



Delivering Self-service Business Intelligence

AUTHOR:
Marc Rueter
with Kevin Winslow

DATE:
July 2010

The authors gratefully acknowledge that the ideas and concepts presented in this paper are based in part on a web seminar titled "Delivering Self-Service BI with SQL Server" presented by Stacia Misner and Marc Rueter on behalf of Tableau Software. Misner is author of six books about business intelligence and SQL Server including "Microsoft SQL Server 2005 Reporting Services Step by Step." Marc Rueter is Director of Consulting for Tableau Software.

Self-service BI shifts the emphasis away from the processes required to manage data in a centralized store and toward a process for finding, accessing and integrating pertinent information as needed.

Why BI?

The promise of business intelligence, or BI, has always been to empower business users with better information than they could obtain from operational systems. While traditional BI continues to fulfill certain needs, *self-service* BI addresses the problem of ever-shifting demands for information.

In any organization, there exist multiple data sources that are constantly capturing information about that business. In addition to these sources (which are often in different formats), there are users with unanswered questions about their business. These users are seeking answers, but first need access to the data. Traditional BI relies on a centralized, cleansed and transformed store of data that users can access through standardized reports and perhaps an ad-hoc query tool. However, the development and maintenance of this type of system takes time and imposes constraints on the types of analysis users can perform – the data must be ‘in the system’ in order to be analyzed. On a day-to-day basis, the types of decisions that business users must make frequently require information that is not yet (or will never be) in the data warehouse. At best, a diligent user will spend the time to manually reformat and integrate the necessary data. At worst, users will make decisions without having all the information relevant to the problem at hand. Neither of these outcomes is best for the business or any of the typical endpoints along the spectrum.

Self-service BI shifts the emphasis away from the processes required to manage data in a centralized store and toward a process for finding, accessing and integrating pertinent information as needed. With self-service BI, decision makers are better able to respond to changes in business conditions quickly.

Common Problems

The problems with BI begin with the data itself. Unfortunately, the structure of the data doesn't always lend itself well to the types of questions being asked. These questions often require summarized data from sources that are frequently in Third Normal Form (3NF). This 3NF form does not summarize well or offer great performance. It can lead to slower query performance and impact users who are trying to run the daily operations of the business. These users will likely be unwilling to accept the impact of querying an operations database.

From here, a number of other problems emerge that can further impede successful BI. One issue is the skill level of individual users. It's not entirely reasonable to expect the average user to learn the query languages used against these systems. This means that either very technical users or specialists are needed, often from the IT department, to produce reports. This scenario brings its own problems as nearly every organization faces bottlenecks in its IT department. Requests for BI queries will only add to the existing, day-to-day backlog at IT.

Finally, once the reports are produced, they often fall short of expectations. They can fail to answer the original question or can even bring up new questions. Many times they don't yield any real, actionable insights. In short, it's rare that these reports can be built and be done with. Changes in the business (or in the competition) or even just the desire to *make* changes in the business will always prompt new questions. This leads to new requirements for new reports and a never-ending cycle of development.

With a single entry point, decision-makers quickly find, organize, automate, deliver and aggregate all forms of information supporting the decision process... without IT assistance.

The Data

Warehouse Institute

May 1999

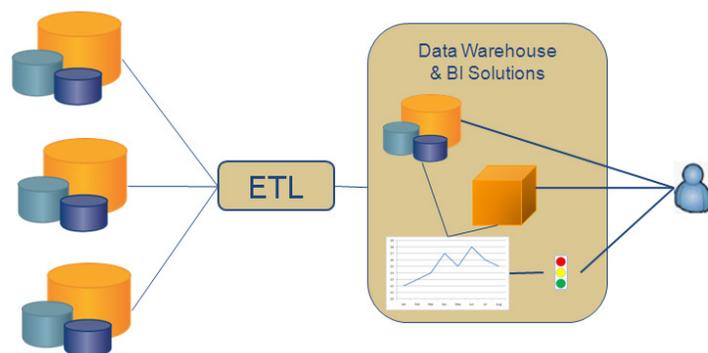
Self-service BI Empowers Solutions

In the face of all these issues, BI is looked upon for solutions and, in particular, self-service BI. The promise of self-service BI is nothing new and has been incubating for more than a decade. Witness the following quote from The Data Warehouse Institute (TDWI), published in May 1999:

A typical BI solution begins by integrating all an organization's various data sources using Extract, Transform and Load tools (ETLs). This process cleanses and consolidates the data sources into a relational database such as full-fledged data warehouse or a more targeted, subject-specific data mart. This data could also go into a multi-dimensional database or "cube." Regardless of the way the data is consolidated, there now exists a separate source that is structured in a way that's easier to query and offers faster query response times. More importantly, this source is now separate from the operational systems, so that the "read" operations of BI queries don't interfere with the "write" operations of day-to-day business transactions.

This BI solution doesn't typically end with developing new data sources. Reports and other tools are likely to be added to make it easier for people to use that data. In this scenario, users can work with the standard reports or choose to have key-performance scorecards and other BI-specialty applications. An additional layer allows users to query the relational or multi-dimensional databases as well. This process and its key components are expressed in the graphic below.

Problem solved?



With traditional BI, a solid, dependable, production-ready system can take months while more complex solutions may take years.

Unfortunately, building this system takes time. A solid, dependable, production-ready system can take months while more complex solutions may take years. All this effort is deemed worthwhile when business users have access to the right information and can help themselves as needed. In this situation, BI has solved at least some information problems for a subset of users.

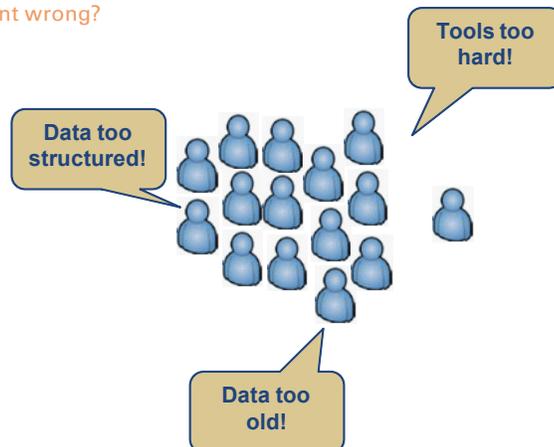
Reality Sets In

Despite the hope that self-service BI offered in 1999, The Data Warehouse Institute eight years later revealed that self-service BI was, in fact, "a myth." The Institute went on to say that though the concept was valid, its implementation was "misguided" and often resulted in "reporting chaos." This was something that many organizations had to learn the hard way. After building their solutions and amassing a collection of reports, organizations found it was still difficult for users to find critical information. And though "myth" might be too strong a word and many organizations have empowered their users through BI, the numbers are still somewhat grim. According to the 2009 BI Survey, slightly more than 8 percent of users in a typical company are regular users of BI applications – this includes static reports. While there are a number of factors that contribute to this scenario, we'll focus on three here.

One issue is that users often find the information is simply too old. In most BI solutions, data is refreshed daily or even hourly. But even in a real-time environment, there will always be data that users have gotten from other sources that needs to be analyzed. This means that despite efforts to grow the database organically, and periodically manage new data sources and requirements, the BI solution will never keep up with users' analysis needs and data requirements.

Two other key issues concern the structure of the data itself and the tools that users have available. Consider the following graphic as a representation of these three key barriers to self-service BI:

What went wrong?



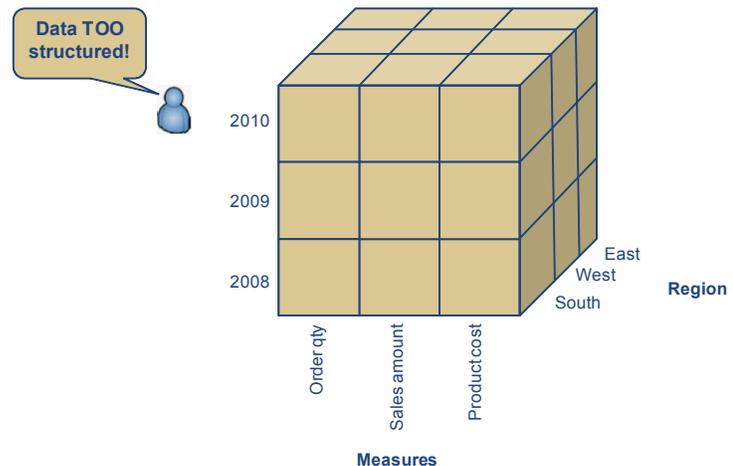
After building their solutions and amassing a collection of reports, organizations found it was still difficult for users to find critical information.

A Problem of Structure

Whether data is put into an OLAP cube or made into a dimensional model in a data warehouse, dimensions and measures are still created that, hopefully, make sense to the end user. However, research has shown that many of these designs are not particularly intuitive. Sometimes this occurs because the developer is new to dimensional modeling and does not yet have a good grasp of the principles. Other times it's because end users don't entirely know what it is they want or haven't fully determined how they're going to interact with the system. The quality of the model, therefore, hinges on anticipating the kinds of questions people might ask. It also assumes users know *what* to ask.

For example, as in the graphic below, data can be sliced and diced any number of ways: by time, by territory and by four additional measures. But this slicing and dicing depends on creating a rigid structure that must be built into the solution. This solution works well with stable, historical data and can empower users to make strong *strategic* decisions.

Cubes may not support ad-hoc analysis.



However, this structured BI solution often is not helpful for day-to-day tactical decisions. Imagine a new business opportunity, such as a new customer or sales territory. How can BI be leveraged to examine how these decisions would affect production levels or change profitability? A manager could work up projections in a spreadsheet, but how would that information be integrated into the historical data? In the exploratory phase of a new opportunity, most organizations wouldn't begin a new project simply to add this information into the data warehouse. This example illustrates the need for finding ways to deal with both the historical (or static) data, as well as the current (or dynamic) data.

A perfect system where users know exactly what to ask and all the necessary requirements have been identified still requires the proper tools. The half life of a question is typically about a third of the time it takes an organization to answer that question. Furthermore, requirements become stale immediately after they are gathered. Thus, the solution to self-service BI begins with better ways of identifying requirements and ends with better tools to get the job done.

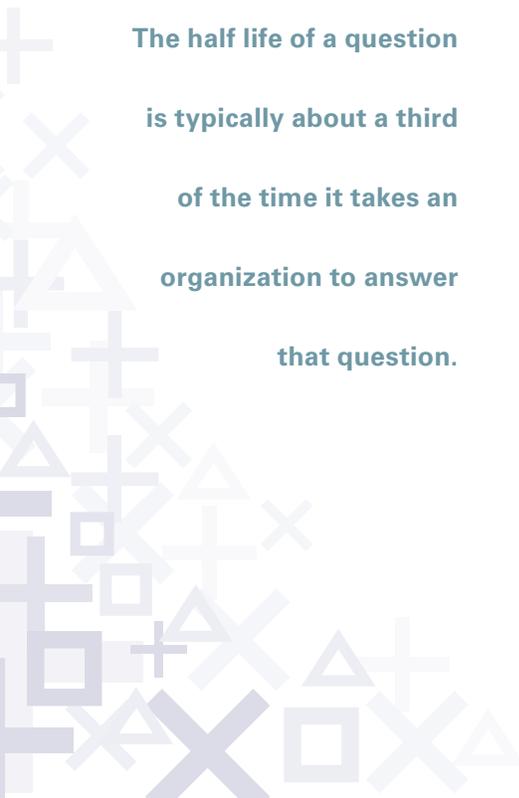
The Problem with Tools

Access to the right tools is a significant barrier to the wider adoption of BI. The problem is simple: users find BI tools too difficult. According to a 2009 Forrester Research study, two-thirds of respondents indicated that BI tools were either “difficult” or “very difficult” to learn. It should be noted, however, that this figure likely pertains more to the average user and not the analysts who spend most of their day sifting through data.

Many users have found success with WYSIWYG (What You See Is What You Get) tools. With this solution, instead of writing queries, users simply drag and drop the information they need onto a screen and perform the analysis. While WYSIWYG tools work well, they still depend on good structure and design running behind the scenes. Users still need a way to go into the data and see a high-level view of all the available information about a customer, product, sales territory, etc., and then have some sort of guided exploration. This way, users don’t have to *start* with a question. They can instead be led to interesting bits of information that enable them to explore and find the answers they’re looking for.

Back to IT

Data that’s too old and/or too structured, and tools that are too hard means IT gets thrown right back into the equation; between the data and the users. The problem with bringing IT back into the mix is that it again introduces the problem of timeliness. According to a 2009 study by *Database Trends and Applications*, only around 15 percent of users get their reports within a day of the initial request. Most are left waiting three days or even weeks before getting the answers they need. This extended wait often renders the initial question moot.



The half life of a question is typically about a third of the time it takes an organization to answer that question.

Spreadsheets to the Rescue?

One common solution to this dilemma is spreadsheets. When users become frustrated, they tend to do an end-run around the processes in place that manage the data. This way, they can get the answers they're looking for instantly. These users often look to spreadsheets to compile and present information. In fact, Forrester's 2009 study showed 88 percent of users rely "heavily" or "exclusively" on spreadsheets. Traditional BI is, therefore, not serving most users' needs in any meaningful way.

Consider the following scenario, which plays out all too often. An advanced Excel user creates an ad-hoc report to answer a specific question. Over time, this report morphs into a mission-critical application with intricacies that only one person (its original author) understands. This report is now very important and, at the same time, very fragile. But is it correct? Is it complete and/or up to date? How can it be changed to meet new demands? And what happens when its author leaves the organization?

In this environment, what's needed is a tool that gives users the freedom to examine problems from different angles and have conversations about different approaches to a problem. And while spreadsheets are great and useful tools, they have significant limitations. Consider the graphic below and the five problems inherent to spreadsheets.

Spreadsheets to the rescue?

	Multiple data sources	Hard to repeat processes	
Inconsistent calculations			Large data volumes
		Difficult to share	

Most users are left
waiting three days
or even weeks
before getting the
answers they need.

1. Inconsistent Calculations

One glaring problem with spreadsheets is that once they're shared, they begin to exhibit variations that lead to inconsistent calculations. For example, two users compile versions of the same data, but apply the calculations in different ways. In this all-too-common situation, the BI system – which was supposed to be the central location for business logic – has been completely circumvented. Even simple values such as revenue or profitability can be corrupted when different users come up with different calculations to arrive at those figures. This situation occurs when users don't have a transparent and consistent tool that empowers them to arrive at solutions on their own.

2. Too Much Data

Large data volumes yield large (and potentially unwieldy) workbooks. And while today's advanced spreadsheets are *technically* capable of storing a million rows of data, they may not necessarily like it. These large data sets often mean slow, poor performance when it comes to attempting complex calculations. Additionally, a large workbook can become unmanageable when its author tries to answer additional questions posed by colleagues that weren't in the initial design. This typically requires a complete restart on the calculations, the logic – even the data itself. Basically the author discards hours of investment and spends additional hours leveraging the knowledge gained in the initial exercise.

3. Multiple Sources

Some of the best analysis comes from using pivot tables in Microsoft Excel. But in order to build a pivot table, all data must be in a single table from all of available sources. To achieve this, it's likely that users are performing some sort of manual data integration. While this is not an impossible task, it's often tedious and error prone.

4. Repeatable Processes

When it comes time to refresh the data, the process may not prove repeatable. This is an especially tricky area. It may be a simple matter of replacing the data, but it could also mean several steps are needed to cleanse and consolidate it. Calculations must also be adapted to new rows or data elements. In this situation, users manually cleanse, filter and consolidate the data. In a data warehouse solution, these procedures would be handled automatically by ETL processes. A reliance on spreadsheets forces users to rely on manual processes and the requisite time necessary to maintain them.

5. Ability to Share

Even if users can overcome the above issues, spreadsheets must still be shared and, given their size, aren't likely to pass easily through e-mail. Sharing also makes security a consideration. Often only portions of a spreadsheet contain sensitive data. This forces users to carve them up to get the right data to the right people. Once this information is distributed, there's no way to guarantee its security. Users now have completely lost control of the contents with multiple versions of the spreadsheet in the hands of multiple users. This situation completely and effectively bypasses all the data governance built into a data warehouse.

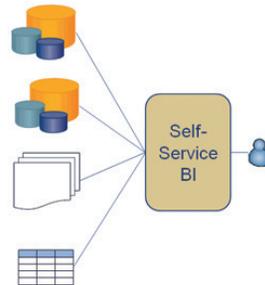
IT professionals must be prepared to deliver a new generation of self-service BI.

Moving Toward True Self-service BI

While all these factors make it appear as though self-service BI will never be a reality, the confluence of changes over the past decade means that it can be. Advances in the desktop PC mean vastly improved processing power and access to more data than ever before. In addition, new technologies in web interfaces offer a more simplified and streamlined user experience. However, in the face of these advances, the real key is to be clear about the problems that need to be solved. Designing systems with more sophistication and complexity only requires more expertise for deployment and administration. What's needed is a tool that empowers business users to answer questions using existing data sources and without the involvement of IT.

Self-service BI will not replace the BI systems we've focused on building in the past. To move forward, we need a compromise and a way to create complementary systems. There is a place for structured solutions and the standardized data warehouse. IT pros can build and maintain these resources, and make them available for users who need fast answers to questions that rely on historical data. However, these traditional systems have many times failed at meeting the ad-hoc reporting needs of users. Thus, IT professionals must be prepared to deliver a new generation of self-service BI. This solution will help to avoid the myriad of problems discussed previously. The graphic below illustrates this idea.

The new BI promise...



Is compromise.

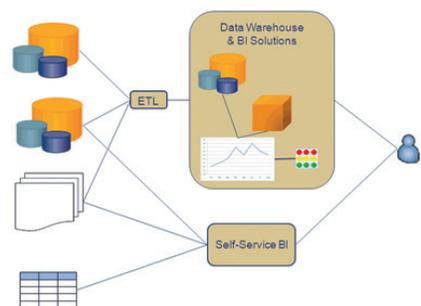


Tableau provides a way for users to answer their own questions and puts organizations on a rapid path to true self-service BI.

Moving Forward

To fulfill the promise of BI, we must create ways to:

- Access external sources and integrate those sources regardless of their origin or format – this would allow us to integrate historical data with data that comes in on a day-to-day basis
- Move easily from summarized views to more detailed views
- Change summarizations on the fly
- Work with large volumes of data, allow users to share this information and support individual users' personalization of data
- Allow IT to manage data in a way that guards against unlimited user access but does not impede the flow of data inward – this would ensure the implementation and maintenance of repeatable processes
- Update, secure and store data while ensuring it remains readily accessible
- Provide solutions that not only can be used by business users, but are also demanded by business users
- Remove the dependence on spreadsheets and provide a modern approach to data access and summarization
- Create repeatable, accurate processes that result in time savings and operational efficiencies while delivering actionable insights

Tableau Server is an internet-grade application with the expected security, robustness and scalability needed to securely share data with individuals or institutions outside your enterprise.

How Tableau Can Help

Tableau's strength is that it complements existing BI standards. While traditional BI offers standardized, consistent and high-volume reporting, Tableau adds an additional layer by giving users the information they need to answer the questions that come up each and every day. And where standard BI offers answers within days (or weeks), Tableau provides a way for users to answer their own questions and puts organizations on a rapid path to true self-service BI. Additionally, the Tableau solution deploys a live connection to the database for analytics, and lets users quickly build and share beautiful, effective dashboards – all without the need for programming or consultants. The result is high user adoption with minimal demand on IT resources.

Getting Started

Tableau works in a variety of ways. Tableau Professional lets users connect directly to relational databases, warehouses and cubes to perform their own self-service analysis. Users can connect to multiple data sources to combine and compile all relevant information about an issue. Once users perform their analysis, it is published to the Tableau Server via the web. Tableau Server then makes a direct connection to the backend data sources to generate the up-to-date results that users request. The publishing itself is embedded into Tableau Professional and needs no extra components.

With Tableau Server, users can share their analysis internally via the web. Because Tableau Server is a web-based system, any standard browser will allow users to consume the analysis that's been uploaded from the database. With this system, Tableau lets users publish their dashboard analysis via the web and share it with customers – either internal or external. Tableau Server is an internet-grade application with the expected security, robustness and scalability needed to securely share data with individuals or institutions outside your enterprise. For clients and/or vendors who may not have Tableau, there is Tableau Reader. The Reader is a free download and lets users share files and applications with anyone.

Additionally, Tableau visuals and other interactive graphics can be embedded directly into products like Microsoft Sharepoint sites and other web portals. Tableau offers products for every user and every organization; solutions range from web-only to free, to paid and robust.

Drag-and-Drop Simplicity

Tableau Professional makes it easy to create reports, to do analysis, to add other data sources, and to publish the entire content to Tableau Server and share it throughout any organization. Tableau supports a wide array of data sources and there's no need for an existing data cube or warehouse. Users simply enter the name of the server, then choose the relevant database and the table, view or set of tables they wish to examine. From there, users can begin their analysis by simply dragging and dropping the appropriate fields onto the screen. Tableau automatically chooses the proper dimensions and measures, and if any fields appear to be miscategorized or out of order, it's simply a matter of dragging them into the right place.

Once a visualization has been created, users can drag and drop information in and out for an instant, dynamic graphical representation of the data. Tableau Professional allows users to browse through an entire database with drag-and-drop simplicity. This applies to data from other sources as well. Even if users obtain information from an outside source, such as a customer or vendor, that data can easily be dragged-and-dropped into Tableau for an instant, updated visualization. Tableau even works with multiple data sources. For example, users can manipulate data from SQL Server and an Excel file simultaneously. Information from those sources can then be combined into a single dashboard.

Additionally, at any point during the session, users can rename any of the data as they see fit. Any labels, any axis, any field names – everything can be renamed at the user's discretion. Users may also select Tableau's "Show Me" capability for a wide array of different views. From there, users can pick from a list of suggested visualizations or choose one at their own discretion.

All of this information can be deployed to Tableau Server with a single click. Tableau then lets users decide which sheets to publish, give permissions to specific users and grant access to specific areas of information. Finally, Tableau Server generates a fully interactive preview of the data which can be opened in a web browser, e-mailed to other users or embedded into other applications.

How Tableau Works

To begin, Tableau first establishes a direct connection to an organization's database. From there, all data fields are displayed on the left-hand side of the screen. Here, users can either double click or drag the fields directly into a visualization. Adding more fields queries the database and refreshes the visualization to reflect the new data. For example, as in the graphic below, if we add the "Sales" field, Tableau sends a query to the database and a single result row is returned. Add in the "Markets" field and Tableau returns four rows of results, one for each market. Regardless of the amount of data in the back-end database, whether it's hundreds or even billions of rows, only four result rows appear in the visualization. Adding the "Time" field generates a new visualization with 96 rows of results – one for each of the four markets in each month over the past two years (4x12x2 = 96). This system allows organizations to leverage investments made in their existing database.

Drag and drop to instantly analyze your data.



To drill down into the data, simply click on any point in the “Time” chart, as in the graphic below. Another query is sent and the detail rows are returned in a new window. All the raw data will be returned for that particular point on the visualization. In the instance below, that single point on the visualization represents a summary of 4,248 rows of data.

Drill down with a click.

The screenshot shows a Tableau interface. On the left, there are dimension and measure shelves. The 'View Data' window is open, displaying a table with the following columns: Sales, Date, Market, Area Code, Market Size, Product Type, and Product. The table contains 4,248 rows of data. A red arrow points from a point on a chart in the background to the 'View Data' window, with the word 'Click' next to it. Below the window, a label '4,248 Rows' points to the data source icons.

Sales	Date	Market	Area Code	Market Size	Product Type	Product
\$128	12/1/2009 12:00	South	995	Small Market	Coffee	Colombian
\$228	12/1/2009 12:00	South	995	Small Market	Herbal Tea	Chamomile
\$106	12/1/2009 12:00	East	978	Major Market	Espresso	Cafe Mocha
\$87	12/1/2009 12:00	East	978	Major Market	Tea	Darjeeling
\$326	12/1/2009 12:00	West	971	Small Market	Espresso	Decaf Espresso
\$110	12/1/2009 12:00	West	971	Small Market	Espresso	Cafe Latte
\$236	12/1/2009 12:00	West	971	Small Market	Tea	Earl Grey
\$166	12/1/2009 12:00	Central	970	Major Market	Coffee	Colombian
\$142	12/1/2009 12:00	Central	970	Major Market	Espresso	Cafe Mocha
\$124	12/1/2009 12:00	Central	970	Major Market	Tea	Earl Grey
\$195	12/1/2009 12:00	Central	970	Major Market	Herbal Tea	Mint
\$326	12/1/2009 12:00	Central	970	Major Market	Herbal Tea	Chamomile
\$389	12/1/2009 12:00	East	959	Small Market	Coffee	Colombian
\$87	12/1/2009 12:00	East	959	Small Market	Tea	Green Tea
\$131	12/1/2009 12:00	East	959	Small Market	Tea	Darjeeling
\$106	12/1/2009 12:00	South	956	Major Market	Coffee	Decaf Irish Cream
\$205	12/1/2009 12:00	West	951	Major Market	Coffee	Decaf Irish Cream
\$334	12/1/2009 12:00	West	949	Major Market	Espresso	Cafe Latte
\$326	12/1/2009 12:00	Central	937	Small Market	Espresso	Cafe Mocha

On top of all this, Tableau features a sophisticated language that performs calculations from within a single row of data or across multiple rows. One common scenario where this would come into play would be finding profit ratio. Taking the sum of profits divided by the sum of sales is very easily expressed with Tableau. No matter how the user decides to slice or filter the results, the profit ratio is still accurate. Another scenario – year-over-year growth – is also just a click away. Both of these examples can be expressed by simply typing in an additional field like in Excel, but through Tableau’s own user interface. This yields calculations that are universally applicable regardless of the user’s scenario or how the data changes.

**Some customers
have reported
gaining insights
from their data
within 15 minutes
of downloading the
free trial.**

Fast, Easy Implementation

Some customers have reported gaining insights from their data within 15 minutes of downloading the free trial: without any training or any IT involvement and using existing data stores. These customers indicated their initial findings represented enough ROI to justify buying the full product.

With Tableau, training is free and available on the web, and there are no hidden costs associated with implementation. A full, enterprise-wide reporting platform with ETL processes and data warehousing would incur myriad additional investments, but self-service BI and a connection to existing data sources is included in the cost of the product. No separate physical server is needed. Some organizations choose the desktop solution, connect it directly to existing data stores and let users begin their analysis. Other than giving users permission to install the product, no other involvement from IT is typically necessary. This scenario represents the real draw of Tableau: it puts information in users' hands and takes IT virtually out of the equation.

The Tableau Server is an optional next step and represents a relatively small installation. Because it's not a data warehouse or store, it does not need to be sized or scaled like other servers. And while implementation of the Server does involve IT personnel, it's designed to require very little maintenance. Twenty minutes is all it takes for a typical set up and install.

Enterprise-grade Security

Tableau offers two different types of security. On the desktop, Tableau honors whatever database security is already in place – it grants no additional permissions. Users retain only their existing permissions and access to data they can already see. Tableau Server layers in row-level security for individual users or groups and enables sophisticated, automated permissions. As new users are added or as data permissions change, Tableau automatically adapts. By combining existing database security with application security and row-level data security, Tableau can satisfy the needs of even the most stringent organizations.

In Summary

Tableau's solution has successfully closed the loop in business intelligence. It complements traditional ETL systems by adding self-service BI, or can stand alone as a self-service BI application. With Tableau, business users can create interactive reports and dashboards, and distribute them throughout any organization.

About Tableau

Get a free trial copy of Tableau Software to test the power of rapid-fire BI at www.tableausoftware.com/trial. Tableau Software, a privately held company based in Seattle WA, provides business intelligence software applications that are fast to install and easy to learn. The power of web-based data visualization and BI enables business people to quickly make discoveries and share insights from all types of databases including large data warehouses.