Tableau and the Enterprise Data Warehouse:

The Visual Approach to Business Intelligence

AUTHOR
Dan Jewett
Tableau Software

DATE
08/01/2009

The imperative for IT organizations today? Empowering the knowledge worker to perform his own data analysis, to identify trends, outliers and anomalies, and to monitor progress of key initiatives.

Organizations focus on acquiring customers, increasing revenues and profitability, building and delivering quality products and services, and in general outmaneuvering the competition. Understanding finances, customer activity, product delivery, and yes, even the impact of yesterday's business decisions all need to occur on a more compressed schedule than ever before.

Companies need to arm their employees with the tools and data to enable immediate investigation and understanding of the business. People need relevant analytics to identify trends, outliers and anomalies, to share findings, and to monitor progress of initiatives. Supporting those efforts is the imperative for IT organizations today.

Over the years, organizations have built enterprise data warehouses (EDWs) to hold the ever-increasing amounts of business data being generated. The success IT has had in capturing every data transaction has made a rich store of data available to help guide business decisions. Instead, the volume of data available in the EDW has overwhelmed and frustrated the business user.

Traditional business intelligence (BI) platforms were developed to solve this problem but typically resulted in stifling the analytical curiosity of the business user and burdening IT with report-writing requests. Far too often, there is a significant gap in the promise and reality of the traditional BI projects.

What's needed is an approach that closes the gap, one that leverages the work of the data steward building the EDW while solving the business intelligence needs of the business user. Tableau Software was designed from the ground up by database and graphics experts to close exactly this gap. The combination of a

The Need	What Traditional Software Delivers
Business users want access to data to explore and understand it.	Traditional applications focus on fixed, sometimes parameterized reports.
Business users want to tease out the interesting and relevant aspects of the data, asking a variety of questions in an iterative process in a search for understanding.	Traditional applications create infrastructures for specialized report developers to create environments to protect the user.
Business users want analytic applications to help them perform analysis and gain understanding.	Traditional applications focus on answering yesterday's questions in the form of static reports.
IT data stewards want to provide clean, correct data to their user communities.	Traditional applications often interfere with data stewardship by either re-hosting data out of the EDW or prescribing specific data schemas that must be created.
IT data stewards want to ensure that access, security, and performance of the EDW remain appropriate for the overall organizations needs.	Traditional applications can add to the data steward's burden on this front rather than relieving it.
Organizations want to gain value, even a competitive advantage from their EDW projects.	Traditional applications add complexity and delays to the project, which too often delays or short-circuits success the EDW can bring to an organization.

powerful EDW with Tableau Software makes for an enlightened solution that can finally deliver on the promise of business intelligence.

Tableau provides powerful, flexible applications that anyone can use to understand data from any source, including high-performance, high capacity enterprise data warehouses. The resulting dashboards, reports and visualizations can easily be shared across organizations in web-based analytics. Using Tableau to access, visualize and share the data in an EDW allows for an analyst to be active with the data, to ask and answer questions, and to naturally benefit from data warehouse investments.

Visualization helps people understand data

Enterprise data sets are large, detailed, and complex. For the spreadsheet user or typical report recipient looking for the underlying patterns, this is a worst-case combination. Spreadsheet applications have a lot of power, but are hard to use against large, complex sets of data. Reports are often designed to serve as detailed lookups so they have rows and columns of summarized data, making it difficult to see overall patterns and trends.

In the example below, drawing a conclusion from a simple cross-tabulation is difficult. Compare the cross-tab report to the graph below it. The graph makes it instantly obvious where sales are higher and profits are a problem.

Visualization is the difference. Data visualization helps communicate information more rapidly due to the fact that the human brain can process and understand a picture faster than it can process and understand a series of numbers that have to be compared and contrasted. As Colin Ware put it so aptly in his 2004 book Information Visualization,

"At higher levels of processing, perception and cognition are closely
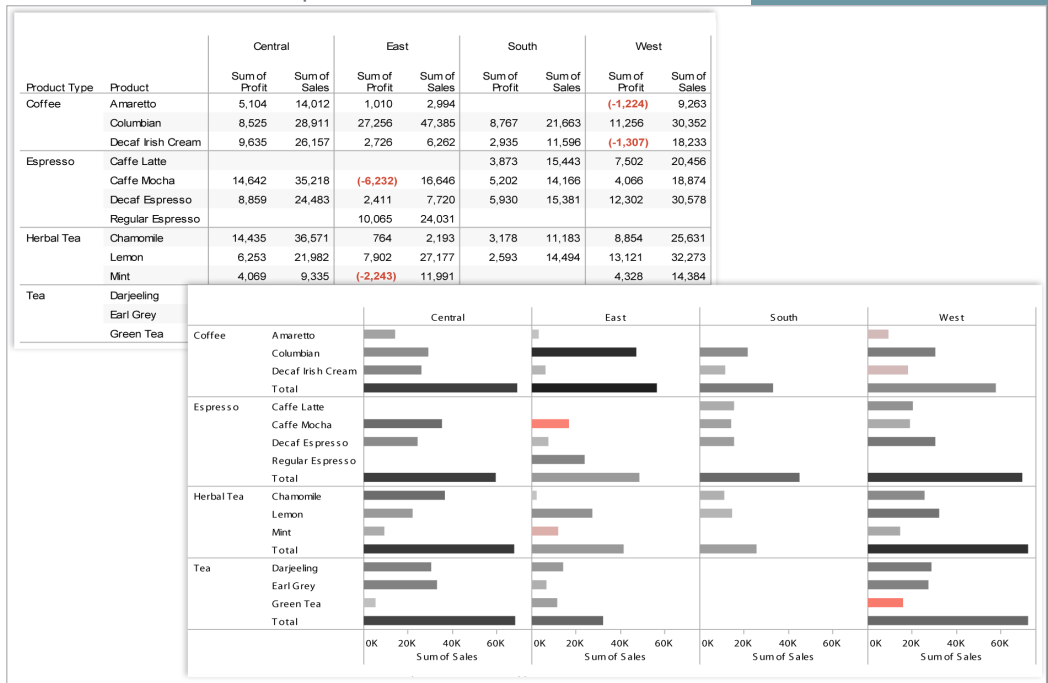
Figure 1: Drawing a conclusion from a simple cross-tabulation is difficult. Compare the cross-tab report to the graph below it. The graph makes it instantly obvious where sales are higher and profits are a problem.

Product Type	Product	Central		East		South		West	
		Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales	Sum of Profit	Sum of Sales
Coffee	Amaretto	5,104	14,012	1,010	2,994			(-1,224)	9,263
	Columbian	8,525	28,911	27,256	47,385	8,767	21,663	11,256	30,352
	Decaf Irish Cream	9,635	26,157	2,726	6,262	2,935	11,596	(-1,307)	18,233
Espresso	Caffe Latte					3,873	15,443	7,502	20,456
	Caffe Mocha	14,642	35,218	(-6,232)	16,646	5,202	14,166	4,066	18,874
	Decaf Espresso	8,859	24,483	2,411	7,720	5,930	15,381	12,302	30,578
	Regular Espresso			10,065	24,031				
Herbal Tea	Chamomile	14,435	36,571	764	2,193	3,178	11,183	8,854	25,631
	Lemon	6,253	21,982	7,902	27,177	2,593	14,494	13,121	32,273
	Mint	4,069	9,335	(-2,243)	11,991			4,328	14,384
Tea	Darjeeling								
	Earl Grey								
	Green Tea								

interrelated, which is the reason why the words 'understanding' and 'seeing' are synonymous."

Traditional BI applications attempt to help by providing graphs, dashboards, and text-based reports. These do not go far enough. Just as users must be able to ask ad-hoc questions, they also need effective ways of slicing, dicing and interrogating the data. The static, inflexible graph or dashboard that led to yesterday's insight may not work for today's problem.

Tableau solves this problem by providing software that leverages best-practices for data visualization and business intelligence. It gives users flexible, easy ways to display data in whatever form makes the patterns most visible. The process of analysis is built on the actual data – not an abstraction. A business user is able to create and share their own dashboards, reports and views that make the underlying data patterns obvious in a glance. Using these live dashboards, reports and views from a web browser, colleagues and managers can manipulate, filter, customize and save these analytical results as highly relevant decision making support tools.

Your EDW provides the computing capability and architecture that allows massive amounts of data or summaries of those data to be delivered. Tableau lets business people appropriately access, manipulate and share those analytical frameworks. Together, they allow the business user to gain insight by changing perspectives and seeing information as data are compared. Thus, the business person does not lose perspective as he mentally navigates from one question to the next. And because he is more effective, he can be more effective in sharing and communicating information for better and faster decision making.

Leveraging the Enterprise Data Warehouse

Tableau accesses data and databases where they live – in the enterprise data warehouse, on the desktop, in the cloud, wherever. The conversation between your data and Tableau is driven by Tableau's proprietary technology, VizQL™. It is the technology that enables users to create stunning graphs and reports interactively and easily from large datasets. VizQL particularly flourishes when this database is fast, large, and has a variety of data subjects stored in it.

Recently featured in an article by Jim Gray published posthumously in the prestigious Communications of the ACM (Association of Computing Machinery), VizQL is a declarative visual query language that unifies the strengths of visualization and databases. People can visualize relationships between data using shape, size, orientation, color, text and texture in all kinds of graphs, and it leverages the advances in database systems to optimize access performance to large datasets. This combination lets users interactively explore the raw data or perform data analysis. It is a major improvement over how analysis is currently done.

Because Tableau takes a subject-oriented view of the data warehouse, it accesses only the data tables that are relevant to the subject at hand as part of a connection. This is a much preferred model by business users, as they have access to just the relevant data items in the subject they elect to work with. Compared to BI tools that require the full database to be logically modeled as an additional prerequisite step, or worse yet actually built as a particular schema type, this design choice by Tableau is valuable, convenient and well understood by DBAs as well as appreciated by end users.

In terms of security, Tableau is bound by all constraints defined at the database level, which helps organizations ensure that access is properly controlled. This is in direct contrast with the design of many other BI solutions which require the data to be moved out of the EDW and stored in their own mid-tier or other data structures. The traditional BI approach has the potential to create silos of special purpose, and possibly incorrectly managed data stores.

Ad-hoc exploration with Tableau

Typical BI applications deliver a fixed set of data that is intended to help answer a fixed set of questions for a given point in time. But business needs and conditions are in constant flux. Users need to explore enterprise data in new ways. They cannot anticipate the new questions they will need to ask, and each question answered can lead to more questions. The traditional approach (of providing a fixed set of data or reports) can never give users the flexibility to answer all the questions they have.

Tableau solves this problem with ad-hoc, on-demand, interactive exploration.

The ad hoc, data exploration experience is enhanced by a well-built database. Speed-of-thought analysis, iteratively asking a question, studying the answer, and asking another question, is the ideal workflow for a knowledge worker. Slow response times, disruptive context switching between systems or tools, convoluted steps to write queries or import data all prevent the natural flow of the analyst.

Although useful in the right context, fixed dataset extracts were often used as a default because they limited problems with slow response times of traditional BI platforms running on large EDWs. But the underlying size of the data is not an issue with Tableau, since the views dynamically query the database to retrieve just the data needed to answer the questions asked. Modify the graph to fine-tune the question and the system automatically determines what queries to run to create the new graph. People move seamlessly through analysis simply by interacting with visual representations of data.

The queries that Tableau automatically generates behind the scenes are optimized to support the visualization being created. Tableau automatically requests the minimum set of data needed to answer each question. This includes asking the database to perform the aggregations, filters, and calculations of the data requested. Some of the concepts used include derived

The underlying size of the data is not an issue with Tableau, since the process of working the views dynamically queries the database to retrieve just the data needed to answer the questions asked.

tables, such as the use of temporary tables, and leverage built-in database functions and SQL.

The Tableau way is in high contrast with traditional tools of query wizards, direct SQL language dialogs, report grids, and even chart wizards.

And because Tableau can use the EDW as its data source, the data are always as current as the EDW. When a business user publishes his interactive dashboards, reports and graphs to share with colleagues, Tableau carries with it the connections and instructions related to querying the EDW.

Summary

Visual analytics is becoming the fastest way for people to explore and understand data. On top of a large enterprise data warehouse, Tableau's visual analytics and BI capabilities can help users serve themselves exactly the data they need when needed, all while adhering to DBA and IT standards.

For organizations with large enterprise data warehouses, Tableau helps people see, understand and share data faster and more easily than ever before. The visual analysis process enables the business user to be more inquisitive, to be more informed, to make smarter, faster decisions and to share their insights across the organization via a browser. The combination of Tableau and an EDW results in a high-performance data access and BI system that is a key enabler of shared understanding.

Tableau customers include Microsoft, Google, eBay, Wal-Mart, Merck, Hewlett-Packard, Countrywide, Fidelity, GE, Charles Schwab, Harvard University, and the U.S. Government. Contact your Tableau representative or visit www.tableausoftware.com.

Frequently Asked Questions

How does Tableau connect to enterprise data warehouses?
Tableau connects directly to the leading EDW databases, including Teradata, Oracle, SQL Server, Netezza, Vertica, Postgres, Essbase and MySQL.

Do I need to do anything special to my database for Tableau to access it?
No. Tableau will access any data that reside in the database "as is." There is no need to create special schemas, special metadata, or alter your database in any other way.

We have hundreds of tables in our EDW. How do business users create the right connections?
Tableau behaves like all good applications and provides security clearance before it is granted access to any tables in your database. Getting "connected" to an enterprise data warehouse is as straight forward as providing an active User Name and Password at appropriate Tableau prompts. User Name and Password then grant access to appropriate tables. In fact, if your EDW is like most deployments, you may allow users to connect to tables directly, but often you prefer to give them views that are optimized to their understanding of the data or to join indexes. Tableau can use any or all of those connection means and even can connect to SQL statements to each of those. Those connections can be saved and published for easy access by others so that users can rapidly see their data the way that you, the DBA, want them to see it.

Without a semantic layer, how does Tableau differentiate dimensions and measures?
During its initial read of the database data dictionary, Tableau applies certain criteria to determine if data should be categorized as either dimensions or measures. This initial categorization is presented to the Tableau user and can be changed as required. For example, it is a simple "drag and drop" move for a Tableau user to re-categorize data between dimensions and measures. Once a Tableau connection is saved, the definitions of these dimensions and measures are stored with the connection and can be shared across the Tableau community.

How do business users decipher the cryptic column names that we use?

For users connecting directly to data, initial "deciphering" may require a consultation with a DBA or at least with the data dictionary. Either way, if a user is not sure of the business meaning of column names as presented in the Tableau data display, they should consult one internal sources.

Alternatively, IT management staff may choose to create a Tableau "data connection". Within this connection, they can literally specify connections to the right tables, views and indexes and specify how the metadata should be presented. This could include connecting to SQL statements, renaming columns, hiding columns, defining roll-up groups, writing custom calculations and aliasing

data stubs or content. The metadata can be customized without changing the underlying data; there's always a traceable path as to the source data elements.

Thus, with regard to understanding the meaning of the display of terms and names on Tableau results, Tableau provides the capability to display aliases for database terms. These aliases appear only on Tableau results.

Some tables have dozens of unusable columns. How do business users manage around this?

One of the great advantages of Tableau is that it does not presume to know the business value of data in a data warehouse. Consequently, within the security policies of a business, Tableau makes any and all data available to a Tableau user who has been granted access to the data. Not all data are relevant. However, decisions with regard to relevance are made by the Tableau user. Part of the definition of a Tableau data connection allows for hiding unneeded fields. An individual user may show/hide the unneeded fields, or the data steward creating the shared connections can hide the unneeded columns. And of course, if the data steward is creating database views for users to access, only the columns defined in the views will be displayed.

I have some hand-optimized SQL that defines the set of data I want to work with. Can Tableau use this SQL?

Yes. When initially establishing a connection to a database, there is a choice to either browse the list of tables (and views) available on the database to work with, or there is the option to provide a custom SQL statement to use as the starting point for your analysis. This custom SQL statement is useful when performing tasks, such as unions or working with highly optimized SQL statements to define a subset of data for analysis.

Can Tableau work with views?

Yes. Creating database views is a very common best practice to help shield the complexity of the database from the business users.

How big can my data be? Are there any size limits when using Tableau?

There is no size limit to the databases that Tableau can access. Tableau's design is such that it doesn't take data out of the database to host it somewhere else. It accesses data directly from the database.

Tableau is a visual analysis application, so the individual visualizations, reports and dashboards that are created in Tableau generally retrieve a relatively smaller number of rows from the database, automatically building the queries to have the database aggregate and filter the data before returning it to Tableau. As such, the data volume returned to Tableau is generally modest compared to the terabytes that are stored in the database.

How does Tableau prevent queries that would cripple the data warehouse?
Because Tableau is a good citizen, it will take full advantage of all database services, both those provided by the technology and those provided by DBAs. There is no substitute for performance testing and ongoing collection of statistics and monitoring to ensure the behavior and performance of any application, and Tableau applications are no exception. However, one of the great advantages of Tableau is the natural restrictions it places on the selection of data simply because of the human constraints of the viewer. Tableau only selects data required to populate the visual in front of the viewer. Generally a Tableau visualization includes only a few Dimensions and Measures simply because the viewer is using visuals to reduce the complexity of the data in the EDW in the first place. Tableau does not ask for all of the members of a Cube to be populated so that it has all of the data that might be required when a Cube is "sliced and diced". Tableau only needs to request the data that are required of the visual in front of the viewer at any moment in time view is constrained the number of dimensions the viewer can visually comprehend. So between good DBA practices and the natural constraints of visualizing data at any moment in time, Tableau's behavior with large databases is generally exceptional.

Is Tableau secure? What data can end users access?
If the database requires a username and password to establish a connection, then Tableau will require the same. Typically an individual user is given grant authority to only a subset of the tables and views in the database, which is how DBAs control the data that can be accessed.

Does Tableau Server also require individual user accounts? I would like to shield those users from having to login again.
Tableau Server has the option for the original author of a set of analysis to specify the username and password that the database which should always be used for that workbook. This is like an 'application password', where anyone accessing that workbook will be automatically connected to the database as a pre-defined username. This capability is optional and may be restricted by the Tableau Server administrator.

Does Tableau support data level security (row and column)?
Yes. Any data level restrictions defined by the DBA as part of view definitions is automatically inherited by Tableau. When using Tableau Server in conjunction with database, additional row level data constraints may be specified to subset the data retrieved based on the individual viewing the content.

Can I work disconnected from my database? Or do I need to be online all the time?
Yes.   Tableau has the concept of an Extract File to allow working offline from the database. The extract file allows you to take a snapshot of some data from the original database and hold it locally on your PC. This allows you to work disconnected from the database. Once you complete your offline work, simply select a menu item to return your workbook to querying the live database, and it will automatically not use the data snapshot any longer.

I have a large number of tables in my database. Do I have to include them all when I make a connection in Tableau?

No. Tableau encourages working only with the set of tables that is relevant to the current topic being analyzed. This is generally called a subject-oriented perspective. The basic principle is to include only the minimum set of tables in the active connection to provide the data for the current subject. For example, if you are doing some retention analysis from an HR database, there is no reason to have your connection definition include tables from a finance or customer database. By taking a subject-oriented view, the business analyst has just the measures and dimensions applicable to their questions at hand. This simple concept is a key best practice in making the data warehouse useful to business analysts. Data stewards and DBAs can help by creating Tableau connection definitions (called TDS files) that are subject oriented and sharing them with the rest of the Tableau community..

Does Tableau require me to have a username?

The desktop version of Tableau does require each individual to have a username and password for each database accessed. In addition, Tableau has capabilities to enable secure means of sharing data.