budget responsibilities and personnel responsibilities.	Supervisor, Non-management
Salary/Hourly	Indicates whether the employee is paid based upon hourly or salaried basis	Salaried, Hourly
Employee Home State	Name of the state of the primary residence for this employee	Iowa, Florida
Employee Home City	Name of the city of the primary residence for this employee	San Francisco, Nashville
Employee Home County	Name of the county of the primary residence for this employee	Cook, Kane
Employee Home Zip Code	The five-digit postal code of the primary residence for this employee	40441, 48060

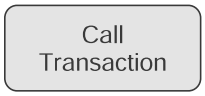

Figure 7-8 Call Center Call dimension

The data definitions for the Call Center Call dimension are included in Table 7-6.

Table 7-6 Call Center Call Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Call Transaction	This is the unique identifier that is assigned to a call when it comes into the call center. This is tracked until the call is completed.	123, 446

Call Outcome Dimension

When handling incoming calls, the purpose and type of call are tracked. The sales, finance, and call center groups each categorize these calls differently. Each of these groupings is shown in Figure 7-9.

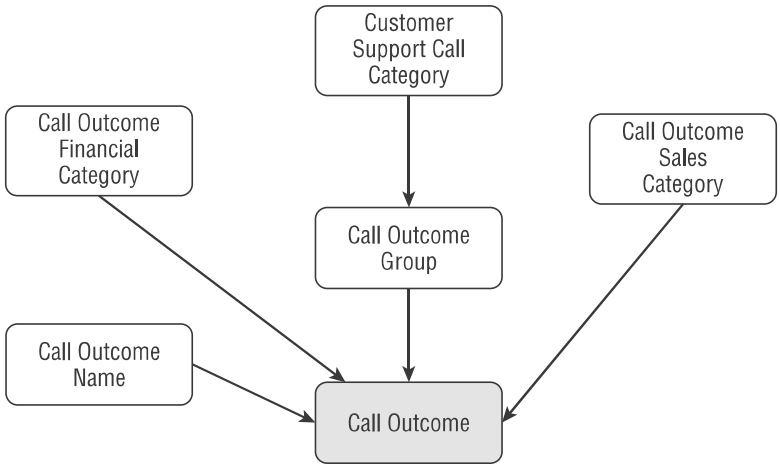

Figure 7-9 Call Center Call Outcome dimension

The data definitions for the Call Center Call Outcome dimension are included in Table 7-7.

Table 7-7 Call Center Call Outcome Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Call Outcome	A unique identifier for each possible result of a phone call	3900, 3901
Call Outcome Name	The label for each of the call outcomes, which describes the purpose of the call	Logged Complaint, Customer Question, Answered Prospect's Question, Placed Order
Call Outcome Financial Category	A grouping of call outcomes used by the financial group	Revenue Generation, Sales Support, Customer Service
Call Outcome Group	A collection of detailed call outcomes	Customer Complaints, Order Related, Prospect Activity
Customer Support Call Category	A grouping of call outcomes used by the call center management	Order Handling, Information Distribution, Problem Handling
Call Outcome Sales Category	A grouping of call outcomes used by the sales department	Customer, Prospects

Employee Task Dimension

Employee time tracking is based upon whatever activity or task that employee was working on. This is needed to be able to help understand the dynamic between employee activity and call productivity. Figure 7-10 shows the Employee Task dimension.

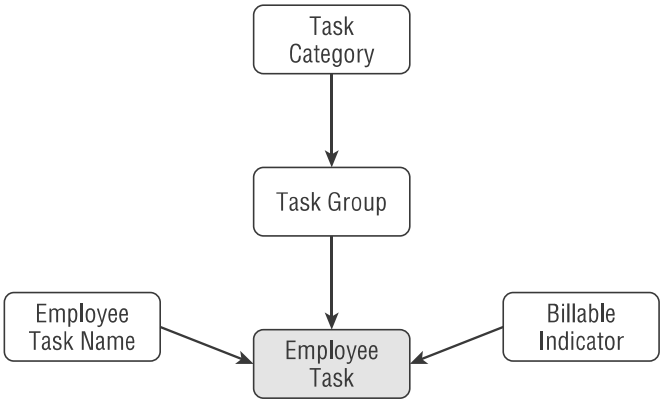

Figure 7-10 Call Center Employee Task dimension

The data definitions for the Call Center Employee Task dimension are included in Table 7-8.

Table 7-8 Call Center Employee Task Dimension Attribute Definitions

ATTRIBUTE NAME	DEFINITION	SAMPLE VALUES
Employee Task	A unique identifier for each specific customer task	500, 600
Employee Task Name	The name of the unique task indicating how an employee's time is spent	Handling Phone Orders, Logging Customer Complaints, Vacation, Training, Call Follow-Up
Billable Indicator	Identifies employee tasks that can be charged to the customer	Billable, Non-Billable
Task Group	A grouping of employee tasks for management reporting	Internal Administration, Problem Handling, Customer Sales and Marketing
Task Category	A high-level grouping of employee tasks for management reporting	Customer, Internal

Call Center Fact Groups

Now that each of the dimensions is defined, it is time to look at the different facts that are needed. These are documented next.

Calls Fact Group

Some of the most critical facts that need to be tracked for a call center are those related to the handling of incoming calls. The Calls fact group is shown in Figure 7-11. Note that not every dimension defined earlier is included in this diagram, only those that are relevant to the calls facts. The minute that the call began for this employee is recorded for the Time dimension.

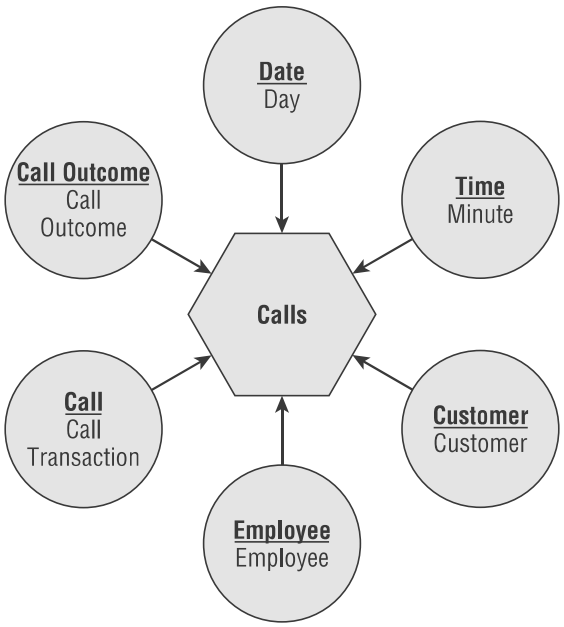

Figure 7-11 Call Center Calls fact group

The specific facts that are included in this fact group are described in Table 7-9.

NOTE **Having the Number of Calls set to one for each row allows a SUM function, rather than having to COUNT instances. The SUM function is much faster than the COUNT function.**

Call Center Time Tracking Fact Group

The ability to understand how employees are spending their time is critical for planning resources. It is also important to be able to track patterns in performance compared to how time is spent. Perhaps more frequent, but shorter breaks make customer service representatives more productive. This is the type of information that needs to be learned by studying the data. To that

end, Figure 7-12 shows the Time Tracking fact group. These facts have some of the same dimensions as the Calls fact group, but the Task dimension is new. In addition, it is important to note that the employee time is not tracked at a minute level of detail but in quarter-hour increments. The facts themselves are defined in Table 7-10.

Table 7-9 Call Center Calls Fact Definitions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Call Minutes	Total number of minutes spent on the phone with a customer	SUM
Wait Minutes	Number of minutes that the customer was put on hold before a customer service representative spoke to them.	SUM
Number of Call Transfers	Number of times that this call was transferred to another customer service representative. The first time a call is being handled, this is set to zero. Each subsequent transfer has this set to one. The Total Number of Transfers can be determined by summing the Number of Call Transfers grouped by the Call Transaction.	SUM
Number of Calls	Number of calls that were handled. This is set to one for each row at the lowest level.	SUM

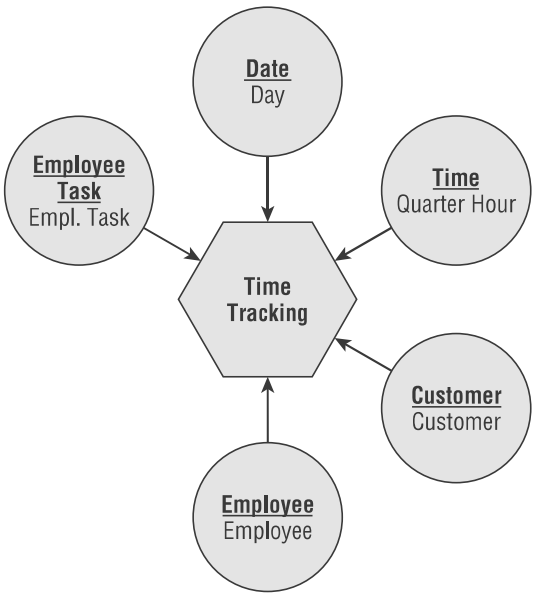

Figure 7-12 Call Center Time Tracking fact group

Table 7-10 Call Center Time Tracking Fact Definitions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Regular Hours Worked	Number of hours worked at this specific task at a regular or salaried pay rate. Time is tracked in quarter-hour increments.	SUM
Overtime Hours Worked	Number of hours worked at this task at an overtime pay rate	SUM

Call Forecast Fact Group

In order to be able to schedule employees effectively, it is helpful to have an estimate of the expected call volume. The forecast is used to develop employee work schedules. Development of these work schedules is operational in nature. The call volume forecast is useful in the data warehouse to compare with actual call volumes. This enables fine-tuning of the forecasting process. Figure 7-13 shows the Call Forecast fact group. Note that the Time dimension is at the hour level of detail. Call volume is not forecast at the minute level of detail. Table 7-11 has the definitions for the Call Forecast facts.

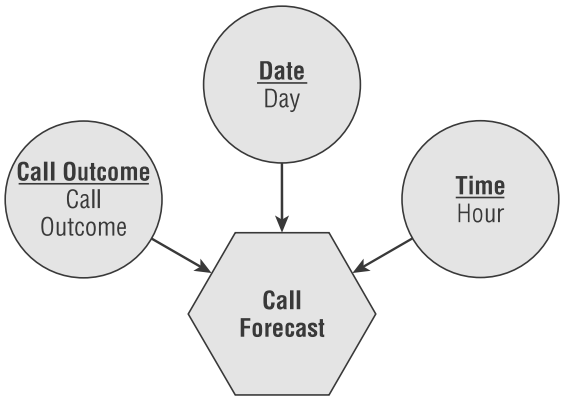

Figure 7-13 Call Center Call Forecast fact group

Table 7-11 Call Center Call FORECAST Fact Descriptions

FACT NAME	FACT DEFINITION	AGGREGATION RULE
Forecast Call Minutes	Expected number of minutes to be spent on the phone with customers	SUM
Forecast Number of Calls	Number of calls expected during this hour	SUM

Working with the Model

Business analysts should work with the diagrams to explore how the model works. This helps ensure that the desired reports and analyses can be run. The following section walks through how to do this using the Call Center business dimensional model. Figure 7-14 shows a calendar monthly report of the number of calls by region. The calendar month perspective aligns with how the call center interfaces with the customers, and the number of calls fact is from the Calls fact group (refer to Figure 7-11).

CENTRAL REGION
# of Calls

REGION	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
EAST	16,124	13,923	10,736	11,282	9,102	15,682	8,624	10,646	11,281	10,851	21,448	25,448
CENTRAL	12,459	10,759	8,296	8,718	7,034	12,118	6,664	8,226	8,717	8,385	16,573	19,664
SOUTHEAST	5,863	5,063	3,904	4,103	3,310	5,703	3,136	3,871	4,102	3,946	7,799	9,254
MOUNTAIN	28,583	24,682	19,032	20,000	16,136	27,800	15,288	18,872	19,998	19,236	38,021	45,112
WEST COAST	10,261	8,860	6,832	7,180	5,792	9,980	5,488	6,775	7,179	6,905	13,648	16,194
Total	**73,291**	**63,288**	**48,799**	**51,283**	**41,374**	**71,283**	**39,201**	**48,390**	**51,278**	**49,322**	**97,489**	**115,672**

Figure 7-14 Sample Call Center Report

The report can be changed easily for an internal look by swapping out the Calendar Month for the Fiscal Month on the report. Both of these attributes are in the Call Center Date dimension (refer to Figure 7-3).

More detail can be pulled by drilling down on the customer regions to get a customer state. By looking at the Customer dimension in Figure 7-6, you can see that regions drill down into states. When you use the drill-down feature of your business intelligence tool, you will see this same report by Customer State.

These first two changes show how you can change a report by using other attributes within the dimensions that are already on the report. Reporting is not limited to these dimensions. Next, you can change the report to see the number of calls (the same fact) by employee education level, which is in the Employee dimension shown in Figure 7-7.

You can also change the fact on the report too. Using another fact from the same fact group, you can replace the number of calls with the call minutes fact from the Calls fact group (refer to Figure 7-11).

Finally, you can change the fact to one that is from a different fact group entirely. This is where you need to ensure that you have consistency between the dimensions that are on the report. In this case, you can change the fact to Regular Hours Worked from the Employee Time Tracking fact group shown in Figure 7-12.

This should help crystallize how dimensional models work. When the model is implemented, the business intelligence tools provide functionality to drill down and modify reports as shown in this example. This should also help you understand that the tool can only perform these functions if the data is organized in a manner to support it.

Business Dimensional Model Index

A business dimensional model can include many different dimensions and fact groups. In order to help understand the contents of the model, a summary matrix can be created. Table 7-12 shows the index for the Call Center example. The columns represent each of the dimensions. Each row of the index represents a fact group. The body of the index shows where the fact group uses that dimension.

Table 7-12 Call Center Business Dimensional Model Index

FACT GROUP NAME	DATE	TIME	CUSTOMER	EMPLOYEE	CALL	CALL OUTCOME	EMPLOYEE TASK
Calls	X	X	X	X	X	X	
Time Tracking	X			X			X
Call Forecast	X	X	X				

Enterprise Considerations

There are often requirements to combine data from different parts of the organization. Dimensional models, when designed and implemented properly, can provide the capability to integrate data at query time. Often, the requirement to integrate data may not have been expressed during requirements gathering. This may be due to limitations of the current environment, so this is not even considered to be possible. This can also happen as the data warehouse grows over time. New data sources can be added that did not even exist when the original data warehouse was built. This section covers two fundamental concepts, conformed dimensions and conformed facts, which are essential to ensuring that all of the parts of the data warehouse work together.

Conformed Dimensions

A *conformed dimension* is one that is shared between two or more fact tables. This enables integration between different fact tables at query time. This is a foundational principle that enables the longevity of a data warehousing environment. By using conformed dimensions, facts can be used together,

aligned along these common dimensions. The beauty of using conformed dimensions is that facts that were designed independently of each other, perhaps over a number of years, can be integrated using these conformed dimensions. An example of the conformed dimension concept can be seen through the Customer dimension. If a customer is assigned an identifier of 1000, then any reference to that customer is made using the identifier of 1000. This enables data about that customer from across the enterprise to be combined whenever it occurs, even if it was not a specific requirement for a data mart. This supports the ability to integrate data from any fact groups along their common dimensions at query time.

The biggest challenge in designing a conformed dimension is that it requires cooperation between different business groups. Too often there is a feeling that this will force everyone in the organization to a single rollup of the data and that only common characteristics can be included. This is not the case.

For example, consider that there are four different data elements to describe Customer Type. Analysis must be done to determine whether there is a legitimate business reason for each. Over time, implementation and enforcement of standards can become lax. Variations in the data can creep in. If three of the Customer Type data elements are truly supposed to represent the same thing, then the data should be cleaned up to reflect this; but if the fourth instance of Customer Type really is different, then this can be included. Each Customer Type must be labeled so that you can tell them apart. Table 7-13 shows sample values from these four data elements.

Table 7-13 Customer Type Sample Values

CUSTOMER TYPE A	CUSTOMER TYPE B	CUSTOMER TYPE C	CUSTOMER TYPE D
Small Retail	SM RTL	Retail	Direct Bill
National Chain	LG CHN	L Chain	Payment on Delivery
Regional Chain	SM CHN	R Chain	Regular Invoice
Single Store	STORE	Single	Electronic Invoice

The first three may be melded into an attribute called Customer Type. The fourth should be renamed to Customer Payment Type.

If all of the attributes in a dimension are conformed, this by default ensures that all of the hierarchies are the same. However, if there are legitimate business reasons to represent different hierarchies, then each can be accommodated. All the attributes within each hierarchy should be uniquely named and clearly defined. This uniqueness must exist between all of the hierarchies, not just within each one. This can often be achieved by simply adding a qualifying word

such as ''Sales'' or ''Finance'' to each attribute within its respective hierarchy. This enables different groups to look at the business in their own way.

In addition to providing integration capabilities, another benefit of conformed dimensions is the sharing of attributes across the organization. Using the preceding example, the Customer Payment Type may have only been available to the billing group in the past. This may be an interesting way for the sales and marketing teams to look at Customer. This sharing can yield fresh insight.

Conformed dimensions should be created once and then shared as appropriate across the enterprise. This leverages development and maintenance efforts and helps ensure consistency.

Conformed Facts

It is also necessary to ensure consistency of facts. A single fact should be defined once and remain consistent if it is used anywhere else in the enterprise. These are *conformed facts*. Again, if there are variances in meaning, then these are not the same fact and each should be uniquely named.

There are cases where the fact is intended to be the same but variations have crept in over time. This provides the opportunity to tighten up the data. Each fact should have a single definition.

TIP When possible, facts should be built and loaded once, and then shared with any other groups that need to access that fact. If, for performance reasons, the fact is loaded into more than one database, the definition and rules to create the fact must be strictly enforced to ensure consistency. The result should be the same regardless of which fact group, across the enterprise, is used.

Practical Guidelines

There is no magic number for the right number of dimensions and fact groups you should have. The number of dimensions and fact groups will be determined by the industry you are in, the data that the organization has, and the specific characteristics of your organization. However, several general guidelines may be helpful.

Guidelines for a Single Dimension

The more attributes that you have in a dimension, the richer the possible analyses. Whenever possible, seek out additional attributes. For example, the claims handling system may contain only a limited set of details about the policyholder. In order to create a complete and robust customer dimension, it would