

Realtime  
publishers

# *The Shortcut Guide<sup>™</sup> To*



# Achieving Business Intelligence in Midsize Companies

*sponsored by*



*Don Jones*

---

# Introduction to Realtime Publishers

---

by Don Jones, Series Editor

For several years now, Realtime has produced dozens and dozens of high-quality books that just happen to be delivered in electronic format—at no cost to you, the reader. We’ve made this unique publishing model work through the generous support and cooperation of our sponsors, who agree to bear each book’s production expenses for the benefit of our readers.

Although we’ve always offered our publications to you for free, don’t think for a moment that quality is anything less than our top priority. My job is to make sure that our books are as good as—and in most cases better than—any printed book that would cost you \$40 or more. Our electronic publishing model offers several advantages over printed books: You receive chapters literally as fast as our authors produce them (hence the “realtime” aspect of our model), and we can update chapters to reflect the latest changes in technology.

I want to point out that our books are by no means paid advertisements or white papers. We’re an independent publishing company, and an important aspect of my job is to make sure that our authors are free to voice their expertise and opinions without reservation or restriction. We maintain complete editorial control of our publications, and I’m proud that we’ve produced so many quality books over the past years.

I want to extend an invitation to visit us at <http://nexus.realtimepublishers.com>, especially if you’ve received this publication from a friend or colleague. We have a wide variety of additional books on a range of topics, and you’re sure to find something that’s of interest to you—and it won’t cost you a thing. We hope you’ll continue to come to Realtime for your educational needs far into the future.

Until then, enjoy.

Don Jones

---

Introduction to Realtime Publishers.....	i
Chapter 1: What Is Business Intelligence, Anyway?.....	1
Business Goals for BI.....	3
BI: Not Just for Big Companies.....	3
BI Terminology.....	4
Data Warehouses and Data Marts.....	4
Data Mining.....	6
Reporting, Dashboards, and Scorecards.....	6
Predictive Analytics.....	9
Business Performance Management.....	9
A Brief Technical Overview of BI.....	10
Data Marts and Data Warehouses.....	10
OLTP vs. OLAP.....	10
Normalized vs. Dimensional Data.....	11
Data Modeling.....	12
Data Transformation and Loading.....	13
In-Memory Analytics.....	13
Reporting, Analysis, and Visualization.....	13
BI: Analysis + Planning.....	16
Analyzing and Reporting on Information.....	16
Making Decisions.....	17
Implementing.....	17
One Size Does Not Fit All.....	17
BI in Large Enterprises.....	18
BI in Midsize Companies.....	18
Coming Up Next.....	19

## Copyright Statement

© 2009 Realtime Publishers, Inc. All rights reserved. This site contains materials that have been created, developed, or commissioned by, and published with the permission of, Realtime Publishers, Inc. (the "Materials") and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of Realtime Publishers, Inc or its web site sponsors. In no event shall Realtime Publishers, Inc. or its web site sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

The Materials (including but not limited to the text, images, audio, and/or video) may not be copied, reproduced, republished, uploaded, posted, transmitted, or distributed in any way, in whole or in part, except that one copy may be downloaded for your personal, non-commercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Realtime Publishers and the Realtime Publishers logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners.

If you have any questions about these terms, or if you would like information about licensing materials from Realtime Publishers, please contact us via e-mail at [info@realtimepublishers.com](mailto:info@realtimepublishers.com).

[**Editor's Note:** This eBook was downloaded from Realtime Nexus—The Digital Library for IT Professionals. All leading technology eBooks and guides from Realtime Publishers can be found at <http://nexus.realtimepublishers.com>.]

# Chapter 1: What Is Business Intelligence, Anyway?

---

For a few years now, *business intelligence* have been buzzwords of the IT industry. Unlike a lot of IT buzzwords—like “Web 2.0”—business intelligence actually carries some weight, has a definite meaning, and brings real value to businesses. Business intelligence—or BI, as insiders like to call it—refers to the practice of helping a business acquire a better understanding of itself. More broadly, BI also refers to the skills, technologies, applications, and practices involved in bringing that understanding to light.

The term *business intelligence* was first used in a 1958 article by IBM researcher Hans Peter Luhn, who defined it as *the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal*. In other words, BI isn't just about technology-centric concepts such as data warehousing or business analytics; BI is really about understanding the relationships between different aspects of your company so that you can guide the company toward specific goals, like increasing market share and improving customer satisfaction.

The IT industry—being technology-centric, of course—uses the term *BI* as a sort of umbrella phrase that covers all the technologies and capabilities used to gather facts about the business, present those facts in a way that makes relationships clearer, and allow manipulation of those facts to project “what if” scenarios—all intended to help guide better decision making.

Let's consider a simple example of how BI can help a business.

### Improving Customer Satisfaction

Widgets, Inc. has been struggling to improve their customer service levels. The company generally gets good customer satisfaction scores on the surveys it conducts, but it goes through periods where satisfaction drops by more than 30%. The company's executives already know that the drops come during periods when the company is extremely busy, fielding far more orders than usual. The company has already spent tens of thousands of dollars improving their distribution center operations to reduce the time it takes to ship customer orders, but it hasn't seemed to make any difference.

The company invests in a BI system. The system gathers information from a number of internal sources, including the main order-processing database. The system also collects data from some external sources, including billing data from the company's shipping vendors, the company's payroll system, and other places.

After using the BI system for a few months, company managers notice something unexpected. During periods of decreased customer satisfaction, their distribution center is actually *less* busy—they can see the increase in customer orders correlate with a *decrease* in payroll for the distribution center. Digging a bit further, they realize that the increased customer orders are mainly for products that are being drop-shipped directly from a couple of specific vendors to customers. Management realizes that it's those *vendors* that are slow to fulfill orders, causing the drop in customer satisfaction. They now know to focus their efforts on improving those vendors' performance, finding new vendors for those products, or stocking those products in their own distribution center, where they'll have better control over shipping times.

The driving force behind BI is that companies are drowning in unrelated facts that come from *silos*: payroll data, financial data, customer data, vendor data, and so on. BI pulls all that data together and correlates it. The data may seem unrelated, but in fact *everything* in the business is related somehow—if some data is *truly* unrelated, then why is it in the business in the first place? BI doesn't generate new data—it simply makes it easier to explore overlooked relationships between data.

## Business Goals for BI

Most companies—even midsize ones—have an incredible amount of data living in transaction-based, distributed systems and databases. The payroll system has one database, the order-processing system has another, and so on. These databases are typically fine tuned for individual transactions, such as retrieving a single customer order, or for specific batch operations, such as processing payroll at the end of each month. What these databases are *not* designed to do is communicate with one another, to allow users to explore data in unusual ways, or to provide high-level summaries of the data in an instant. The main goal for BI, then, is to provide exactly those things:

- The ability to view data from multiple sources in a single view, such as seeing sales information correlated with shipping expenses and facility utility bills
- The ability to quickly see summaries of data from different places—such as the total payroll spent, along with the total sales, for a given time period
- The ability to see data over time, comparing data from yesterday, last month, the past three quarters of the past 5 years, to see how things have changed over time
- The ability to ask “what if” questions and have answers generated based on historical data; for example, you might want to know if raising sales by 10% for a sustained period of time will necessitate a raise in utility costs or payroll expenses

Ideally, the answers to these questions can feed directly into the company’s planning systems, helping define budgets, sales goals, and other planning elements. Doing so allows historical trends to drive business decisions, and those decisions automatically drive business planning.

One of the reasons BI has become so popular a term in the past few years is that it is an *incremental* IT investment. You don’t have to tear apart your existing systems in any way—in fact, those systems won’t know that they’re participating in a BI solution. The fact that BI doesn’t impact your production systems means implementing BI is relatively low risk—you’re not likely to disrupt daily operations while implementing a BI solution. Best yet, a properly-designed and -implemented BI solution can quickly deliver a high return on investment (ROI), something executives really appreciate.

## BI: Not Just for Big Companies

There’s a perception that BI is just for major enterprises, and it’s certainly true that major enterprises have a lot to gain from implementing BI and that those big companies have the extra cash needed to implement a BI solution. Implementation in major enterprises can take months, and it’s rarely an inexpensive undertaking.

But that doesn't mean midsize companies can't benefit from BI, and it doesn't mean midsize companies have to spend as much time or money as huge companies do. After all, midsize companies often use less-expensive, easier-to-implement solutions for things like payroll, corporate bookkeeping, and customer relationship management; BI doesn't need to be any different.

But there's an important corollary to that statement: Just as midsize businesses *don't* use the same bookkeeping software that a major enterprise does, a midsize business *won't* use the same BI solution that a gigantic company uses. Bringing a large-scale BI solution into a midsize company isn't any smarter than bringing a gigantic facilities-management package into a midsize company. What midsize companies need is a midsize BI package. Huge companies are generally accustomed to lengthy implementation times for *any* new solution package—whether it's Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), payroll, or whatever. Midsize companies typically take advantage of more prepackaged solutions that are less expensive and much easier to implement. The same approach can work for BI.

In fact, you can make an argument that midsize companies can see a quicker, better return on a BI investment than large companies. After all, large companies can be slow to change, even when they've got good intelligence, smart decisions, and a good plan. Even with a fully-implemented BI system, large companies often take longer to implement the decisions that their BI leads them to. Midsize companies, in contrast, are known for being somewhat simpler in structure and can often react more quickly to change. With the right BI solution, a midsize company could easily develop solid BI, make smart decisions about the company's future, and *act* on those decisions more rapidly—meaning a faster ROI from that BI solution.

## BI Terminology

Let's briefly step back from the business aspects of BI and look at some of the major elements that go into a BI solution. We'll start with the next five sections, where I'll outline the major aspects of a BI solution at a high level.

### Data Warehouses and Data Marts

Central to most BI efforts is a *data mart* or a *data warehouse*. They're both basically the same thing under the hood: a specialized kind of database that's designed to support business analytics and to contain data from one or many different sources. The difference between a mart and a warehouse is their scope: A *data mart* only seeks to serve the needs of a portion of the company, such as the marketing department or finance department. A *data warehouse* seeks to serve the entire company.



There are two design approaches that come into play when you're talking about marts and warehouses. The *top-down* approach directs us to build the data warehouse first, considering the needs of the entire company when designing it. That's obviously a complex task, as you really need to look at how the entire company works and how its various pieces fit together. With the warehouse in place, you can create data marts that extract a portion of the warehouse to meet the needs of specific audiences, such as finance or marketing.

The *bottom-up* approach directs the opposite, having us build data marts for each audience within the company, then combining those to form a companywide data warehouse. This is often easier to implement, simply because it's easier to wrap your brain around the needs of a specific department—such as marketing—than to grasp the companywide view of things. However, this approach is not without its detractors. In a 2000 article, “Data Mart Does Not Equal Data Warehouse,” author William Inmon offered this analogy: “You can catch all the minnows in the ocean and stack them together and they still do not make a whale.” What he's saying is that the silo-specific data marts will have gaps between them, and they'll never, in sum, provide the real companywide view you need.

These sorts of design issues are what make BI implementation in large companies so complex. In midsize companies, however, there's an advantage: Prepackaged solutions can take much of this controversy out of your hands by providing a pre-made data warehouse.

Ralph Kimball, another noted data warehouse design expert, favors what many describe as a bottom-up approach. He suggests that, first and foremost, you need to focus on your business. Building smaller data marts that each focus on a particular subject within your business helps constrain the task to something more manageable, and helps you keep the business—not the ultimate data warehouse—more firmly in mind. He recommends building data marts not around business units but around business processes, such as orders, shipments, payments, and so forth. Kimball refutes the use of the “bottom-up” label for his approach, pointing out that his approach doesn't follow the traditional bottom-up approach of designing around organizational units within the business but instead around how the business works and what it does.

### Resource

A full discussion of Ralph Kimball's approach is beyond the scope of this book, but you can learn more at <http://www.ralphkimball.com/>. Especially informative is his “Fables and Facts” page at <http://www.ralphkimball.com/html/controversies.html>.

So what *is* a data warehouse? I'll get into the technical details a bit later in this chapter, but for now, think of a data warehouse as a place where *all* your company's data is *copied to* and then rearranged so that relationships are easier to perceive and summaries are easier to generate.

## Data Mining

Data mining is simply the process of extracting patterns from data—in BI terms, from a data warehouse. Want to find out why the distribution center processes orders more slowly at certain times of the month? Mine your data warehouse looking for patterns—perhaps you'll find that things slow down as the supply of cardboard boxes dwindles because the box vendor isn't fulfilling *their* orders quickly enough.

It's important to realize, however, that data mining can't reveal patterns in data that aren't present in the data being mined. That sounds obvious, but it can be deceptive because data mining can often *seem* to reveal patterns that aren't really there. What you can find yourself looking at is a pattern comprised of *symptoms* rather than *causes*, when the causal data isn't present in the data warehouse. This isn't to say that data mining isn't useful—it's at the heart of BI, in fact. Rather, you just need to be aware of what's in the data, and what isn't, and take common-sense approaches to verifying and validating the conclusions to which your data mining leads you.

## Reporting, Dashboards, and Scorecards

With all the data that a data warehouse can contain, most users will need simplified means of looking at commonly-examined information. Reports are one obvious product of a data warehouse, and they can range from high-level summaries to extremely detailed analyses. Another option is a *dashboard*, which provides a summary of common metrics possibly from multiple sources, and often contains planning and actual comparisons, often visualized in a slick, simplified user interface (UI)—such as the example Figure 1.1 shows.

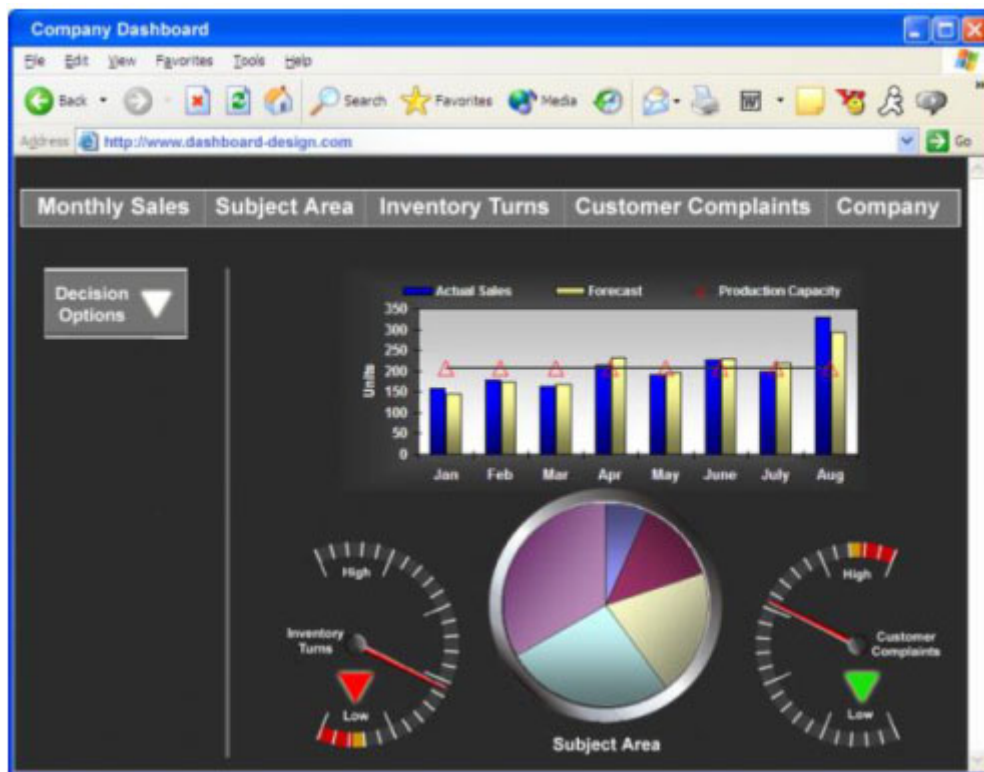


Figure 1.1: An example dashboard.

Dashboards don't often drive direct decisions; rather, they let individual users get a feel for general performance, such as sales, inventory turn, customer complaints, and so forth. Dashboards let you know if everything is all right or if further investigation is warranted. Dashboards may also include trends—such as day-by-day sales figures compared with goals or plans. Again, the dashboard isn't going to let you know *why* sales are where they are—but if they're significantly off-plan, you'll be able to tell at a glance and initiate further investigation.

Dashboards are useful because they can help new and less-experienced users quickly start taking advantage of the data in a data warehouse—or even the data in a transactional processing system, such as an order-entry application. The learning curve for a dashboard is usually pretty short and shallow, so more users are likely to use them more effectively.

Another useful visualization is a *scorecard*. This custom UI links internal and external data to the organization's goals. Basically, it tells you how far along you are in terms of achieving your goals. Figure 1.2 offers an example, showing how far various departments are in achieving a specific company goal.

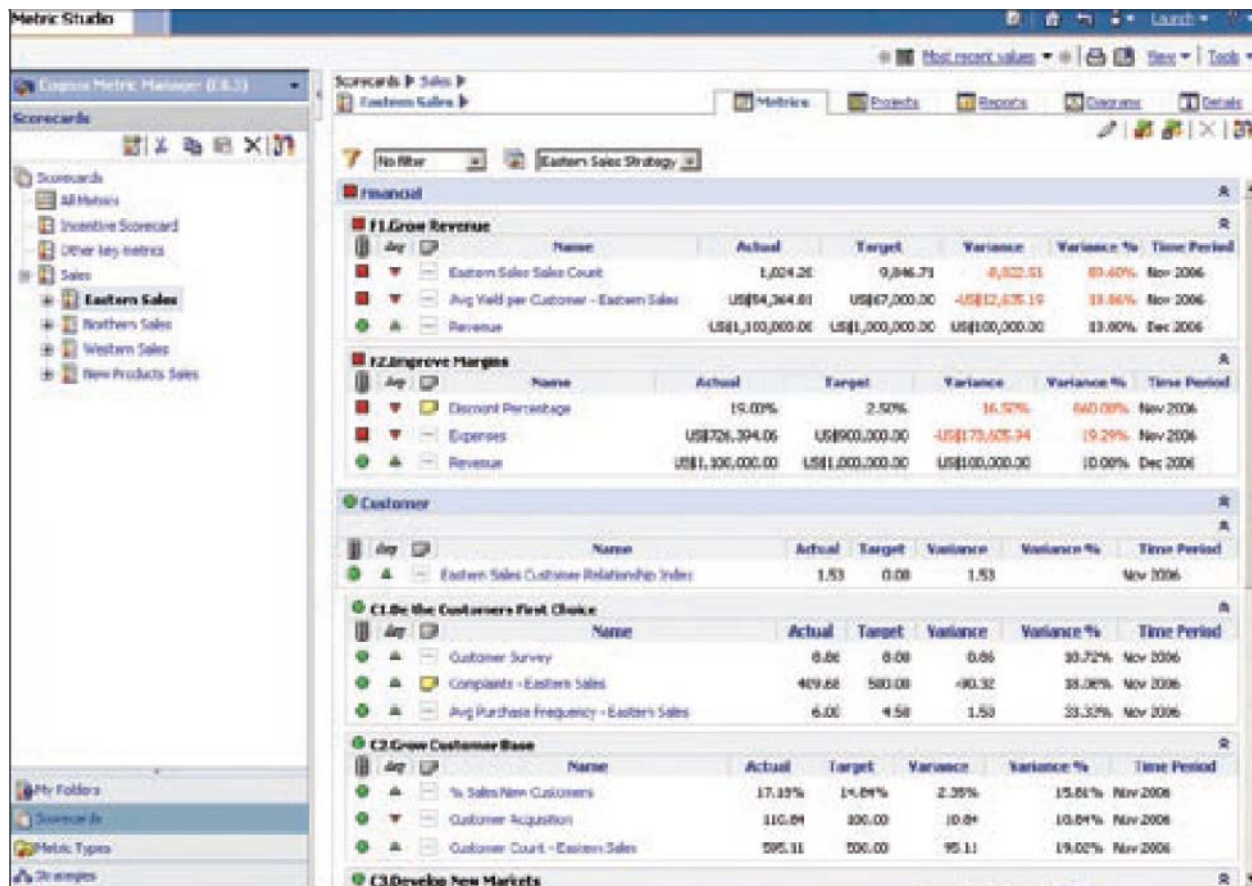


Figure 1.2: An example scorecard.

Scorecards are useful precisely because they show you where you are in relation to your goals. In this example, the Sales department is doing well; the Human Resources department may need some investigation and assistance. Scorecards help you focus your efforts, spot problem areas, and manage *to your goals* on a daily basis.

Dashboards and scorecards serve a similar purpose, and some BI solutions present them in such a way that there's no practical difference. If you need to make a distinction between them, a dashboard usually just shows you where you are in absolute terms; a scorecard shows you where you are relative to your goals. Both are useful, and both may be presented in a "dashboard-style" UI.

### **Empty Beds?!?!**

Consider this example of how dashboards can help management focus on the right areas for their attention. Nice Hotels, Inc. traditionally managed itself like most hotels do—their primary metric was room occupancy. 100% occupancy was always the goal, although most of their properties rarely achieved that on a consistent basis.

After implementing a BI system, managers were given a new dashboard that showed them total occupancy as well as occupancy on a *per-bed* basis. The dashboard also showed the number of guests turned away in a "no vacancy" situation. They found that their front desk agents were assigning single guests to rooms with double beds; later in the evening, they would turn away families simply because the only rooms left at that point contained single beds.

Managers started managing to *per-bed* occupancy instead of per-room occupancy. The scheduled room cleanings so that an equal number of doubles and singles would be cleaned at roughly the same time, ensuring that the front desk would have the best chance of having the "right-size" room for whatever guest was checking in. They trained the front desk to not use double rooms for single guests.

Over time, some properties were still seeing higher-than-desired quantities of empty beds each night, so they began planning to convert some double rooms to singles during refurbishments. Single rooms require less time to clean, are less expensive because they contain less furniture and soft goods such as sheets, and those properties clearly had a surplus of double rooms.

By being able to see per-bed occupancy at a glance, managers began to manage more aggressively and intelligently, customizing each property to its historical clientele and realizing additional revenue and savings over time.

## Predictive Analytics

Predictive analytics is exactly what it sounds like: mining data, looking for patterns, and making predictions about future events based on historical facts. Credit scoring is one of the most well-known forms of predictive analysis, and credit reporting companies make use of some of the world's largest data warehouses. The reporting companies' proprietary scoring algorithms consider factors such as credit history, payment history, loan applications, and other customer data to assign numeric scores that are a prediction of the customer's ability to properly manage and service their debt.

Cross-sell—such as the “you might also like” product suggestions from e-tailers like Amazon.com—is another example. By analyzing the total purchases made by all of its customers, a company can predict, with some degree of accuracy, what products you are likely to purchase based upon those you have purchased or those you are considering. By offering you those items rather than making you go hunt for them, the company can take advantage of impulse buying, helping to increase overall revenue.

Internally, many companies use predictive analytics without even thinking about it. For example, many retailers know that their holiday sales will be some multiple of their pre-holiday sales; they also know that a certain number of additional employees will be needed to handle the additional holiday sales volume. By looking at historical data and plotting a trend into the future, those companies can make an informed guess about each year's holiday hiring needs. Most companies face numerous ad-hoc decisions every day; predictive analytics can't *make* those decisions, but it can help *inform* them—with more accuracy than a “gut” decision.

## Business Performance Management

Business Performance Management (BPM), or Corporate Performance Management (CPM), rolls up reporting, predictive analytics, and other BI practices together with planning, budgeting, and forecasting. It's designed to provide a framework that organizes information, delivers new insight, takes action, and optimizes performance.

In “The Next Generation of Business Intelligence: Operational BI,” author Colin White described the link between BI and business performance management:

The biggest growth area in operational BI analysis is in the area of business performance management (BPM). Operational BPM applications not only analyze the performance... but also compare the measured performance against business goals and alert business users when actual performance is out of line with business goals.

As with scorecards, *goal* is the operative keyword for BPM. BI is great; comparing the information from BI to your goals allows you to *manage*.



Although the fancy name is new, BPM as a concept isn't all that new. Most sales managers, for example, are in the habit of looking at the prior days' sales when they sit down for work in the morning. They might drill into each salesperson's individual sales as well. They compare all that data with their sales goals, and typically know how much they still have left to sell in the month before they meet their goal. BPM and BI simply take that to a higher level, automating the collection of the data, automatically comparing it with goals, and not only displaying the result in a dashboard but also alerting management when performance is out of line with goals.

BPM might even drive automated decisions. For example, if sales are markedly higher than planned, a BPM system might accelerate the velocity of products orders placed with vendors to ensure sufficient stock is on hand to meet the growing demand trend.

BPM is a continual loop, meaning a BPM system updates itself. As the company responds to current events, the BPM system continuously collects its data and changes its analyses, helping the company immediately see the near-term and predicted long-term effects of its efforts.

## A Brief Technical Overview of BI

Now, let's spend some time looking at the technology that lives under the hood of a BI solution. Many of these technologies are really just clever extensions of proven, decades-old technologies used in new ways to help achieve greater results.

### Data Marts and Data Warehouses

A data warehouse—or a data mart subset—is really just a normal database. Typically, they live in the same relational database management systems—such as Oracle, SQL Server, DB2, or whatever—that “normal” databases live in. However, data warehouses and data marts are structured quite a bit differently.

### OLTP vs. OLAP

An Online Transaction Processing (OLTP) or transactional database—has several key features:

- It contains detailed information. For example, it contains not only summary information such as the total amount of an order but also detailed information such as how much each item costs.
- It is designed to process transactions—meaning it's typically dealing with one piece of data at a time: one order, one product, one customer. It may be used to generate basic reports from this data, but it's structure is optimized to support rapid access to small chunks of data, therefore containing many tables.

- The schema is rigid and unchanging.
- It contains up-to-date information and is updated in realtime. It may purge data once it is no longer needed or relevant, therefore containing only current data.
- The quality of input data is often very high, meaning applications and other elements ensure that correct data goes into the data warehouse.

A data warehouse is an Online Analytical Processing (OLAP) database, and often includes these features:

- Some data may be summarized, meaning detail is not available. You may be able to tell the total amount of a given order but not the cost of each product contained in the order.
- Its purpose is to drive analysis and decisions. Access is optimized for examining large quantities of data in order to see trends.
- The information is historical and is only as up to date as the loading schedule. The emphasis is on past data and trends, more so than immediate, realtime data. Historic data from months or years ago will often be stored within the database.
- Data is often “cleansed,” meaning errant or “edge” data may be removed to make trends clearer. Data warehouses tend to focus on sets of data rather than on individual elements.

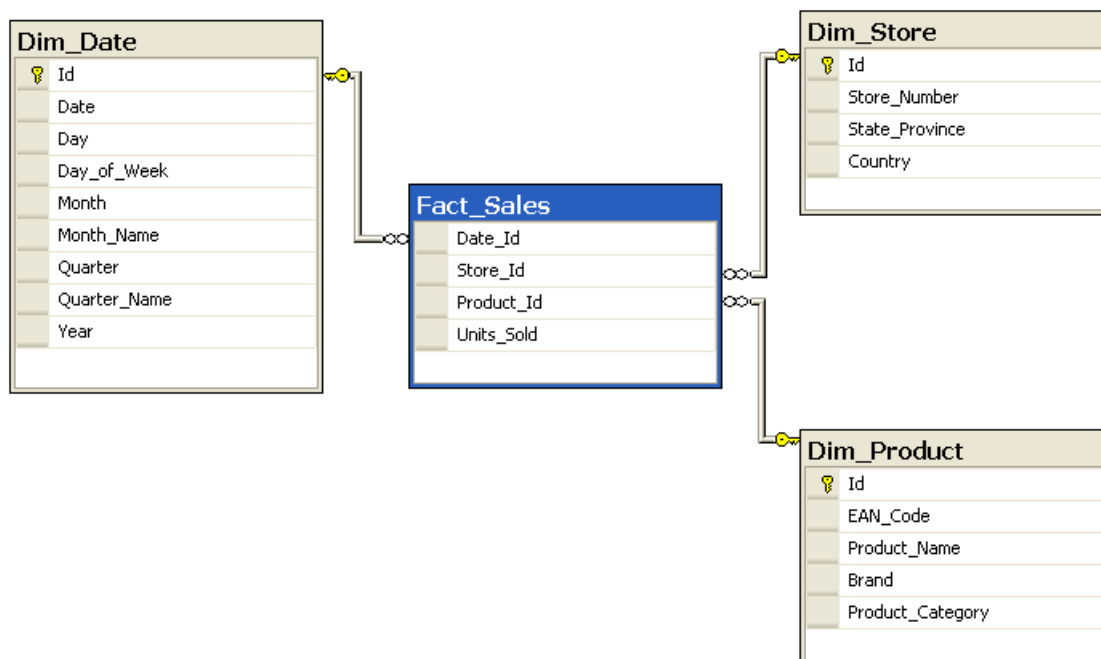
Companies with a data warehouse will always have one or more “normal” databases that feed the data warehouse.

### Normalized vs. Dimensional Data

OLTP systems rely on *normalized* data, meaning they seek to reduce data duplication and rely on dependencies between sets of data. Customers place orders, for example, and so an OLTP database might have one table for customer data and another for order data and would use relationships to connect customers to their orders. OLTP databases work well because in many situations only small pieces of data are being manipulated—entering a new customer or retrieving one order, for example. OLTP databases are tuned to support this behavior.

However, when it comes to BI, you don’t always need the fine details—like which customers ordered what products. Instead, you’re looking at bigger trends—like how many of a particular product were sold in the preceding quarter. OLTP databases can provide that information, but they’re a bit slower at it because their database structure isn’t fine-tuned to summarize and aggregate data in that fashion. In fact, *asking* an OLTP database to provide that information can impact performance of normal database operations—meaning your quest for BI will actually slow down business.

A data warehouse uses a different database structure entirely. Figure 1.3 shows a simple example of a *star schema*, one of the simplest data warehousing structures.



**Figure 1.3: An example star schema.**

Here, all the possible dates are pulled into one table—called a *dimension* table. All possible products are in a second dimension table, and all the company’s stores are in a third dimension. A central *fact table* links the three dimensions so that you can discover which products were sold in which store on what date. This structure is not optimized to reduce data redundancy—in fact, there will likely be a lot of duplicated data. That’s okay; data warehouses trade size for speed, meaning they often contain a great deal of redundant data, but that redundancy helps them produce results like reports much more rapidly. Using a schema like this, for example, makes it much easier to see the days on which televisions sold more quickly or whether there are any stores that sell laptop computers especially well.

### Data Modeling

The trick to an effective data warehouse is in the schema—how you *model* the data. You need to know in advance, to a degree, what questions you want answered by the data warehouse. For example, the example in Figure 1.3 won’t help you figure out whether more customers use Visa or American Express to purchase refrigerators because you didn’t include payment information in the schema. If that’s the type of thing you want to know, it needs to be included in the data model.

Data modeling is where large companies spend a lot of time when they’re implementing a data warehouse. Fortunately, midsize companies often share a lot of common data sources and business needs. Just as an off-the-shelf bookkeeping package tends to work well with just about any kind of midsize business, an off-the-shelf BI solution can also work well—without all the time-consuming up-front data modeling.



## Data Transformation and Loading

With your data warehouse schema ready, the next task is to get data into it. One reason a data warehouse works well is that it *copies* data from production systems into the warehouse; that means you can pull reports from the warehouse all day long without impacting your production systems. It also means that your warehouse's data will always be slightly out of date—exactly how much depends on how often you copy new data into it. However, most BI works from long-term trends, so not having up-to-the-moment data isn't usually an issue. There *is* such a thing as realtime BI, but it is beyond the scope of this book.

The actual process of copying data is usually referred to as an *extract, transform, and load* process (ETL). The extract step connects to the source database and pulls the required data from it. A good BI solution will have the capability to connect to most common databases; long-standing connectivity standards, adopted by most database vendors, makes this easier. The transform step rearranges the data into the schema used by the data warehouse. This may involve summarizing certain pieces of data, if the data warehouse won't contain line-item details, and it usually involves spreading the data out into the different table structure used by the data warehouse. Finally, the load step actually places the transformed data into the data warehouse. This may be a separate database running on a relational database management server or a proprietary database contained within the BI solution itself.

ETL isn't an all-or-nothing process. You may load data from certain sources once an hour, or every night; other sources—such as accounting systems—may become involved on a monthly or quarterly basis.

## In-Memory Analytics

A relatively new development in the field of BI is *in-memory analytics*. Rather than copying all your data to a different location—like a data warehouse—and performing analysis there, data is simply read into a server's memory and analyzed right then and there. Actually, the *idea* of in-memory analytics isn't new, but it's only recently—with the availability of more powerful computing resources at lower prices—that it's become practical. I'll discuss it in more detail in the next chapter.

## Reporting, Analysis, and Visualization

The last step, of course, is to *use* the data in the warehouse. This is where BI systems really come into play: Simply having a lot of data sitting in a database isn't terribly useful; the BI system contains the smarts to turn that *data* into useful reports, dashboards, scorecards, and other forms of *information*.

Good BI systems will allow users to work with the tools that they're already comfortable with. I've described how dashboards and scorecards provide intuitive summaries of data; some users may prefer straightforward reports, while others might prefer to work with pivot tables in a spreadsheet application such as Microsoft Excel. Figure 1.4 shows an example of these, and how a single BI solution can provide all these forms of output.

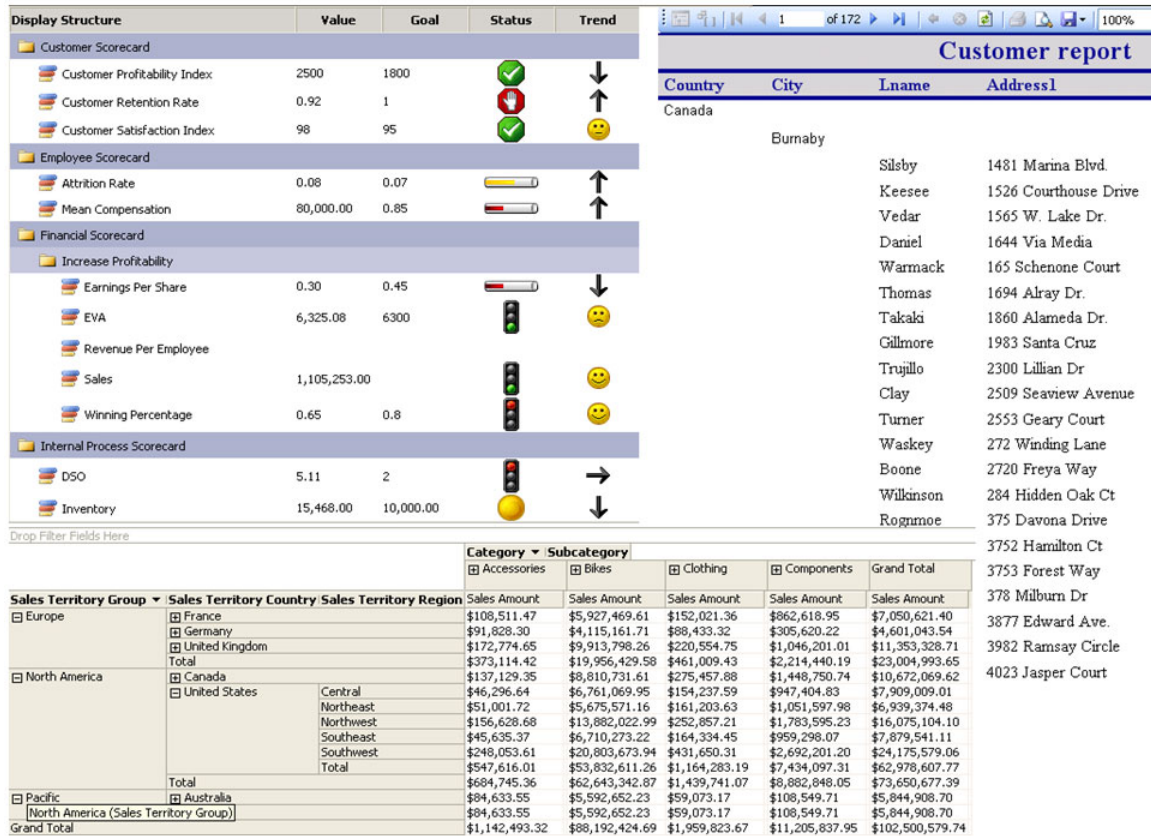
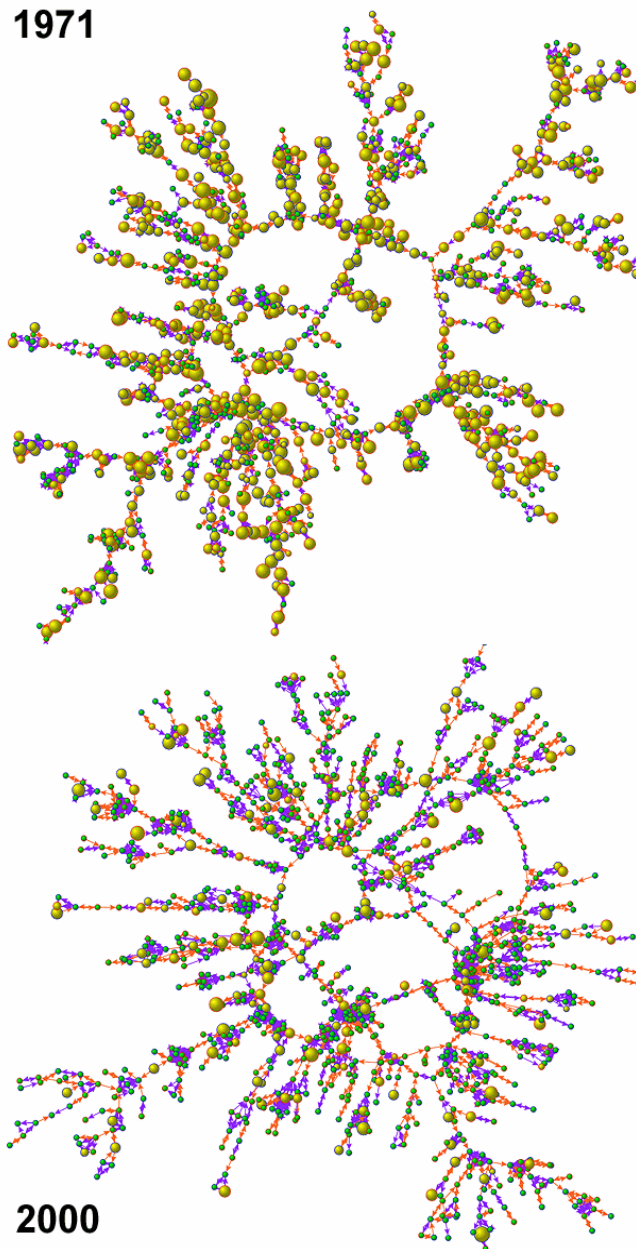


Figure 1.4: Common examples of data warehouse output.

Pivot tables—the “PivotTable” feature in Microsoft Excel, for example—can be useful analysis tools for users already comfortable working with spreadsheets. We’ll discuss them more in later chapters, but for now, keep in mind that pivot tables allow users to construct custom output, summarize key data on demand, and rearrange data to see different relationships.

BI systems can also provide more powerful visualizations for data. For example, Figure 1.5 shows a relationship chart, which helps visualize patterns in data. In this example, the yellow nodes represent individuals in a community who smoke cigarettes. Researchers used this chart to study the effects of the smokers’ relationships on their smoking, and found that after almost 30 years, most stopped smoking. The ones who continued smoking had few close relationships. This suggested a pattern of quitting—if your friends quit, you will too. It’s a type of relationship that only a data warehouse, and this specific kind of visualization, can unveil.



**Figure 1.5: An example relationship chart.**

Reporting, analysis, and visualization can all seem like the same thing—and they certainly all serve similar goals. So what are the real differences?

- **Reports** can address a variety of business needs, can be personalized and targeted, and can be easily distributed via many means. They can be adapted to run from different data sources as well, making them very flexible.
- **Dashboards** translate complex information into high-impact, at-a-glance displays. They're more intuitive than reports, and help keep decision makers aligned with organizational goals. Dashboards help you spot problems more quickly, making them good daily-use tools.

- **Scorecards** also deliver at-a-glance information but show how well you're doing compared with specific targets. They're good for communicating strategy and progress, and can help increase accountability.
- **Analysis** lets you explore information from across the business, and helps you dig deeper to discover the "why" behind what you're seeing on a report, dashboard, or scorecard. You can move through summary information to deeper-level information, and model "what-if" scenarios to help find the right action for improvement.

## BI: Analysis + Planning

All the business "intelligence" in the world is useless if you don't *use* it to change something about your business—or at least to validate that what you're currently doing is the best path. Good BI solutions facilitate this goal by not only providing intelligence but also by helping you analyze that intelligence, make decisions, and actually implement those decisions.

### Analyzing and Reporting on Information

Analyzing information can happen in many ways. You might use reports—or, more specifically, you might use an interactive *drill-down* or *drill-up* report. These reports let you click an item of data to see the detail behind that item, making it easier to dig to the source of something that is problematic. Or you might look at information based on ad-hoc queries, which can be presented in a variety of formats. You might also use dashboards or other visualizations, like the breakdown shown in Figure 1.6.



Figure 1.6: An example sales summary report.

If you're comfortable in Excel, you might create a PivotTable, or use one of the many BI add-ins for Excel. These can help you analyze "what if" scenarios, like seeing the effect of a change in product sales for a given month. All this analysis will help you draw conclusions about why things are the way they are within your company; you can then make informed decisions about which things need to change in order to meet your goals.

### Making Decisions

Making decisions can be the tough part. The trick is to make sure you're looking at *all* the data. BI solutions can help by correlating data from many sources, making it easier to see previously-siloed information in a single, cohesive view. See product sales next to manufacturing costs alongside payroll and facilities overhead. Drill-down reports help you see the nitty-gritty details easily and then "drill up" to summary views that show you the "big picture" results of proposed changes.

"What if" is where BI really comes in handy. As you look at projections and trends, BI solutions make it easy to plug in different numbers: What if we increased payroll and hired new shipping clerks? What if we use an additional vendor to source critical components? What if unit sales go down 10%? What if we offer bigger discounts for bulk orders? By punching in proposals, you can see—almost instantly—how the numbers fall out, to determine whether your proposals make a positive or negative impact.

### Implementing

Implementing is where the rubber meets the road: It's where you take your best proposals and put them into action. Doing so may require updating corporate budgets, financial outlooks, and other planning and support systems; a good BI solution will provide *write-back* capabilities so that your "what if" proposals can be "accepted" and linked back to business optimization and planning tools. In other words, just as a BI solution can *read* data from planning systems, it can also *write* data back into those systems to adjust assumptions and plans. It's a great way to "close the loop" in BI, making your "what if" proposals a reality, sooner.

### One Size Does Not Fit All

BI—like most business technologies—is not a one-size-fits-all affair. I actually get a little surprised when I see midsize businesses attempting to implement—or even considering—the same BI solutions used by a giant company like Ford Motor Company, Aetna, or Home Depot. Most midsize companies would never even look at the ERP, CRM, and other applications suitable for a giant company; most solution vendors have specific products geared toward midsize businesses. Why should BI be any different?



## BI in Large Enterprises

No two large enterprises are exactly alike—indeed, few of them are even vaguely similar, even when they're in the same industry. They may have all started as small companies, but they grew up in very different ways. They handle payroll differently, structure their accounting differently, have different manufacturing models and philosophies, and so on. Large enterprises are often organized into business units and divisions which themselves operate almost as independent entities. In some large enterprises—GE is a good example—there are divisions that have absolutely nothing to do with other divisions, such as GE's medical equipment division and their television broadcasting division (although GE's medical equipment does get suspiciously good placement in their networks' television shows).

All of this means that BI in a large enterprise is *complicated*. You're dealing with thousands of data sources in the ETL process and may have hundreds of different audiences that need different reports, dashboards, and scorecards. Every BI deployment in one of these enterprises is a custom affair, from designing the data warehouse schema to writing the ETL routines to developing the final output. Many BI vendors who work at this level don't sell "products;" they sell *toolkits*, along with deployment and implementation services—experts who use those tools to build a custom BI solution for each new customer.

Sound expensive? Sure—but don't forget that big enterprises have a lot of money on the line. If they spend a couple of million dollars implementing a BI solution, they may be looking to save *tens* of millions from that solution. But it *doesn't scale down*. Follow that same implementation process, using the same vendors and toolkits, and you'll spend a lot of money—and midsize companies aren't looking to *save* as much, so the BI investment doesn't seem worth it.

## BI in Midsize Companies

Midsize companies are a totally different animal. They're more likely, for example, to use off-the-shelf or lightly-customized solutions for key tasks such as bookkeeping, payroll, CRM, and so forth. Midsize companies tend to focus on similar broad business questions: Where are sales? How's inventory? What does payroll and other overhead look like? Midsize companies tend to have a lot of broad similarities with one another, in other words.

Thus, BI vendors can offer prepackaged BI solutions designed for midsize businesses. These solutions may require little or no outside expert services to implement, and they're designed in much the same way that midsize bookkeeping software is designed—to meet a common set of needs, with enough room for customization to ensure a good fit on most businesses. The BI vendor is able to spread the cost of designing the data warehouse, output, and ETL routines across all of its customers, resulting in a lower-cost, off-the-shelf product. It wouldn't be at all suitable for a Fortune 500 company—but it can work great for midsize businesses.

Prepackaged BI solutions often offer the most compelling features from the “full-sized” BI toolkits, such as Web-based dashboards, support for pivot tables in spreadsheets, scorecards, and so on. We’ll explore some of these features and help build a BI solution shopping list in an upcoming chapter.

### Coming Up Next...

So exactly what is involved in making BI happen inside your organization? How do you introduce a BI solution, and what sort of outputs will you need? How will you analyze that output, and what sorts of decisions can you expect to make? In the next chapter, I’ll attempt to answer all of these questions by explaining the basic process by which BI is introduced into an environment. We’ll dig a little deeper, technically, than in this chapter, but I’ll also focus on laying out the steps you’ll be taking. I’ll look at why BI is traditionally complicated and expensive, and outline some ways in which midsize companies can leverage certain economies of scale and similarity to save both time and money.

### Download Additional eBooks from Realtime Nexus!

Realtime Nexus—The Digital Library provides world-class expert resources that IT professionals depend on to learn about the newest technologies. If you found this eBook to be informative, we encourage you to download more of our industry-leading technology eBooks and video guides at Realtime Nexus. Please visit <http://nexus.realtimepublishers.com>.