



Tableau Software and Teradata:
The Visual Approach to the Active Warehouse

AUTHOR

William Smith
Claraview, a division of
Teradata

Dan Jewett
Tableau Software

DATE
04/30/2008



Empowering the knowledge worker to perform her own research into the data to identify trends, outliers, anomalies, and to monitor progress of initiatives is the imperative for IT organizations today.



Tableau (www.tableausoftware.com) is a powerful, flexible application for visually analyzing and understanding data. Teradata (www.teradata.com) is the industry leader in high performance, high capacity enterprise data warehouses. The combination of using Tableau to access, visualize and analyze the data in Teradata allows for an analyst to be active with the data in a process of asking and answering questions – naturally taking advantage of the Teradata Active Data Warehouse concepts.

Now more than ever, companies need to arm their knowledge workers with the tools and data to enable immediate exploration of the state of the business. Understanding the financial position, the activity of customers, and the impact of yesterday’s business decisions all need to occur on a more compressed schedule than ever before. Empowering the knowledge worker to perform her own research into the data to identify trends, outliers, anomalies, and to monitor progress of initiatives is the imperative for IT organizations today.

Over the years, organizations have built enterprise data warehouses (EDW) 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. Ironically, the volume of data available in the EDW has the potential to overwhelm the data analyst.

The traditional software response to meet the data access and business intelligence problems causes challenges for both the analyst and the stewards of the EDW. Far too often, there is a significant gap in the promise and reality of the EDW projects.

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 and 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 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, too often delaying or short-circuiting success the EDW can bring to an organization.



Tableau was designed from the ground up to solve the needs of the business user. Teradata Database was designed from the ground up to solve the needs of the data steward building the EDW. The combination of the two products together makes for an enlightened EDW solution.

Let's further look at these needs and how Tableau and Teradata bring a modern approach to answering them.

The value is in understanding the data, not just tabulating it

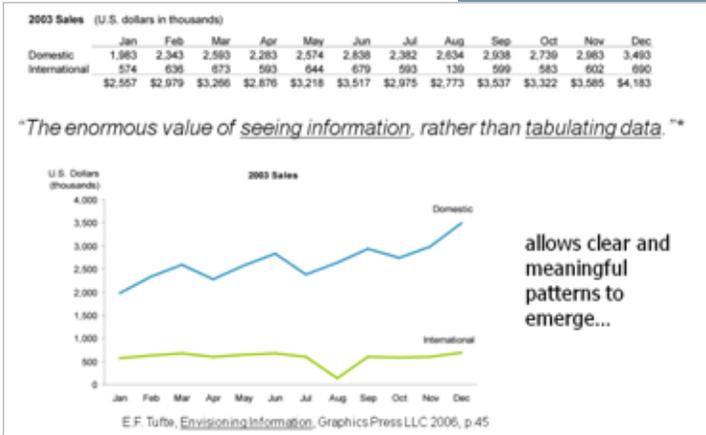
Enterprise data sets are large, detailed, and often complex. For the spreadsheet or typical report recipient looking for the underlying patterns, this is a worst-case combination. Spreadsheet applications have a lot of power, but are hard to use against large, detailed data. Reports are often designed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

In this example, drawing a conclusion from a simple cross-tabulation showing three rows of data by 12 columns is difficult. Compare those three rows of data to the visualization below it: it's instantly obvious that domestic sales are climbing while international sales are flat.

Unprecedented advances have been made during the past several years in the application of visual analysis. We now have a much better understanding of how we humans visually make sense of the world around us. We have applied that knowledge to create sophisticated technology that enables people to visualize and understand the data that are swamping our world. Many researchers are now convinced that, as Colin Ware put it so aptly in his 2004 book *Information Visualization*, "At higher levels of processing, perception and cognition are closely interrelated, which is the reason why the words 'understanding' and 'seeing' are synonymous". In conversations with others, how often do each of us say, "I see" when we mean "I understand"?

Traditional 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 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-practice applications for data visualization and visual analysis. It gives users flexible, easy





ways to display data in whatever format makes the patterns most visible. By this we mean the process of working with data is directly live with the data itself – not an abstraction of the data. Working with Tableau is a natural process of interacting with a visual representation of your data to perform the analysis.

Tableau allows analytical frameworks to be manipulated. It allows an analyst to gain insight by changing perspectives and seeing information as data are compared differently. Teradata provides the computing capability and architecture that allows massive amounts of data or summaries of those data to be delivered to changing analytic contexts efficiently. Thus, the analyst does not lose perspective as she mentally navigates from one question to the next. Teradata allows data to be interactively manipulated as Tableau allows analysts to experiment with analytical frameworks.

Analysts are able to use Tableau for visual comparisons because Teradata enables not only conventional multi-dimensional online analytical processing (MOLAP) where the data are pre-aggregated into cubes; it also enables multi-dimensional visual analysis which does not require cube aggregation (but can work with cubes). While Tableau allows an analyst to manipulate analytical frameworks, a Teradata Database provides the massively parallel processing (MPP) database power required by these analytic sessions. A Teradata Database enables comparisons to be seen in real time.

A Teradata database can provide the speed needed for ad hoc, interactive data exploration. Teradata is legendary for the volume of data and high performance query engine – the perfect complement to an interactive exploration journey.

Ad-hoc exploration

Traditional applications deliver a fixed set of data, intended to help answer a fixed set of questions that were defined at a given point in time. But business needs and conditions are in constant flux. This leads to an ever-changing set of challenges and opportunities. In turn, users will always need to explore enterprise data in new ways. They cannot anticipate the questions they will need to ask, and each question answered can lead to more questions. A fixed set of data can never give users the flexibility to ask whichever questions they need for exploring information, refining their understanding as they go and asking further questions.

Tableau was designed from the ground up to solve this problem with ad-hoc, on-demand, interactive exploration.

But the data exploration experience should be enhanced by a fast database. Speed-of-thought analysis, iteratively asking a question, studying the answer, and asking another question is the typical work flow a knowledge worker cherishes. Slow database response times, disruptive context switching between systems or tools, convoluted steps to write queries or import data all prevent the flow of the analyst.

A Teradata Database can provide the speed needed for ad hoc, interactive data exploration while Tableau provides the analytical interface. Teradata Database, a

high-performing query engine, is legendary for the volumes of data it can process – making it the perfect complement to an interactive exploration journey in Tableau. A active data warehouse based on Teradata, where the organization’s operational data are immediately available to the analytic process enabled by Tableau, further enhances the business analyst’s experience.

The underlying size of the database 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. Modify the picture to fine-tune the question and the system automatically determines what queries to run to create the new picture. People move seamlessly through analysis simply by interacting with pictures of data.

The queries that Tableau automatically generates behind the scenes are optimized to support the visual being created. To do this, Tableau generates sophisticated requests to Teradata Database. The analytic power of Teradata Database is especially well-suited to rapidly and efficiently answer these queries.

Tableau automatically requests from Teradata Database the minimum set of data needed to answer each question. This includes asking Teradata Database to perform the aggregation, filtering, and calculations of the data requested. Some of the SQL concepts used include derived tables, the use of temporary tables, leveraging the built-in Teradata functions and SQL that is quite happily further optimized by the Teradata optimizer, including the use of Aggregate Join Indexes (AJIs), partitioning and other general optimization techniques with which most Teradata DBAs are familiar.

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

The Enterprise Data Warehouse

Tableau was designed to work with the modern database and data warehouse concepts. Tableau is an interactive analytic application that has a rich conversation with your database. Tableau flourishes when this database is fast, large, and has a variety of data subjects stored in it. These are precisely the characteristics of an enterprise data warehouse EDW based on Teradata technology.

Tableau accesses the databases where they live – in the Teradata data warehouse. Tableau is bound by all the security constraints defined at the database level, which helps organizations be sure that access is properly controlled. Contrast this to the design of many other BI solutions which require the data to be moved out of Teradata Database and stored in their own mid-tier or other data structures. This has the the potential to create silos of special purpose, and possibly incorrectly managed data stores.

Because Tableau takes a subject-oriented view of the data warehouse, it includes only the data tables that are relevant to the subject at hand as part of a

Tableau is an interactive analytic applicaiton that has a rich conversation with your database. Tableau flourishes when this database is fast, large, and has a variety of data subjects stored in. These are precisely the characteristics of Teradata EDWs.



The combination of Tableau and Teradata results in a high-performance data access system that is a key enabler of data understanding.

connection. This is a much preferred model by the business users, as they have access to just the relevant data items in the subject they elect to work with. Contrasted with the 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, well understood by database DBAs and appreciated by end users.

Tableau customers include Microsoft®, Google, eBay, Wal-Mart, Merck, Hewlett-Packard, Countrywide, Fidelity, GE, Charles Schwab, Harvard University, and the U.S. Government. Norfolk Southern uses Tableau running on Teradata Database in its AccessNS system, supporting 22,000 customers; Wells Fargo Bank uses Tableau running on Teradata Database as well.

Summary

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

For Teradata customers, Tableau helps people see and understand data faster and more easily than ever before. The visual analysis process enables the business user to be more informed, more inquisitive, and make smarter, faster decisions. Teradata is the respected industry-leader in enterprise data warehouses. Teradata Database has been proven time and time again to house the largest amounts of data yet still provide speed-of-thought query responses to enable business applications. The combination of Tableau and Teradata results in a high-performance data access system that is a key enabler of data understanding.

Frequently Asked Questions

How does Tableau connect to Teradata?

Tableau connects directly with the native Teradata ODBC driver. You can download the ODBC driver from the Teradata website at <http://www.teradata.com/DownloadCenter/>

Do I need to do anything special to my database for Tableau to access it?

No. Tableau will access any data that reside in Teradata 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 Teradata applications and provides security clearance before it is granted access to any tables in Teradata Database. Getting “connected” to an enterprise data warehouse based on Teradata 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 Teradata tables. In fact, if your EDW is like most Teradata instances, 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 within Tableau so that users can rapidly see their data the way that you, the DBA, wants them to see it.

Without a semantic layer, how does Tableau differentiate dimensions and measures?

During its initial read of the Teradata 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. The important distinction in Tableau is that the value of measures will be dependent on the dimensions selected for the visual.

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 Teradata DBA or at least with the Teradata 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 of the other 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 visuals, Tableau provides the capability to display aliases for database terms. These aliases appear only on Tableau visuals.

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 to all visuals. However, decisions with regard to relevance are made by the Tableau user. After initial presentations of data and creation of initial visuals it may be determined that some data will never be relevant to a particular analysis of business process. Should that be the case, a Tableau user can consult with the Teradata DBA to

Interactive data visualization can help to communicate this even 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.



have relevant views of data presented to Tableau based on User Name and Password and/or have separate databases, within a Teradata Database, created. In addition, Tableau itself provides the capability to join, group and arrange data so that only relevant data are more readily visible to a Tableau analyst.

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 Teradata 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.

Can Tableau work with a join index?

Yes, Tableau can work with an AJI just as if it's any other table, index, view, stored procedure or user-defined function (UDF) in Teradata Database. Tableau behaves as a good citizen of Teradata Database and presents SQL like any other application. In addition Tableau presents SQL to Teradata Database to take full advantage of the Teradata Optimizer and Workload Manager. This means that Teradata DBAs can collect statistics, with regard to any SQL submitted by a Tableau application as if it were any Teradata application. Based on the DBA's analysis of the performance of Tableau applications and statistics collected by the Workload Manager, a Teradata DBA can take any and all performance actions that might be taken for any application requesting Teradata services, including the establishment of AJIs.

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. This is one of the reasons that Teradata and Tableau work so well together. Teradata Database excels at large data volumes and still retaining fast query response times. Tableau's design is such that it doesn't take data out of Teradata Database and host it somewhere else. It accesses data directly from Teradata Database.

Tableau is a visual analysis application, so the individual visuals that are created in Tableau generally retrieve a relatively smaller number of rows from the database, automatically building the queries to have Teradata 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 Teradata Database.



Does Tableau access the Teradata SQL Query Optimizer?

Yes. The Teradata Optimizer is the most robust in the industry. It is parallel-aware meaning it has knowledge of distribution of data on nodes as well as utilization of computing resources. It manages mixed query workloads and determines the least expensive plan to process them and responds to them in parallel. The SQL created by Tableau looks to the Teradata optimizer like any other request for service and is processed with the same capabilities as any other SQL.

Does Tableau take advantage of the Teradata Workload Manager?

Just like the Teradata optimizer, the Teradata Workload Manager sees all SQL created by Tableau as standard SQL and processes it accordingly.

How does Tableau prevent queries that would “cripple” the data warehouse?

Because Tableau is a good Teradata citizen, it behaves like all Teradata applications and takes full advantage of Teradata services, both those provided by the technology and those provided by Teradata DBAs. Tableau abides by the Teradata concept of strategic vs. tactical queries where the DBA can schedule certain types of queries to run based on resources required. The DBA literally can specify and take charge of when a query is allowed to run. With Teradata’s governing system called the workload manager, Teradata Database will govern exactly how data is queried, run and returned and Tableau will abide by that.

There is no substitute for performance testing and ongoing collection of statistics and monitoring to ensure the behavior and performance of any Teradata 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 visual 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 Teradata DBA practices and the natural constraints of visualizing data at any moment in time, Tableau’s behavior with Teradata Database is generally exceptional.

Is Tableau secure – what data can end users access?

Tableau requires a username and password to the Teradata Database to establish a connection. This is a standard Teradata username which can have all the normal Teradata security constructs applied. 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 Teradata 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 the Teradata 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 Teradata 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 Teradata DBA as part of view definitions is automatically inherited by Tableau. When using Tableau Server in conjunction with Teradata Database, additional row level data constraints may be specified.

Can I work disconnected from my Teradata 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 line, simply select a menu item to return your workbook to querying the database live, 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 Teradata username?

The desktop version of Tableau does require each individual to have a database username and password.

My DBA created some AJI's and told me that makes queries run fast. Do I need to change my Tableau workbooks to use them?

No. Tableau will inherit the benefit of these automatically.