

The *fact table* in a dimensional model stores the performance measurements resulting

from an organization’s business process events.

Fact Tables for Measurements

The *fact table* in a dimensional model stores the performance measurements resulting

from an organization’s business process events. You should strive to store the

low-level measurement data resulting from a business process in a single dimensional

model. Because measurement data is overwhelmingly the largest set of data,

it should not be replicated in multiple places for multiple organizational functions

around the enterprise. Allowing business users from multiple organizations to access

a single centralized repository for each set of measurement data ensures the use of

consistent data throughout the enterprise.

The term *fact* represents a business measure. Imagine standing in the marketplace

watching products being sold and writing down the unit quantity and dollar sales

amount for each product in each sales transaction. These measurements are captured

as products are scanned at the register, as illustrated in Figure 1-2.

Each row in a fact table corresponds to a measurement event. The data on each

row is at a specifi c level of detail, referred to as the *grain*, such as one row per product

Data Warehousing, Business Intelligence, and Dimensional Modeling Primer **11**

sold on a sales transaction. One of the core tenets of dimensional modeling is that

all the measurement rows in a fact table must be at the same grain. Having the discipline

to create fact tables with a single level of detail ensures that measurements

aren’t inappropriately double-counted.

The most useful facts are numeric and additive, such as dollar sales amount.

Throughout this book we will use dollars as the standard currency to make the

case study examples more tangible—you can substitute your own local currency

if it isn’t dollars

Facts are often described as continuously valued to help sort out what is a fact

versus a dimension attribute. The dollar sales amount fact is continuously valued in

this example because it can take on virtually any value within a broad range. As an

**12 Chapter 1**

observer, you must stand out in the marketplace and wait for the measurement before

you have any idea what the value will be.

As examples are developed throughout this book, you will see that all fact table

grains fall into one of three categories: transaction, periodic snapshot, and accumulating

snapshot. Transaction grain fact tables are the most common. We will

introduce transaction fact tables in Chapter 3: Retail Sales, and both periodic and

accumulating snapshots in Chapter 4: Inventory.

All fact tables have two or more foreign keys (refer to the FK notation in Figure 1-2)

that connect to the dimension tables’ primary keys. For example, the product key in

the fact table always matches a specifi c product key in the product dimension table.

When all the keys in the fact table correctly match their respective primary keys in

the corresponding dimension tables, the tables satisfy *referential integrity*. You access

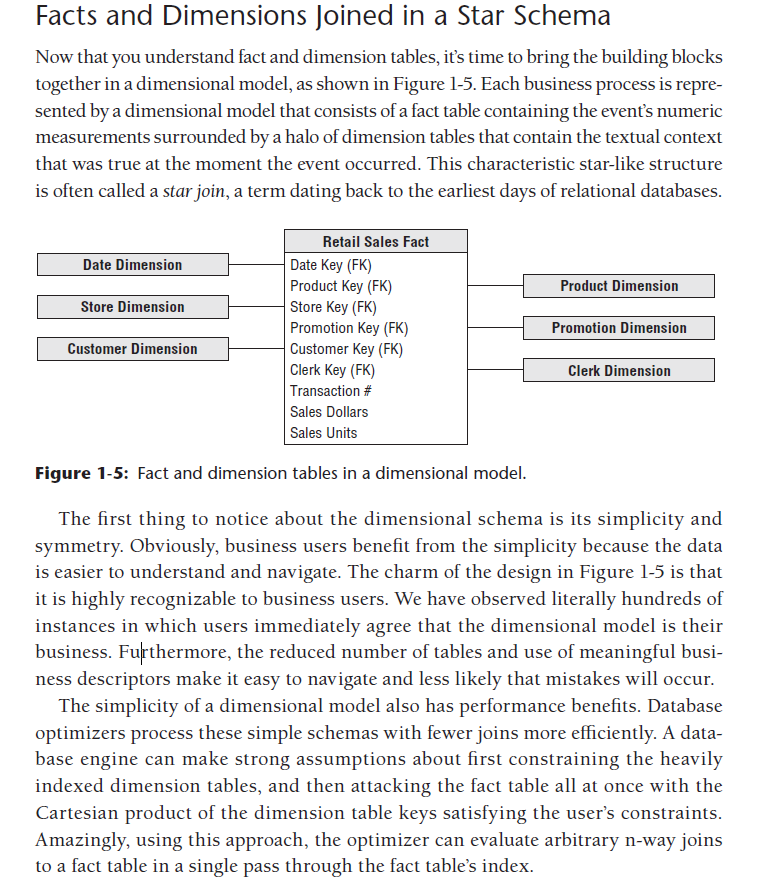
the fact table via the dimension tables joined to it.

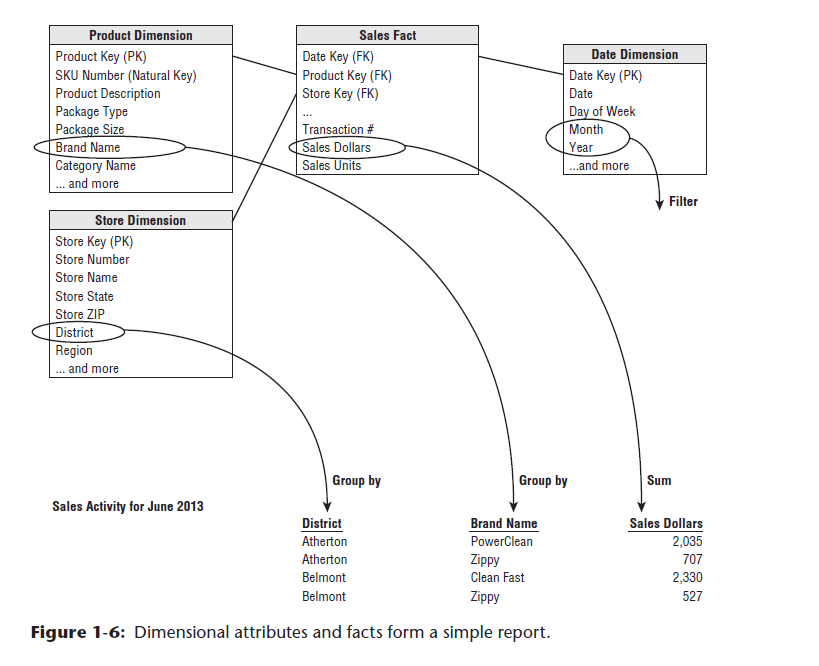
Dimension Tables for Descriptive Context

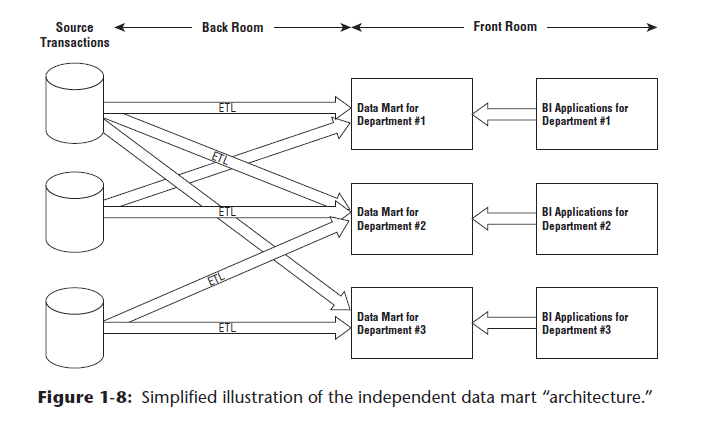
*Dimension tables* are integral companions to a fact table. The dimension tables contain

the textual context associated with a business process measurement event. They

describe the “who, what, where, when, how, and why” associated with the event.







Step 1: Select the Business Process

A *business process* is a low-level activity performed by an organization, such as taking

orders, invoicing, receiving payments, handling service calls, registering students,

performing a medical procedure, or processing claims.

Step 2: Declare the Grain

Declaring the *grain* means specifying exactly what an individual fact table row

represents. The grain conveys the level of detail associated with the fact table

measurements. It provides the answer to the question, “How do you describe a

single row in the fact table?” The grain is determined by the physical realities of

the operational system that captures the business process’s events.

■ One row per scan of an individual product on a customer’s sales transaction

■ One row per line item on a bill from a doctor

■ One row per individual boarding pass scanned at an airport gate

■ One row per daily snapshot of the inventory levels for each item in a warehouse

■ One row per bank account each month

Step 3: Identify the Dimensions

Dimensions fall out of the question, “How do business people describe the data

resulting from the business process measurement events?” You need to decorate

fact tables with a robust set of dimensions representing all possible descriptions

that take on single values in the context of each measurement.

Step 4: Identify the Facts

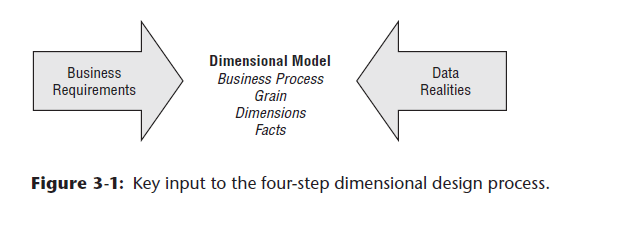
Facts are determined by answering the question, “What is the process measuring?”

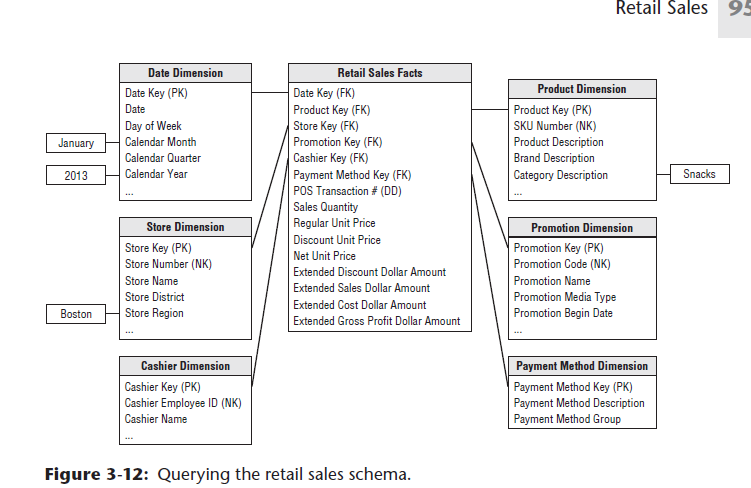
Business users are keenly interested in analyzing these performance metrics. All

candidate facts in a design must be true to the grain defi ned in step 2. Facts that

clearly belong to a diff erent grain must be in a separate fact table. Typical facts are

numeric additive fi gures, such as quantity ordered or dollar cost amount.





The unique primary key of a dimension table should be a *surrogate key* rather than

relying on the operational system identifi er, known as the *natural key*. Surrogate keys

go by many other aliases: meaningless keys, integer keys, non-natural keys, artifi -

cial keys, and synthetic keys. Surrogate keys are simply integers that are assigned

sequentially as needed to populate a dimension. The fi rst product row is assigned a

product surrogate key with the value of 1; the next product row is assigned product

key 2; and so forth. The actual surrogate key value has no business signifi cance. The

surrogate keys merely serve to join the dimension tables to the fact table. Throughout

this book, column names with a Key suffi x, identifi ed as a primary key (PK) or

foreign key (FK), imply a surrogate.

Step 1: Select the Business Process

In our retail case study, management wants to better understand customer purchases

as captured by the POS system. Thus the business process you’re modeling

is POS retail sales transactions. This data enables the business users to analyze

which products are selling in which stores on which days under what promotional

conditions in which transactions.

Step 2: Declare the Grain

A DW/BI system almost always demands data expressed at the lowest

possible grain, not because queries want to see individual rows but because queries

need to cut through the details in very precise ways.

You should develop dimensional models representing the most detailed,

atomic information captured by a business process.

Step 3: Identify the Dimensions

After the grain of the fact table has been chosen, the choice of dimensions is straightforward.

The product and transaction fall out immediately. Within the framework

of the primary dimensions, you can ask whether other dimensions can be attributed

to the POS measurements, such as the date of the sale, the store where the sale

occurred, the promotion under which the product is sold, the cashier who handled

the sale, and potentially the method of payment. We express this as another design

principle.

A careful grain statement determines the primary dimensionality of the

fact table. You then add more dimensions to the fact table if these additional dimensions

naturally take on only one value under each combination of the primary

dimensions. If the additional dimension violates the grain by causing additional

fact rows to be generated, the dimension needs to be disqualifi ed or the grain statement

needs to be revisited

Step 4: Identify the Facts

The fourth and fi nal step in the design is to make a careful determination of which

facts will appear in the fact table. Again, the grain declaration helps anchor your

thinking. Simply put, the facts must be true to the grain: the individual product

line item on the POS transaction in this case. When considering potential facts,

you may again discover adjustments need to be made to either your earlier grain

assumptions or choice of dimensions.