SPECIAL EDITION

TERADATA®

Magazine

ADVANCING the Data Warehouse

TERADATA Magazine

TABLE OF CONTENTS

SPECIAL EDITION

ADVANCING the Data Warehouse

by Imad Birouty

The Teradata® Database pioneers cutting-edge technology innovations. This collection of articles demonstrates how Teradata has been at the forefront of data warehousing, advancing it as a strategic business tool.

Advancing the Data Warehouse

The Teradata® Database pioneers cutting-edge technology innovations to deliver more business value.

by Imad Birouty

Since its inception, the data warehouse has been a positive disruptive force. Never before had organizations been so well-equipped to make insight-rich business decisions.

In 1984 Teradata delivered the first database appliance built specifically for data warehousing. Its powerful, cutting-edge software and hardware technology pioneered the realm of decision support. Naturally, as the need for business analytics grew, so did the need to advance early technologies.

The high level timeline above demonstrates that through its powerful new data warehouse technologies and industry-first innovations Teradata has continually pushed the leading edge in the drive to help companies address their ever-widening range of business and data requirements.

An Asset for All

The data warehouse used to be an asset reserved for a few analysts that ran long, complex queries or generated canned reports for others to consume. Today, it's an integral part of daily operations and a valuable business asset, for everyone.

- **Business analysts** use the data warehouse for ad-hoc analysis and self-service rapid testing of new data and business theories.
- **Application developers** use it as a source of data and as a scalable processing engine. They need an interface that will allow them to use the tools they are familiar with, such as an Eclipse development environment, or REST interface for driverless applications.
- **Data scientists** specialize in advanced analytics. They bring a combination of business knowledge and technical skills to discover new insights. They use languages such as Python, Perl, or Ruby scripts in addition to SQL. Performance is paramount as time is critical for their analysis. The faster they can test their theories, the more theories they can test.

Mission-Critical

Many data warehouses have become an integral part of an organization's daily operations making it a mission-critical system. This means that high availability features are an essential part of the data warehouse, both in hardware and software features. To deliver the highest levels of availability while protecting against disaster situations, a second system that is geographically separated is needed. The Teradata Dual Active solution provides both high availability as well as disaster recovery, while delivering the full performance value of both systems; maximizing ROI on the data warehouse investment.

Flexibility is Essential

IT organizations are under a lot of pressure. First, they must meet the growing analytical needs of the business while maintaining or reducing their budgets and controlling the complexity of their environment as well the efforts to run it. Those conditions can be met by running multiple data warehouses on a single platform.

In addition, IT is also being asked to reduce capital expenditures (CapEx) in favor of higher operating expenses (OpEx). By outsourcing the daily operations of platform maintenance to Teradata Cloud, organizations can shift expenses and allow IT to focus on their core competencies instead of environment management.

Not all jobs in the data warehouse are created equal so service level agreements (SLAs) can vary greatly. Some workloads require low SLAs while others may demand very high performance and guaranteed response times while others are somewhere in between. In-memory computing will deliver on very stringent SLAs but most data warehouses are too large to fit into memory. Teradata Intelligent Memory is an integral part of the Teradata Database that provides an integrated, high-performance, scalable, and economic in-memory solution that overcomes the size challenge.

The Analytical Ecosystem

As the source for enterprise analytics, users depend on the data warehouse for access to the information they need to support their decisions.

Over time the sources, volumes and types of data have changed. The hardware and software used to store and analyze it have changed, too. And it will all continue to change.

The Teradata Database has evolved in not only storing these data types, but enabling native analytics on this data along with all other data in the warehouse. Teradata QueryGrid takes this one step further. It gives users self-service access to data and analytics across multiple heterogeneous systems. Plus, it enables users from across the organization to access and analyze the data they need, including Apache™ Hadoop®.

Continuing the Innovation

As evidenced by the collection of articles in this special edition of *Teradata Magazine*, Teradata has been at the forefront of data warehouse technology, advancing it as a strategic business tool. Going forward, the company will continue to deliver and pioneer cutting-edge technology innovations that empower customers to derive more and more value from their data. T

Imad Birouty is the Director of Teradata Product Marketing. He is responsible for Teradata software and hardware products including the Teradata Database, Teradata Platform Family, Teradata Aster Analytics, Teradata QueryGrid, and Teradata Unity.

2010
Teradata
Temporal

Keep Track

New features in Teradata 13.10 enhance support of temporal data.

BY RICHARD T. SNODGRASS

Information is *the* key asset of many companies. For most, this asset contains time-referenced data, which in the past was difficult to track and manage. Languages, such as standard SQL, have few facilities for such data. Fortunately, many new features introduced with Teradata 13.10 SQL provide temporal support.

These new SQL features make it easier to express data modifications (SQL's insert, delete and update statements) and greatly reduce the length and complexity of such modifications. As a result, developers can now more easily convert their applications to support time-varying data as well as more simply create new applications that exploit stored data about the past for deeper insight into the future.

Bitemporal Tables Supported

Take, for example, a mortgage company that is challenged with achieving high data quality on information stored concerning customers and their loans. A customer service representative (CSR) reports an error to the IT department; errors are also discovered by batch jobs producing quarterly reports. The more information the IT department has access to, the better it can analyze and correct these errors.

For this reason, it is important that changes to critical tables concerning customers and their loans be tracked. This implies that each such table has "transaction-time support" to maintain a history of changes. Since this table also needs to model changes in reality, it requires "valid-time support" to indicate when the data was considered valid. The result is a bitemporal table, reflecting these two aspects of underlying temporal

support. Teradata 13.10 supports bitemporal tables, greatly easing the development of such applications.

With bitemporal tables, IT can first determine when the erroneous data was stored (a transaction time), roll back the table to that point and examine the valid-time history. It can then determine the correct valid-time history. With that history, IT can tell the CSR what needs to be changed. Or, if the error was in the processing of a user transaction, IT may update the database manually.

Because transaction-time support is included, these changes will be logged as well, enabling someone later to see what happened when the change itself was in error. The support for valid time and transaction time permits a sophisticated analysis of the evolution of the table, with all of the data directly at hand. The alternatives—going back through paper records to reconstruct the sequence of changes that were made, or attempting to extract that sequence from backup tapes or other secondary data sources—are simply not practical in such a dramatically changing environment.

Follow the Property

A property owner table captures both the history in reality of the owner(s) of a property over time, as well as the sequence of database states, denoting the transactions applied to this table. This bitemporal table is easy to specify in Teradata SQL (note the new keywords VALIDTIME and TRANSACTIONTIME):

```
CREATE MULTISET TABLE prop_owner (
customer_number INTEGER,
property_number INTEGER,
property_VT PERIOD(DATE) NOT NULL AS VALIDTIME,
property_TT PERIOD (TIMESTAMP(6)) WITH TIME ZONE)
NOT NULL AS TRANSACTIONTIME)
PRIMARY INDEX ( property_number );
```

The valid-time timestamp is specified as having a granularity of day, as a property cannot change hands multiple times in a single day. The transaction-time timestamp is identified at a granularity of microsecond, to differentiate rapidly executing transactions.

The property number column constitutes a primary key in valid time and transaction time. Specifically, the state of the table at any day in valid time, as stored at any instant in transaction time, should include at most one row in the table for any particular property, meaning that that property has one owner at that valid time, as recorded at that transaction time. Because the table was declared to be bitemporal (via inclusion of both valid and transaction time), this temporal integrity constraint will be checked automatically by Teradata 13.10.

Let's follow the history, over valid time and transaction time, of an apartment in Boston at 123 Main St. for the month of January 2010. On Jan. 10, this apartment was purchased by Eva Nielsen. We record this information as a current valid-time, current transaction-time insertion.

When the database management system (DBMS) starts up, the default temporal qualifier is exactly that: current in valid time and current in transaction time. The English is in italics, followed by the SQL statement, using the new Teradata functionality.

Eva Nielsen (whose customer number is 145) buys the apartment at 123 Main St. in Boston (whose property number is 7797) today (which happens to be Jan. 10, 2010).

```
INSERT INTO Prop_Owner (customer_number,
property_number)
VALUES (145, 7797)
```

This information is valid starting now and was inserted now. The DBMS encodes this information using the special valid-time and transaction-time columns, all under the covers. Note that the transaction-time extent of all modifications is from "now," in this case "2010-01-10," to "until closed," which is encoded as "9999-12-31." (See table 1.) This is also the valid-time extent, given the default temporal qualifier (which can be changed by the user).

TABLE 1 BITEMPORAL STATE

CUSTOMER NUMBER	NUMBER PROPERTY	PROPERTY_VT	PROPERTY_TT
145	7797	(2010-01-10, 9999-12-31)	(2010-01-10, 9999-12-31)

Let's examine a simple update:

Today (which happens to be Jan. 15) Peter Olsen (whose customer number is 827) buys this apartment, transferring ownership from Eva to him.

```
UPDATE Prop_Owner
SET customer_number = 827
WHERE property_number = 7797
```

The figure (page 53) shows the bitemporal time diagram corresponding to this update.

The horizontal axis tracks transaction time and the vertical axis tracks valid time. Information about a row, or about multiple rows associated with a primary key value, are depicted as two-dimensional polygonal regions in the diagram. Arrows extending rightward denote "until closed" in transaction time; arrows extending upward denote "forever" in valid time. Here we have two regions, one associated with Eva Nielsen and one with Peter Olsen. The left-hand region starts at time 10 (all times are relative to January 2010, so "10" corresponds to Jan. 10, 2010) in transaction time and extends to "until closed," and begins also at time 10 in valid time and extends to "forever."

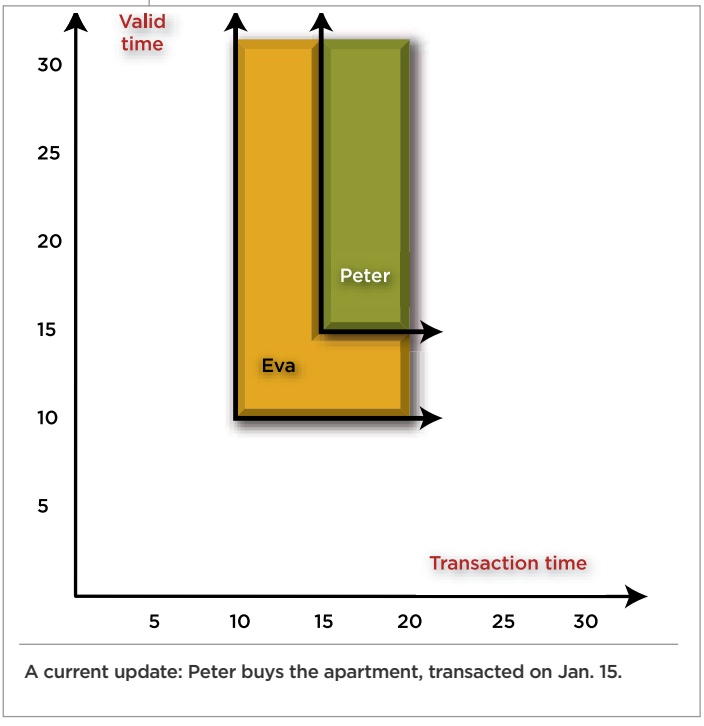
It also illustrates how this update affects the time diagram. From time 15 on, Peter owns the property; from time 10 to 15, Eva owned the property. Both regions extend to the right to "until closed." This time diagram captures two facts—Eva owning the apartment and Peter owning the apartment—each associated with a bitemporal region.

Additionally, the figure captures the evolving information content of the property owner table. Consider a transaction time-slice, which returns the valid-time history at a given transaction time. Such a time-slice can be visualized as a vertical line intersecting the x-axis at the given time.

At transaction time 5 (Jan. 5), the table has no record of the apartment being owned by anyone. At transaction time 12, the table records that the apartment was owned by Eva from Jan. 10 to "forever." If we time-traveled back to Jan. 12 and asked for the history

Many new features introduced with Teradata 13.10 SQL **provide temporal support.** These new SQL features make it **easier to express data modifications** ... and **greatly reduce the length and complexity** of such modifications.

FIGURE **Bitemporal Time Diagram**



A current update: Peter buys the apartment, transacted on Jan. 15.

of the apartment, that would be the response. We *thought* then that Eva owns the apartment, and that is what the property owner table recorded then. At transaction time 17 the table records that the apartment was owned by Eva from Jan. 10 to Jan. 15, at which time ownership transferred to Peter, who now owns it to "forever." And that is the history as best known (denoted by the right-pointing arrows). It is what we think is true about the valid-time history.

Automatic Time Handling

In addition to contending with valid time, we also must ensure that the transaction-time extent of the modification is from "now" to "until closed." One important property of tables with transaction-time support is that they are append-only.

As these tables capture the state of the stored table over time, once we have recorded what the state was at a particular time, we can't go back and change it later, because we can't change the bits stored on the disk at that prior time. The changes always accumulate in the table with transaction-time support. The practical ramification is that we never physically delete a row from such a table; only physical modifications are allowed to insert rows into the table and to change the transaction-stop time of a row from "until closed" to "now," thereby logically deleting the row.

Teradata 13.10 handles this automatically. The resulting property owner table contains three rows. A careful matching of the dates in table 2 to the time diagram will aid in understanding how a bitemporal state table encodes the regions found in the time diagram.

TABLE 2 **BITEMPORAL STATE**

CUSTOMER NUMBER	NUMBER PROPERTY	PROPERTY_VT	PROPERTY_TT
145	7797	(2010-01-10, 9999-12-31)	(2010-01-10, 2010-01-15)
145	7797	(2010-01-10, 2010-01-15)	(2010-01-15, 9999-12-31)
827	7797	(2010-01-15, 9999-12-31)	(2010-01-15, 9999-12-31)

In previous versions of the Teradata Database, all of this must be done manually. You are encouraged to implement this simple update in conventional (nontemporal) SQL. It requires a surprisingly complex series of five INSERT and UPDATE statements, 31 lines in all.

Pared Down

We examined two simple variants of temporal modifications, both current in valid and transaction time: an insertion and an update. In conventional SQL these statements can be long and complex: The worst case is one in which a non-temporal update of only a few lines is expanded to some 60 lines of SQL. All are very natural to write given the temporal extensions provided in Teradata 13.10, requiring but a few lines. T

Richard T. Snodgrass, a professor of Computer Science at the University of Arizona, has researched temporal databases for 30 years.

This article originally appeared in the Q4 2010 issue of *Teradata Magazine.*

ONLINE
Discover more on this topic by downloading "A Case Study of Temporal Data," by Richard T. Snodgrass, on **Teradata.com.**

2011 Hybrid Storage with SSD

The Heat Is On

Teradata's temperature-based data warehouse solution makes smarter decisions faster.

by Betsy Huntingdon

Enterprise leaders have an ever-expanding need to use the current business activity data provided by their integrated data warehouse to make accurate, real-time decisions. They also need long-term historical data to help identify trends and perform comprehensive analytics. Getting the best of both worlds efficiently and economically with one system requires the power of a hybrid storage solution.

The Teradata Active Enterprise Data Warehouse (EDW) delivers. It provides solid-state drives (SSDs) and hard disk drives (HDDs) fully utilizing Teradata Virtual Storage, which automatically migrates data between the two types of disks to achieve optimum performance.

Check out this "Teradata Mixed Storage - Real Value" video on YouTube to find out how this data warehouse solution keeps up with the increasing pace of data.

This solution tracks data use and intelligently moves it to the appropriate disk—"hot" data on the faster SSDs and the least used "cold" data on the slower HDDs. The result is a platform that makes smarter decisions at hyper speeds.

Now that this revolutionary system has entered the marketplace, the heat is on for organizations to integrate and smartly manage the gamut of data temperature to dramatically speed up their decision-making capabilities and boost their competitive advantage.

Delivering Business Success

SSD is an innovative technology for data warehouses. An SSD device leverages the speed of semiconductor flash memory technology and robust management logic to provide fast, immediate, direct access to hundreds of gigabytes of data. Just a small group of SSD devices is needed to supply the storage performance required by data warehouse platforms.

Traditional electromechanical HDDs are an economical way to store and analyze vast amounts of data. However, their slower retrieval speeds mean a very large number are needed to meet the performance requirements for real-time analysis.

How can integrating these storage systems help businesses succeed? Consider this example:

Company A just loaded its current week's sales figures into its mixed storage data warehouse. The data is fresh and hot, so it goes directly into the system's SSD storage, where employees throughout the enterprise can run near real-time reports and make lightning-fast decisions about what should change (in marketing, operations or elsewhere) to improve next *week's sales performance.*

Meanwhile, the system automatically migrates colder data that is weeks or months old from SSD into the larger-capacity, lower-cost HDD storage area. Employees can still access and analyze it quickly, just possibly not as quickly as the hot data. As data heats up again—such as sales figures for the same period last year—the system automatically moves that data into SSD storage so it's ready when needed.

An Optimized Solution

The Teradata Active EDW optimizes the use of the hybrid platform with the unique Teradata Virtual Storage feature that moves hot and cold data to the appropriate location based on frequency of use. (See graphic, page 19.) Data is automatically transferred and maintained in the optimal storage drive, making the system precisely suited to support current analytics and event response while enabling long-term strategic analysis and reporting. >>

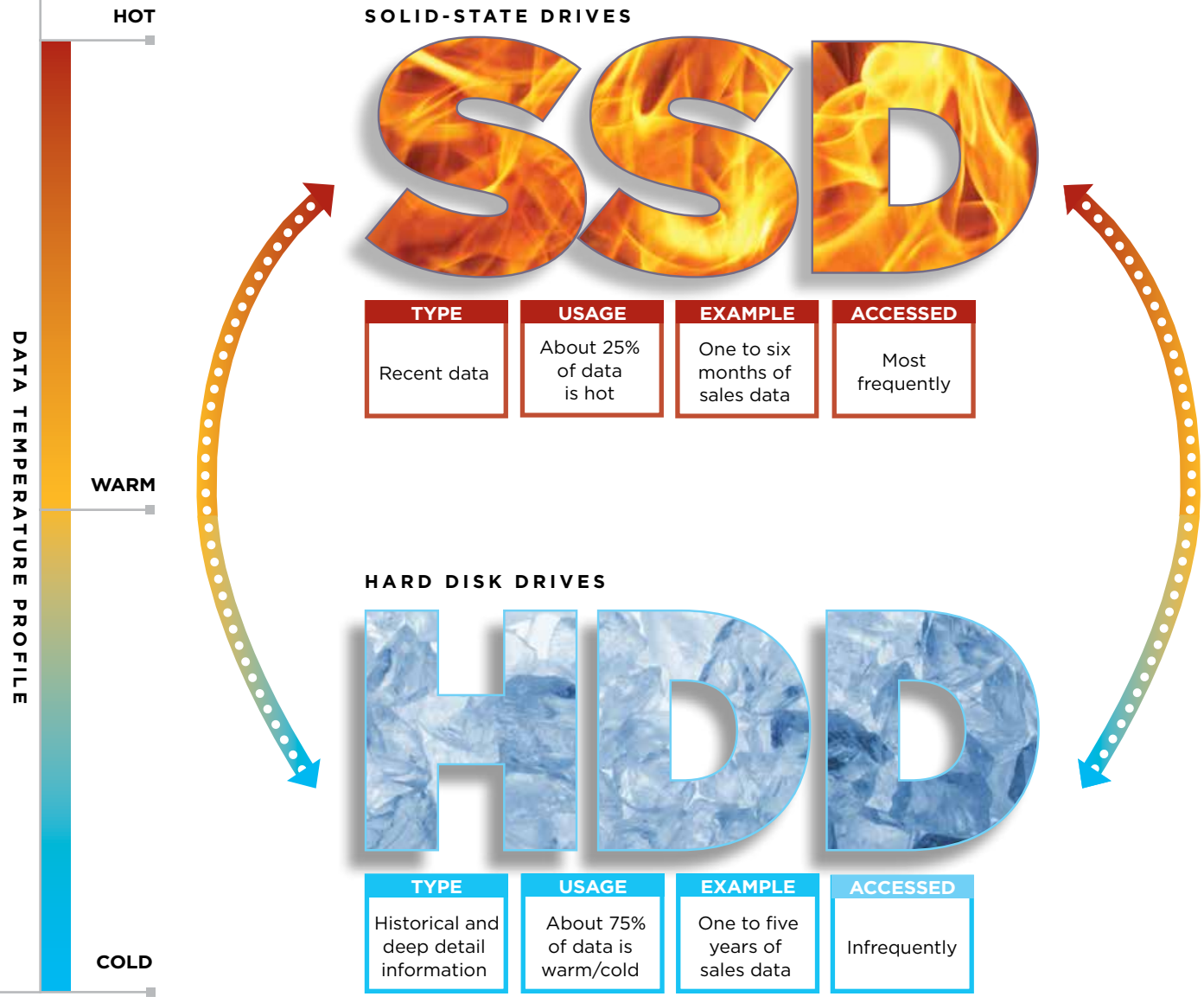
ONE DATA WAREHOUSE: TWO TYPES OF DISKS

An Industry First

This Teradata Active Enterprise Data Warehouse is the first to combine SSD and HDD technologies with the industry's only intelligent virtual storage solution that automatically migrates data between disk types for optimum performance.

Data is automatically placed on **SSD**s or **HDD**s depending on its temperature. This industry-unique feature enables unassisted hot and cold data placement between the storage types.

- Optimum use of storage resources
- Makes smarter decisions at hyper speeds
- Gives companies a competitive advantage
- No database administrator guidance required
- Boosts efficiency, profitability and business agility
- Uses 75% less energy
- Has a 75% smaller data center footprint



Any business that needs to make fast, short-term decisions in addition to conducting longer-range reporting and analysis can benefit from this mixed storage solution.

Get More From Your Data Warehouse

Teradata's mixed storage offering empowers organizations to stay ahead of their competitors by giving leaders better information faster. It also enables them to grow the use of their data warehouse across an expanding spectrum of applications and functions by applying increased performance to integrated data.

A data warehouse based on hybrid storage can be used to:

Populate executive dashboards. The storage solution is ideal for informing increasingly popular business intelligence (BI) dashboards using up-to-the-minute data to display accurate current-state snapshots. The data can also be used to display relevant long-term projections and strategic analyses.

Bolster cybersecurity. Hours of Internet packet data can be quickly loaded to solid-state drives (SSDs) to help spot malicious activity such as network attacks or spam outbreaks. Older data can be migrated to hard disk drives (HDDs) for trend analysis and long-term attack identification and prevention.

Boost manufacturing operations. Companies can do a better job facilitating current inventory planning and troubleshooting. Long-term yield and quality can be improved with quick reaction to process anomalies.

Aid retail planning and strategy. Weekends are often the highest-grossing period for retail organizations. With this storage system, retail leaders can analyze weekend sales data on "manic Monday," in time to adjust prices and marketing messages for the current week and upcoming weekend. Long-term data could help these same organizations spot trends, target marketing campaigns and inform buying decisions.

Teradata Virtual Storage helps ensure that data is placed on the most appropriate disk type. But data movement doesn't stop there. The system migrates it as needed: As cooler data "heats" up in use, it is automatically moved to the SSD storage area. As it cools, it is transferred to the HDD storage area as "hot" data space is needed.

Teradata analysts conducted extensive research to determine the actual usage pattern of data within the warehouse. They determined that although data is constantly changing, at any given time, typically only about 25% of an organization's data is hot, and the remainder is warm or cold. This information is used to determine the optimum ratio of SSD to HDD storage capacity, thereby enabling a balance of both performance and data capacity for the most economical Teradata Active EDW solution.

Short- and Long-Term Benefits

Whether organizations need to make immediate operational decisions or plan an effective long-term strategy, the Teradata Active EDW offers "temperature-based" information management that improves productivity as it speeds up reporting and ad hoc analysis for power and casual users alike.

This helps boost efficiency, profitability and business agility, even when the competitive heat is on. Any business that needs to make fast, short-term decisions in addition to conducting longer-range reporting and analysis can benefit from this mixed storage solution. T

Betsy Huntingdon is the product marketing manager for Teradata hardware platforms.

This article originally appeared in the Q3 2011 issue of *Teradata Magazine.*

Found in Translation

Teradata Plug-in for Eclipse lets SQL developers speak another language.

Writing a SQL algorithm and then turning it into a Web service can be as challenging as trying to speak German when you are fluent only in French. The Teradata Plug-in for Eclipse lets SQL developers prototype working examples of Web-based services using their native language of SQL. It even automatically generates code that is acceptable to application developers.

Mike Coutts, CTO for Client Software Engineering at Teradata, explains how the free plug-in works and why SQL developers should welcome it into their toolkits.

You're responsible for directing the development of Teradata's next generation of tools. What does that encompass?

We develop all of the pieces that lie between the database and the applications—utilities for load and unload, database connectivity, client access, dual-active processing and single operational view, as well as tooling, like the integrated development environment (IDE).

I think about it as a continuum, with systems management at one end and development at the other. Our current focus is on giving customers the best operational tools for monitoring and managing the Teradata system with products like the Viewpoint and Ecosystem Managers, along with data-management tools like Data Mover and BAR [backup archive and restore]. At the developer end of this continuum, the Teradata Plug-in for Eclipse helps both Teradata SQL and application developers be more productive.

Can you explain how Eclipse and the new plug-in work?

Eclipse is an open-source environment into which the Teradata plug-in fits. Basically, we expanded the Eclipse Data Tools Project [DTP] to include Teradata-specific features. In this latest release [version 13.10], the tool allows developers to access tables and macros, create objects, grant and revoke user rights, create user-defined functions [UDFs], define stored procedures, etc.

Many traditional tools are SQL-focused. While we support that, we're moving into embedded features and *how* you use SQL in real applications. With the Eclipse plug-in, we can take a SQL statement and generate the Java code rather than make the developer write it. This allows faster development and better collaboration among SQL developers and their application colleagues.

Can you describe typical development activities enabled by the Teradata Plug-in for Eclipse?

Developers create a connection profile for their Teradata user account, which allows them to browse the database, tables, views, macros, etc., as a tree structure, much like the folder explorer in Windows 7. They just right-click on any element to get a menu of relevant options.

The plug-in includes an array of wizards designed to help create various DDL [data definition language] and DML [data manipulation language] artifacts as well as create the structure for embedded functionality like Java Stored Procedures and UDFs. Developers can, of course, just open up a SQL Scratchpad document, point it at the connection profile and start typing SQL with all the usual content-assist features like SQL parsing and prompting.

Then why not just use existing tools like Teradata Administrator, SQL Assistant or even BTEQ?

First, Eclipse-based tools are flexible and easy to use. Another advantage is multi-platform support (Windows, Linux, Mac) so I'm not tied down to one OS. Eclipse also supports version control plug-ins and team development—agile or otherwise. I consider everything we develop to be source code, from Java and XML to SQL and build scripts. Internally, Teradata Labs uses Eclipse and the Teradata plug-in extensively. It helps with consistent coding, team-based development and version control.

You touched on application development earlier. Can you elaborate?

While most Teradata developers are comfortable using SQL, it only delivers relational data sets. In the application world, programmers think in terms of object-oriented paradigms. Crossing between these two paradigms is best achieved with the help of an object relational mapping tool or infrastructure. We

> “[Teradata Plug-in for Eclipse] not only assists SQL developers in writing their algorithms, but it also enables them to **prototype in Java without having to actually write in Java.**”
>
> —Mike Coutts, CTO for Client Software Engineering, Teradata

Three Things You Didn’t Know About Mike Coutts

1. Part of the Clan Farquharson, his formal dress is this Scottish kilt. “Kilts seem to be making a comeback.”
2. He rides his Kawasaki ZX-636 on the track. “This lets you do the things you should never do on the road in a protective environment.”
3. He wakeboards on the Colorado River and snowboards in the California mountains. “I’ve also been able to snowboard in the Alps, Canada and New Zealand.”

14 REASONS TO GET PLUGGED IN

The free Teradata Plug-in for Eclipse offers a variety of benefits to developers, including:

1. **Teradata-focused** SQL development
2. **Automatic code** generation
3. **Fast prototyping** of Web-based services
4. **Embedded processing** development
5. **SQL-level access** to the underlying Teradata Database
6. **Easier collaboration** among SQL and application developers
7. **Wizards that automate** enterprise development
8. **Java user-defined** function phase templates
9. **Result Set Viewer**
10. **SQL History View**
11. **Support for Data Tools** Project 1.8 and Eclipse Helios 3.6
12. **Multi-platform support** (Windows, Linux, Mac)
13. **Teradata support** (for subscribers)
14. **Teradata-validated** open source

To download the Teradata Plug-in for Eclipse, visit **Downloads.Teradata.com** and select Tools.

have developed a number of wizards that make this process a simple form-filling exercise rather than a repetitive coding exercise. These wizards allow a SQL developer to generate the Java Data Access Objects [DAOs] associated with their queries, which application developers can then integrate into larger applications.

We realized that during the process of creating the DAO, we could actually derive sufficient information from the query and the underlying database to create a Web Service Definition Language [WSDL] and schema—and, therefore, a Simple Object Access Protocol [SOAP] Web service representation of the query. This series of interconnected wizards allows a SQL developer to right-click the SQL, fill in some forms and generate part of an enterprise application.

Can you give us an example of how this process works?

Consider a banking environment where a table in the database is kept up-to-date with live credit-score information. I would want to be able to access that information in various ways and use it in my applications.

For example, call-center representatives want to know if the customer on the phone has a good or bad credit score so they can determine which products and services to offer. At the ATM, the system needs to know whether the customer has a good credit score before it extends overdraft protection. With Internet banking, we want to show customers relevant banner ads when they log in.

Those three application environments often use different coding languages. Now, I can craft a single SQL statement that can gather the required information, create a Web service definition and make it available to other developers. I can prototype the SQL statement and quickly craft a Web service that I can test and hand off to an application developer.

We want to deliver tactical access to the information in the database without having to write three sets of code. This tool not only assists SQL developers in writing their algorithms, but it also enables them to prototype in Java without having to actually write in Java. The plug-in generates all the code.

That sounds like it will help SQL developers become more agile and integrated with the applications, and save a lot of time. How can readers learn more?

I wrote a series of online articles called the "Friday Night Project" on the Teradata Developer Exchange. We took some of the best practices from our own enterprise applications and illustrated them within a sample application to help our customers connect the Java/HTML application to the data warehouse. A lot of what was hand-coded in that exercise has now been embedded within the plug-in's various wizards to make development even easier. T

This article originally appeared in the Q2 2011 issue of *Teradata Magazine*.

ONLINE
To access the "Friday Night Project" article series, learn more about best practices for integrating Teradata products into the larger IT ecosystem and join discussions regarding Teradata tools, visit **Developer.Teradata.com.**

2011
Hybrid
Row/
Column

Rows vs. Columns: Why Not Both?

Teradata Columnar allows row, column or hybrid reads for ultra-efficient queries.

by Paul Sinclair

Teradata Columnar, introduced in Teradata Database 14.0, offers a new approach for organizing the data of user-defined tables and join indexes on disk. Teradata's unique implementation of columnar has the ability to partition a table or join index by column. This column partitioning feature can be used alone in a single-level partitioning definition or with row partitioning, enabling queries to access data by column, row or both.

The hybrid row and column feature provides a new paradigm for storing data that changes the cost-benefit trade-offs of physical database design choices and their combinations. Teradata Columnar also benefits businesses by reducing I/O for certain queries while simultaneously decreasing space usage.

Divide and Conquer

Row partitioning in the Teradata Database allows users to group rows horizontally. Each row partition contains a subset of the total rows in a table, such as one day of transaction data. Users can define one or more levels and define a partitioning expression for each level. If they have a query that specifies a value or a range of values for the partitioning column, fewer rows need to be accessed, compared with a full table scan.

With column partitioning, each column or groups of columns in a table will become a partition containing the values of that specific grouping. This is a simple approach because there is no need to define partitioning expressions, and determining partition elimination is very easy. Here is an example that creates a column-partitioned table that vertically partitions the data:

```
CREATE TABLE Sales_CP (
   TxnNo     INTEGER,
   TxnDate   DATE,
   ItemNo    INTEGER,
   Quantity  INTEGER )
PARTITION BY COLUMN,
UNIQUE INDEX (TxnNo);
```

A primary index is not specified, so this is a no primary index (NoPI) table. Each column is its own partition, so every partition value is just a value of that individual column.

Efficiencies of Column Partitioning

The key benefit in defining row partitioning for a table is when queries access a subset of rows based on constraints on the partitioning columns. Using column partitioning improves the performance of queries accessing a subset of the columns, either for predicates or projections. Because sets of columns can be stored in separate partitions, only those containing the columns referenced by the query need to be scanned.

The advantages of column and row partitioning can be combined in Teradata Columnar, further reducing I/O. Fewer data

FIGURE 1 Column and Row Partitioning

TxnNo	TxnDate	ItemNo	Quantity	
100	05-29-2011	756	1	One row partition
100	05-29-2011	124	3	
290	05-30-2011	437	1	One row partition
450	05-30-2011	110	1	
530	05-30-2011	815	2	
One column partition	One column partition	One column partition	One column partition	

Partitioning by rows and by columns results in less data being accessed, which means faster query times and reduced I/O.

blocks need to be read, since more information is packed into the blocks. For example, the following shows how to get a list of items sold on May 29, 2011:

```
CREATE TABLE Sales_CPRP (
   TxnNo     INTEGER,
   TxnDate   DATE,
   ItemNo    INTEGER,
   Quantity  INTEGER )
 PARTITION BY (
    COLUMN,
    RANGE_N(TxnDate
       BETWEEN DATE '2011-01-01'
         AND DATE '2011-12-31'
         EACH INTERVAL '1' DAY) ),
 UNIQUE INDEX (TxnNo);
```

With both column and row partitioning defined on a table, the query needs to access only the column partitions containing essential items. (See figure 1.)

Another way to look at the advantages of partitioning is to contrast the data accessed when different types of partitioning are defined. For example, this table definition has various physical database designs:

```
CREATE TABLE mytable (
  A INT, B INT, C CHAR(100),D INT, E INT,
  F INT, G INT, H INT, I INT, J INT, K INT,
  L INT );
```

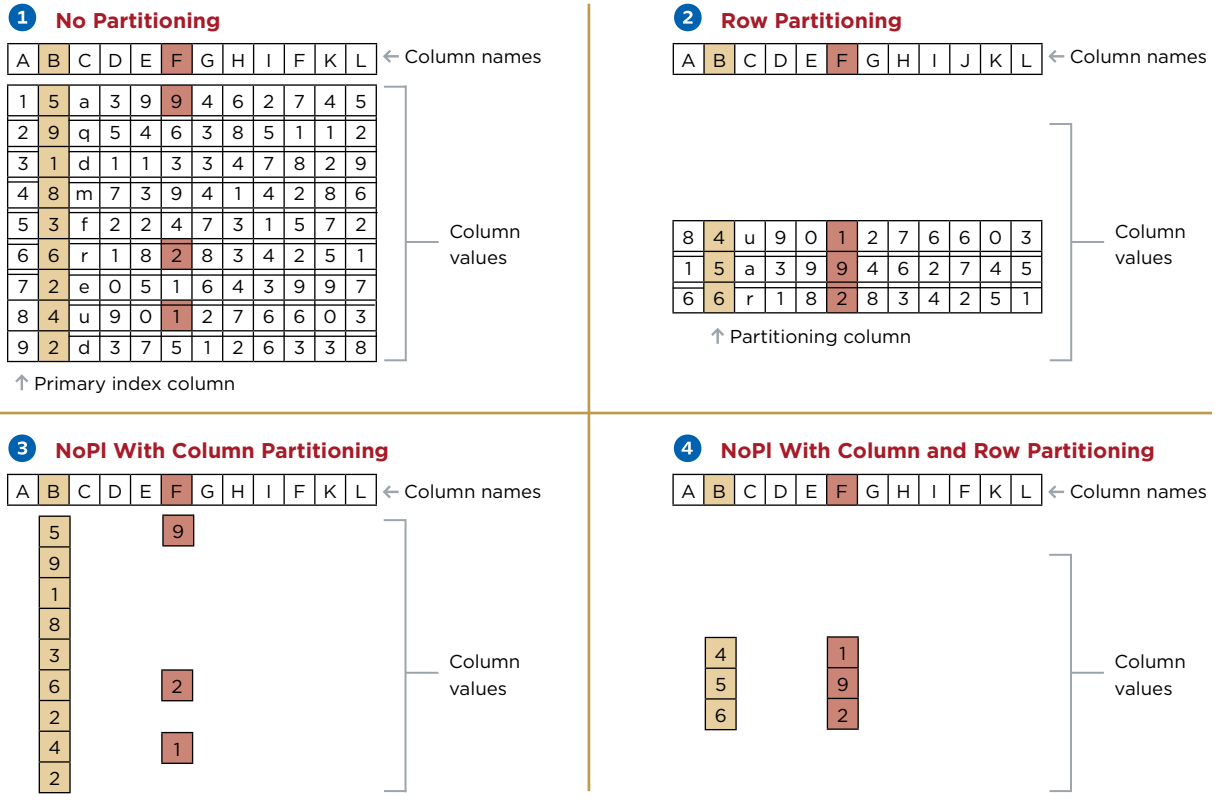
The query below is based on the previous table called mytable:

```
SELECT SUM(F) FROM mytable WHERE
B BETWEEN 4 AND 7;
```

In figure 2 on page 17 only columns F and B are referenced by the query, even though the table has 12 columns. The figure illustrates the data that has to be accessed when there is:

1} no partitioning
2} row partitioning on column B
3} column partitioning
4} both column and row partitioning

FIGURE 2 **Comparison of Physical Database Design Choices**



When a table has no partitioning, as in No. 1, all data is accessed. By combining column and row partitioning, as shown in No. 4, only the necessary data is accessed.

Column-Partitioned Characteristics

A table or join index that is partitioned by column has several key features:

- Primary index, primary AMP index, or no primary index
- Column partitions composed of single or multiple columns
- Column partitions, each of which usually contains multiple physical rows
- Columnar storage and autocompression implemented with a new physical row format column called a container
- Row-storage implementation that may have physical rows with row format called a subrow

If the table has 4 million rows, the query reads approximately this many data blocks:

- 9,987 with no partitioning
- 4,529 with row partitioning
- 281 with column partitioning
- 171 with column and row partitioning

Decreased I/O usually comes with higher CPU usage. Since I/O is often relatively expensive compared with CPU, and CPU is getting faster at a much higher rate than I/O, this is usually a reasonable trade-off.

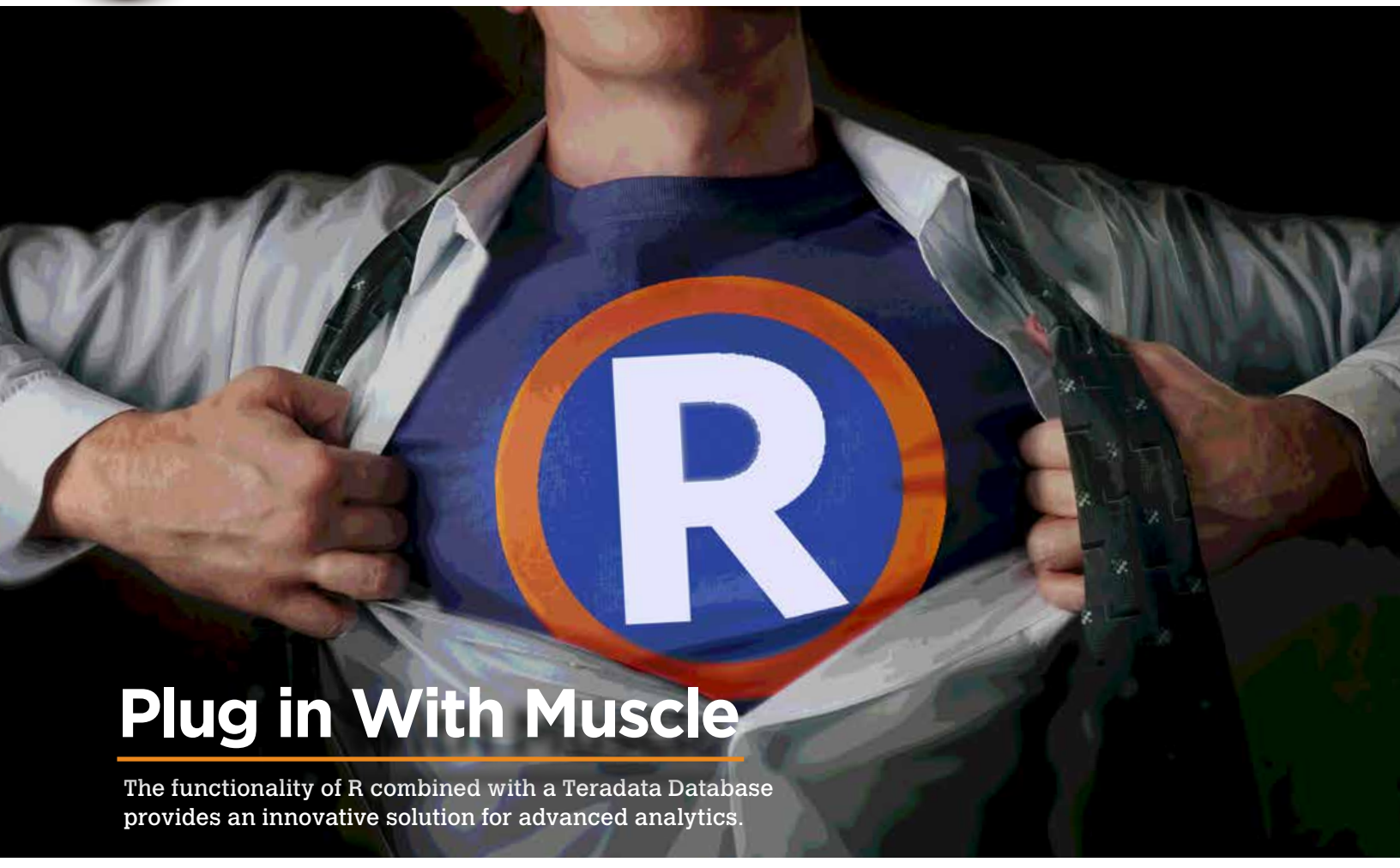
Fewer Touch Points

Teradata Columnar offers the ability to partition a table or join index vertically by column, horizontally by row or a hybrid of both. To eliminate the need to access all data in all rows and columns, Teradata Columnar breaks the table into smaller sections so only the individual column data for the proper rows is scanned to complete a query. T

Paul Sinclair is an architect at Teradata Labs. He designed partitioning features that are the basis for Teradata Columnar.

This article originally appeared in the Q2 2012 issue of *Teradata Magazine.*

2011
R Integration



Plug in With Muscle

The functionality of R combined with a Teradata Database provides an innovative solution for advanced analytics.

by Karl Krycha

The "R project," commonly known as "R," is a powerful solution to implement analytic methods for business applications such as churn, cross-selling and credit risk analysis. With this solution, users can easily enact all of the required steps to prepare, run and interpret statistical analysis.

Teradata supports R as a cost-effective option for companies. Like Linux, Apache and Firefox, it is an open-source program—free for anyone to use and modify—encouraging organizations to explore analytic techniques and experiment with analytic applications without procurement and licensing of software.

Leverage In-Database Functions

Analysts are required to move data into the R environment, which can be a challenge depending on the volume and data source. To address this, Teradata developed an add-on that enables users to push key analytic tasks directly into the database for processing. This eliminates the need to move information from the data warehouse into an R data frame. The R add-on allows users to easily connect to the Teradata Database, establish data frames to tables within the database, and use the more than 45 in-database analytic functions callable from R. The add-on takes a unique approach to data frames by establishing a pointer (virtual table) to Teradata Database tables, which eliminates the need to move the entire table into the R environment.

{ definition }
Data frames are the R matrix or table-like structures in which the columns can be of different types and rows represent an observation. R data frames closely resemble the SAS or SPSS data set.

The add-on also provides the programmer with the opportunity to leverage the processing power of the Teradata Database with

TABLE 1

Summary of Analysis

minutes_of_use	age
Min. : 0	Min. :13.00
1st Qu.: 7467	1st Qu.:28.00
Median : 17242	Median :42.00
Mean : 1	Mean :1
3rd Qu.: 32152	3rd Qu.:57.00
Max :144157	Max :89.00

TABLE 2

Attribute Coefficients

Coefficients:	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.045e+00	5.355e-01	-1.952	0.051
minutes_of_use	7.790e-06	9.981e-06	0.781	0.435
age	9.688e-03	1.040e-02	0.932	0.352

the R interface. The advantages of using R in-database include:

- Keeping data movement to a minimum
- Supporting big data processing
- Executing R process steps in parallel

R at Work

The end-to-end process of analytical modeling starts with business specifications. The process addresses statistical data preparation, the actual modeling, and preparation of (recurrent) scoring after the modeling.

To be understandable for statistical methods, information mostly needs to be organized in a data set or matrix form. In the case of churn prediction, for example, the most basic element of information is the line level Customer Analytic Record (CAR). Data preparation behavioral details, such as the number of calls or minutes of use, are aggregated to a weekly or monthly level. Finally, information about whether the individual line has churned or not is attached. Typically, a CAR covers 100 to 300 attributes per line.

To be used by regression analysis, which is usually the preferred option for analyzing churn, a certain number of churned lines are combined with a number of lines still active in order to prepare modeling. The resulting sample of records is called an analytic data set (ADS).

An essential part of the modeling process is the preparation of the ADS. Data preparation is recommended to take place completely in-database. The best practice is to use Teradata ADS Generator. To initiate modeling with R, users can create the modeling ADS R Teradata data frame.

For example, if the model will be used for scoring, R offers predictive model markup language (PMML) to import the regression model into the Teradata Database. Other options include parameter handover using command *coefficients*, plus code parsing or scoring with R.

Get Results

This code example shows the syntax used to generate histograms (see figure) for a churn analysis and to create the results in table 1:

```
library(teradataR)
tdConnect("Teradata")
tdf <- td.data.frame("CHURN_ADS",
  "CHURN_SOURCE_DB")
td.hist(tdf, "minutes_of_use")
td.hist(tdf, "age")
summary(tdf[c("minutes_of_use",
  "age")])
```

Table 1 provides an analysis summary. This code provides the corresponding R output for the churn model:

```
A_churn_model <- glm(formula =
  churn_event ~ minutes_of_use +
  age, family = binomial, data = tdf)
summary(A_churn_model)
```

Table 2 shows the estimated regression coefficients for minutes of use and age attributes. A regression analysis typically uses many more types of attributes.

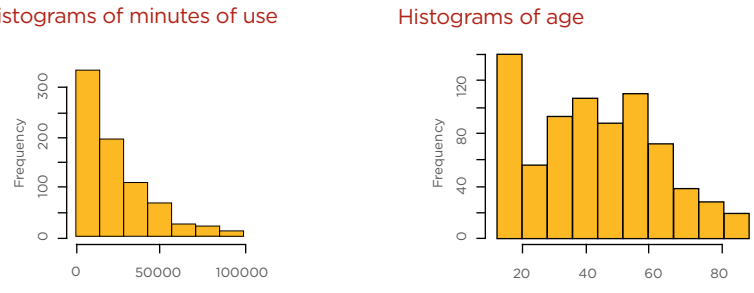
Increased Benefits, Reduced Costs

The functionality of R and its free access to numerous statistical techniques gives users of Teradata systems a powerful environment for advanced analytics. Teradata's add-on package allows users to capitalize on the benefits of R and leverage in-database processing for analytical experimentation, prototyping and development, then deploy models using commercial tools. This reduces development costs, delivers emerging analytic techniques and accelerates delivery with reduced risk. T

Karl Krycha is a managing consultant at Teradata. He heads up the Europe, Middle East and Africa Advanced Analytics Professional Services Center of Excellence.

This article originally appeared in the Q3 2011 issue of *Teradata Magazine.*

FIGURE: Histograms for a Churn Analysis



These histograms show important attributes in the analytic data set. The analysis results help identify the basic distribution of those attributes.

NetApp®

22 Year Partnership

NetApp and Teradata

- Proven storage reliability — more uptime and more business results
- High-performance storage — faster business results
- Advanced data protection and security — critical business data is available when you need it

Click for more information on the Teradata Active Enterprise Data Warehouse 6800.

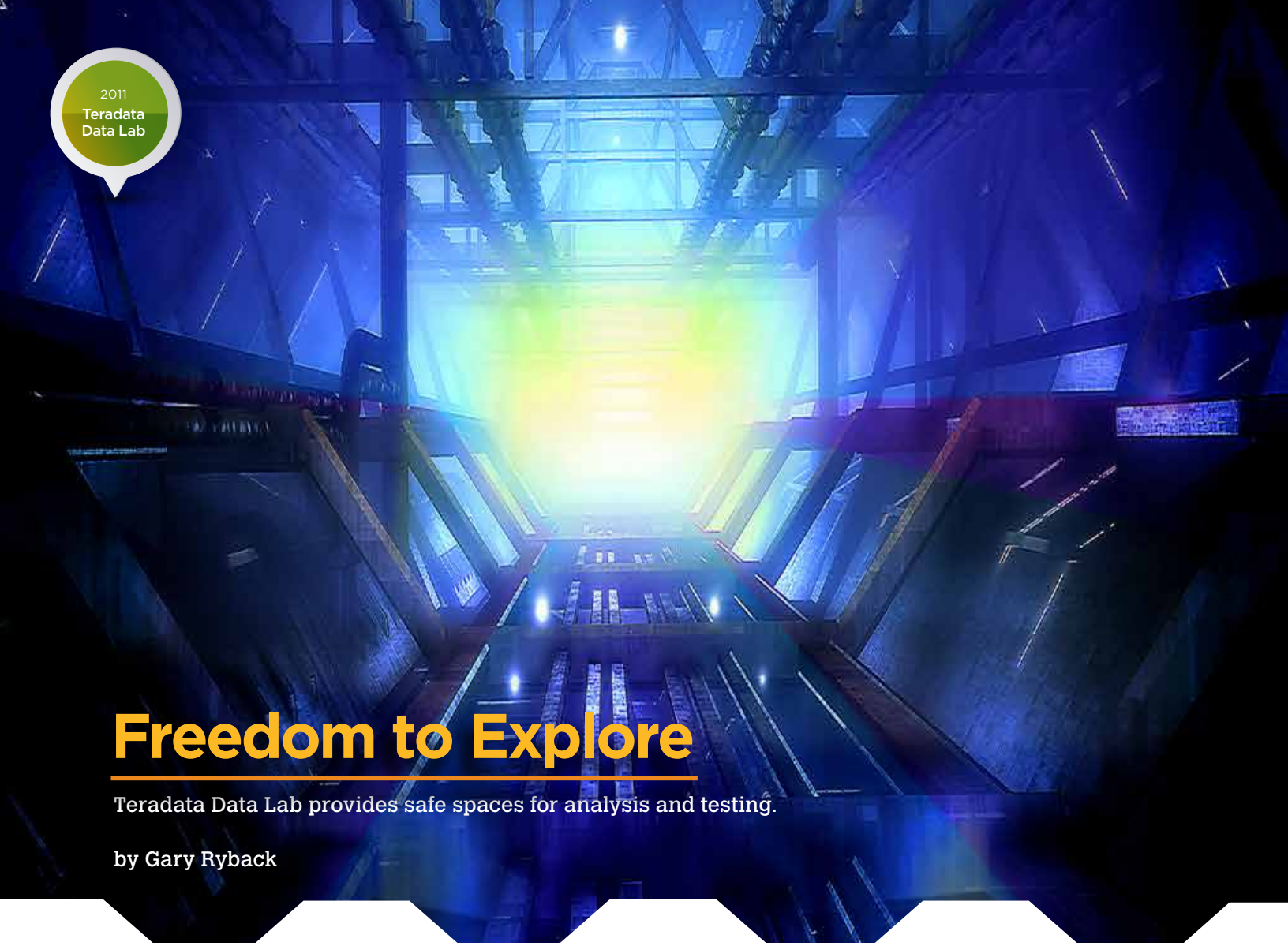
Go further, faster®

22 Years of Partnership

TERADATA

NetApp®

Teamwork • Cooperation • Innovation



Freedom to Explore

Teradata Data Lab provides safe spaces for analysis and testing.

by Gary Ryback

Any organization with a production data warehouse faces the challenge of how to allow for data proofing and new value investigations while maintaining the relationships to production data. One solution is internal "sandboxes," which are places to play, create and mold new designs.

However, bringing concept data into production often creates contention between the need for flexibility and the IT processes necessary to protect and stage data within that environment. Moreover, DBAs may be unable to take on the added responsibility of managing this additional workload, especially since it may have a negative impact on the performance and service levels of production operations.

Now there's Teradata Data Lab—a solution that allows for the exploration of new data with easy linkage to production data. It also has mechanisms for distributed automated management, provisioning and governance assurance that the production environment will be protected.

Data Lab Beneficiaries

Teradata Data Lab helps DBAs, power users and end users—everyone involved in the staging of data—determine whether it makes sense to add the information to production:

- **DBAs.** The main concerns for DBAs are the integrity and health of the production system. They also want to share the workload where it makes

sense. Data Lab provides integration aspects into Teradata workload management strategies—in particular, Teradata Active System Management. This allows the Data Lab environment to be managed appropriately through wildcard classifications. Once workload management is set up and a portion of database space is allocated, management can hand off to Data Lab owners.

- **Power users and Data Lab owners.** These people act as liaisons between DBAs and the business analysts and end users. And as the key enablers for the Data Lab architecture, they also realize the same benefits as the business analysts and end users. In many cases, power users and Data Lab owners are already acting as IT representatives for their team or group where they manage the external repositories for data proofing. With the Data Lab framework and appropriate Viewpoint portlets, these management responsibilities are simplified and more efficient with more productive results.
- **Business analysts and end users.** As users, hands-on access is enabled to explore the data and test theories providing confidence that the results will be consistent when implanted in production. Data proofing is made easier for the analysts since it can be managed through email. The Data Lab accelerates and eases the burden of the process flow to migrate new data into production information. And issues of stale data, waiting for periodic data loads from production and instability of "under the desk" infrastructures are eliminated.
- **Other beneficiaries.** Individuals responsible for project management, data design and integration find that Data Lab offers the advantages of providing more precise feedback to requirements, improved data quality results and earlier usage input. In the end, these users can make faster, more informed decisions.

Explore, Analyze and Test

Teradata Data Lab refers to workspaces in the data warehouse where users can explore and analyze new data and test theories. Users can add new data and then link it with existing data inside the warehouse with the assurance that mechanisms are in place for protecting production. Unique Teradata Viewpoint portlets, built to support the Teradata Data Lab, can assist in governing, auto-provisioning, and monitoring data exploration and proofing in production.

The Portlet Advantage

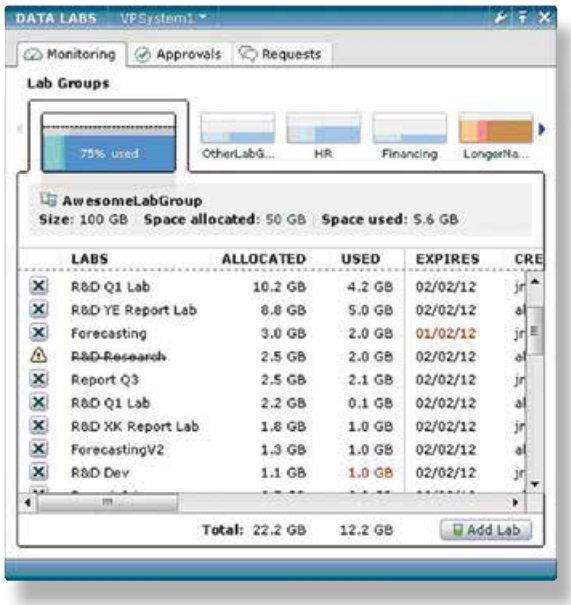
Teradata Data Lab offers unique Teradata Viewpoint portlets and capabilities. A Data Lab is used for viewing, creating or editing a lab group, which is the space in production carved out for data exploration. Within this portlet, users define the overall size of the lab group and the default size for individual lab requests. Users also establish the lab group default expiration time frame, the distributed ownership, user access, thresholds and management settings.

The Lab Group Setup portlet sets the framework for how the Data Lab area will be managed, accessed and partitioned. This portlet is the key administrative aspect of Data Lab and is tailored to DBA foundation operations. Tabs are set for owners, privacy, default users and approval permissions.

Data Labs, another Teradata Viewpoint portlet, is the workhorse of the architecture and the primary interface for the Data Lab owners, business analysts and end users. The portlet has three tabs—Monitoring, Approvals, Requests—that allow these user operations to be performed once the appropriate permissions have been granted:

- Create new lab groups
- Request increase or decrease of lab group size
- View lab group expiration and request an extension
- Request the addition or removal of a user
- Ask to assign a new owner or remove an owner
- Send other requests
- View and approve requests
- Request to promote or drop a table
- View existing tables and table locations
- View space usage

FIGURE **Data Labs Monitoring Tab**



The Monitoring tab allows users to add and remove groups and see every group in the system.

Define, Monitor and Manage

Teradata Data Lab definition, monitoring and management are supported via unique Teradata Viewpoint portlets: **> Lab Group Setup.** View, create and edit lab groups **> Data Labs.** Monitor and manage access to lab groups and/or specific data labs

The Monitoring tab lets users see all the lab groups on the system that they have been authenticated to view. (See figure.) This is where users can add and delete lab groups and view lab status.

The Approvals tab provides an easy interface for approving or denying Data Lab requests and reviewing their status. This tab is primarily used by Data Lab owners to manage incoming requests and is complemented by the underlying email notification architecture.

The third tab, Requests, is used by everyone. Here, Data Lab owners will see all requests for their particular Data Lab, and users will see their requests and the status of each one.

In addition, preferences can be set to easily identify conditions such as lab group space thresholds, aging approvals and requests, upcoming expirations, and table skews of concern. Users also have the option for automated email notifications on approaching lab expirations, lab space utilization, and status of approved or denied requests. Significant efficiencies are gained and simplicity is achieved through the automated request and approval processing being offered in an easy-to-access common location, which is Teradata Viewpoint.

Have It All

Teradata Data Lab provides self-provisioning and management features for DBAs, business users and others. With it, DBAs can easily configure a Data Lab within the lab group hierarchy to simplify governance. They can also assign owners and users to the Data Lab based on database permissions to maintain security and assign feature permissions based on the user's database comfort level.

Business users will have an easy-to-use collaboration and management environment to rapidly explore unproven data and test theories without long procurement processing. Users also benefit from agility in a well-protected and governed environment. Who says you can't have it all? T

Gary Ryback is senior product manager for Teradata Viewpoint and Teradata Active System Management.

This article originally appeared in the Q4 2011 issue of *Teradata Magazine.*

2012 Teradata Unified Data Architecture™ & Teradata Unity™

The Best Tool for the Job

The Teradata Unity™ portfolio offers functionality that improves performance and eases data management tasks.

by David Gilbreath

The Teradata® Unity™ portfolio keeps your data fresh, reliable and highly available so you can process and monitor data when and where you need to, using the tools of your choosing. The product suite also allows you to make full use of all of the computational and storage assets across your multi-system warehouse environment.

The latest releases of these tools increase the reliability and accessibility of data and comprise the next steps in building a comprehensive toolset that enables data movement, duplication, synchronization and monitoring in the Teradata Unified Data Architecture.™

Updated Functionalities and Abilities

A variety of enhancements have been added to Teradata Unity. The suite of integrated products offers intelligent query management and data synchronization in multi-system solutions such as the Unified Data Architecture. New features and advancements include:

UNITY DATA MOVER

This tool now supports the Unified Data Architecture by providing parallel bi-directional data movement between Teradata Database systems and Apache™ Hadoop® or the Teradata Aster Database. This is, of course, in addition to transfers between Teradata Database systems.

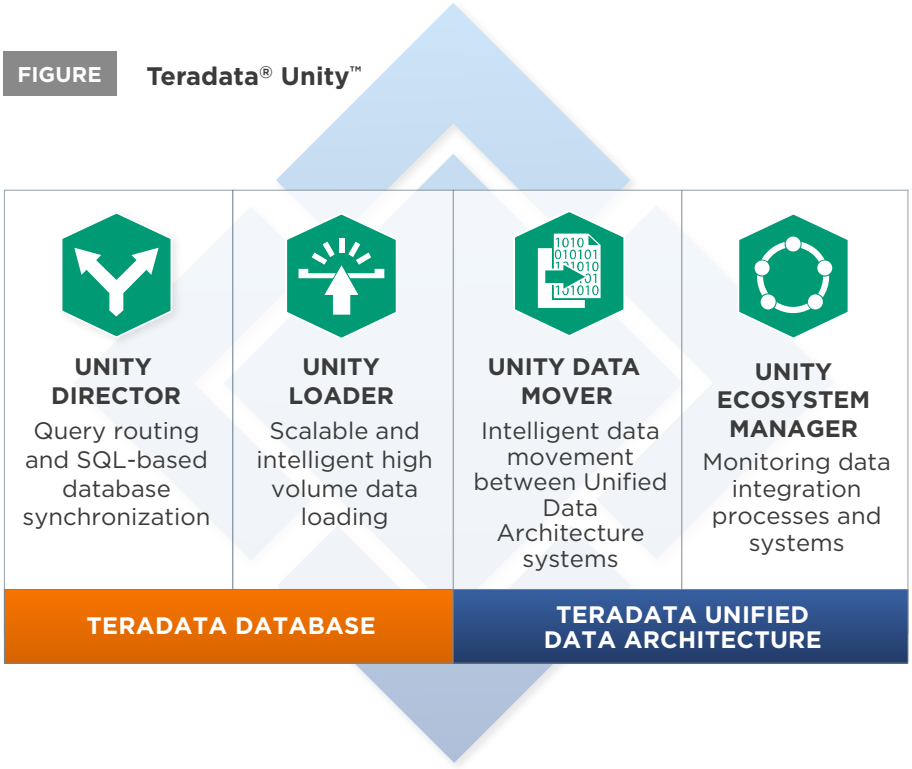
This is an important capability since data analytics can be performed anywhere in the data center. Unity Data Mover can transform loosely typed data in Hadoop into actionable insights or keep what-if and predictive analytics in the Teradata Aster Database refreshed with the deep stores of information from the data warehouse. Results and answers to business questions can be sent back to the data warehouse to take advantage of a wealth of applications. All of this interactivity is made possible by a common interface that automatically selects the optimal data transfer method.

Unity Data Mover maintains its easy-to-use, point-and-click user interface. You can construct data movement jobs on the fly without having to gain a full understanding of the differences between Teradata installations. This allows you to ignore details such as version variations in your Teradata system that might otherwise make it difficult or impossible to use certain utilities.

UNITY ECOSYSTEM MANAGER

One of the most difficult tasks in a multi-system environment is the relationship, tracking and correlation of all the parts of the ecosystem: hardware, processes, data and the state of each component. Unity Ecosystem Manager automatically tracks process jobs so you'll know if any are running behind or if your data is out of sync. Installing this tool enables the automatic creation of job execution histories. It also now

FIGURE **Teradata® Unity™**



Teradata® Unity™ is an integrated portfolio of products that routes queries, synchronizes the data and monitors the whole multi-system environment supporting the Teradata Database and the Teradata Unified Data Architecture™.

monitors MapReduce jobs in Hadoop as seamlessly as other jobs to give you full notifications and alerts. In addition, you're provided with a single view of the status of all jobs in your data warehouse without regard for which system runs them. This lets you "see" your Unified Data Architecture and make the most informed decisions for managing it.

The latest version of Unity Ecosystem Manager offers a unique perspective through an updated user interface. It lets you easily switch between views of your jobs, tables and systems in a way that flows logically, letting you discover more about the data environment. When tables come out of sync, your perspective can change to the jobs that act on that data or the systems and sensors that are the source of the data to find the problem or optimize the solution. This capability gives administrators a unique advantage in quickly and easily drilling into issues for root cause analysis.

UNITY DIRECTOR

Unity Director, the premier tool in the product portfolio, has a basic design method that supports a simplified list of Teradata processes and features. It is constantly being expanded to add more underlying database protocols with specific improvements to the handling of the data definition language (DDL). This support now includes performance enhancements for high-volume DDL while requiring less management of the Unity Director dictionary. All of this results in higher availability of database objects and providing evenly distributed workloads across multiple Teradata systems. These benefits, combined with a more functional and simplified user interface, make Unity Director easier than ever to manage and transparent to users.

UNITY LOADER

Unity Loader now includes support for the full set of Teradata client load utilities with the addition of Multiload and Teradata Parallel Transporter Update. You can direct even more of your load jobs at this solution to automatically keep all of your data warehouses in sync. There is no longer a need to load one system and then repeat that same load job on multiple systems. This means you are no longer forced to choose between fresh data and data that excludes live transactions. And more importantly, loading to multiple systems is automatic, transparent and provides all the fault handling you'd expect.

Performance on Multiple Platforms

Enhancements to the Teradata Unity portfolio include support for a broad range of tools and utilities that you already use in your data warehouse. The updates make interacting with the Unified Data Architecture much easier. You can manipulate data directly on systems in the Unified Data Architecture while keeping the data in sync with the Teradata Database. In addition, Teradata Unity improves performance and simplifies the management of all your jobs, whether they are on Hadoop, the Teradata Aster Database or the Teradata Database. **T**

David Gilbreath is the Teradata Unity suite architect responsible for Unity Director, Loader, Data Mover and Ecosystem Manager. He has more than 25 years of experience in software design and development, and has been with Teradata eight years.

This article originally appeared in the Q3 2014 issue of *Teradata Magazine.*

2013 Teradata Intelligent Memory

Queries at the Speed of Business

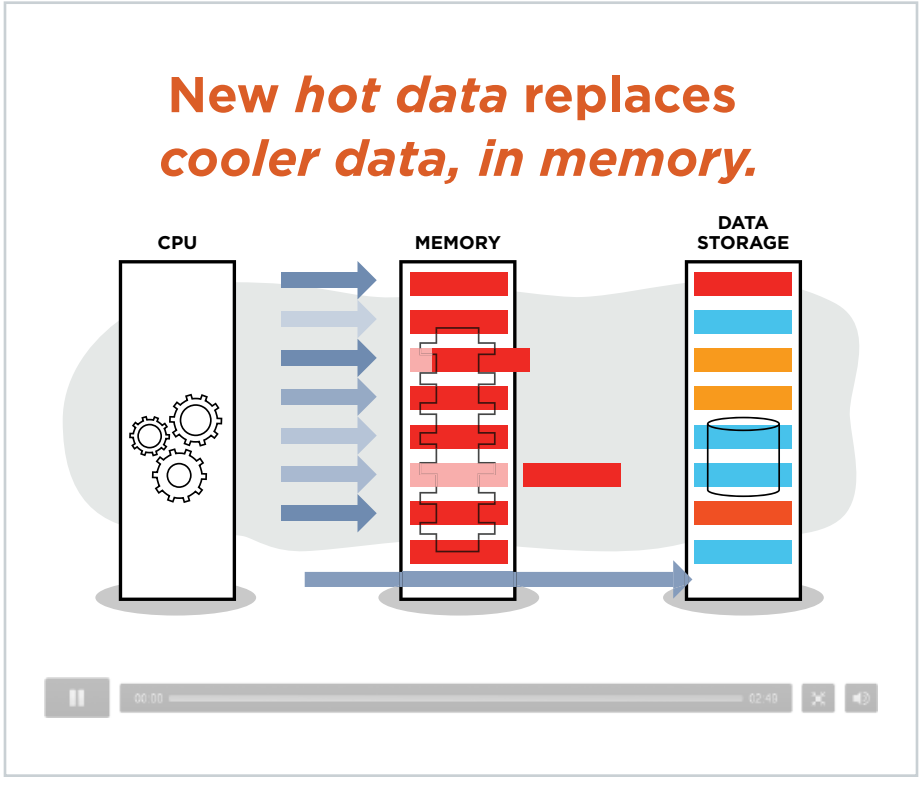
Teradata Intelligent Memory maximizes performance while increasing the value of memory.

by Alan Greenspan

Teradata® Intelligent Memory, an innovative technology built into Teradata Database 14.10, speeds data warehouse query performance and maximizes the value of system memory. It creates extended space beyond the cache and responds dynamically and intelligently to data access frequency. The result is the highest query performance at a fraction of the cost—and without the capacity limitations—of in-memory databases.

Teradata has taken a system-level approach to multi-temperature data management. Intelligent Memory is the latest feature in the list of temperature-aware features in the Teradata Database. It automatically and transparently keeps the most frequently used, or "hottest," data in memory. As usage patterns change, with some data being used less frequently, or "cooling," and other data being accessed more often, it reacts to ensure the hottest data is always available in memory for rapid access by the CPU for processing, which improves performance.

VIDEO



3,000 Times Faster

The Teradata Database continuously tracks the temperature of all its data. The most frequently accessed information is identified within the database on a "very hot" data list, which is modified and stays current as workload and data access patterns change. Whenever data on the list is read from disk during query processing, a copy is kept in the extended memory area for future use. When this data is needed by another

query, the database automatically looks to Teradata Intelligent Memory. This eliminates the need to access a physical disk where I/O speed is up to 3,000 times slower.

I/O for reading data from disk for CPU processing and the CPU processing cycles are the two major computer system resources in which constraints generally impact performance. By keeping the most frequently used data in memory and thereby eliminating needed I/O, performance is enhanced.

If the system is I/O-constrained, performance can be increased when the I/O requirement is reduced. However, if a system is CPU-resource constrained with its workload, then reducing needed I/O is unlikely to deliver the performance increase. The system will continue to be governed by the lack of CPU resources to process the data. Therefore, the benefits achieved by the solution will vary with workload and system configuration.

A combination of Database Query Log (DBQL) and ResUsage data can be used to understand the data usage and I/O patterns on a system. This analysis can help determine the benefit from enabling Teradata Intelligent Memory.

Make the Most of Memory

Teradata Intelligent Memory automatically copies information to its extended memory area whenever data on the very hot list is retrieved from disk for query processing. DBA intervention is not required to identify hot data or to tell the system which data should be copied and which should only be kept in its persistent storage location. The Teradata Database file system knows what data is available in memory and automatically uses that copy just as it would use data out of cache instead of going to disk. Query changes are not necessary, and no special query syntax is needed to take advantage of the solution. Business intelligence (BI) tool choice, user training and physical data design are also unchanged.

The solution makes the most efficient use of the memory it has available. If one of the database's many compression techniques is used to reduce the size of the data stored on disk, that compression will be carried into memory. This allows the maximum amount of user data to be kept there. Columnar storage format and columnar compression techniques will also be maintained.

Experience has shown that in most data warehouses, close to 50% of the query activity accesses just 1% of the data. Teradata Intelligent Memory leverages this observed fact in an innovative approach to achieve high database performance without the cost of purchasing enough memory to store the entire database. By keeping a copy of the most frequently used data in memory, physical disk I/O to solid-state disk (SSD) or hard disk drive (HDD) can be reduced dramatically.

Since all data is still kept on disk storage and remains available for use, the data warehouse continues to meet its full potential. Restrictions are not placed on user queries or data capture and retention in order to meet an artificial limit of storing all data in memory. This gives users the answers they need with the performance they want.

Rapid Access Equals Better Decisions

Teradata Intelligent Memory ensures that the most frequently used data is available for rapid access by the CPU for processing, which improves query performance. This gives users more timely insights so they can make better decisions and increase the value that organizations get from their data and their warehouse. T

Alan Greenspan joined Teradata in 1988 and is the product marketing manager for the Teradata Database and other core software products.

This article originally appeared in the Q4 2013 issue of *Teradata Magazine.*

WORKING WITH CACHE

Teradata® Database FSG cache (i.e., FileSeGment cache, the primary internal cache within the Teradata Database) is an extremely efficient, short-term database cache. As with most computer cache techniques, FSG cache is very fluid. Data is kept there for short periods—seconds or at most minutes.

Teradata Intelligent Memory works with FSG cache to focus on long-term data usage. The extended memory area works alongside the existing FSG cache to provide a more stable, temperature-based collection of data that will satisfy many queries over an extended period of time. To ensure the most efficient system operation, guidelines are available for minimum amounts of both FSG cache and Teradata Intelligent Memory. If enough memory is not available for both, the latter should not be enabled.

The New Norm

TERADATA INCREASES IN-DATABASE ANALYTICS CAPABILITIES WITH NEW DATA TYPES AND FUNCTIONS THAT 'EXTEND' THE DATABASE.

by Arlene Zaima

An industry rule of thumb is data management and analytics teams will dedicate approximately 70% of their development time to accessing, exploring and preparing data to ensure analytic models accurately represent their business. That leaves only a small amount of time for actual analysis.

Since most information is stored in a data warehouse, it makes sense to process the data where it resides instead of extracting it into a spreadsheet, flat file, data mart or proprietary datasets. That's why in-database analytics is no longer an emerging trend. It's now the norm for organizations that use analytics to drive their business and customer interactions. >>

EVERYTHING MANUFACTURERS NEED TO KNOW

Teradata's new in-database functions and capabilities, in conjunction with a number of analytic partners, integrate technologies to help organizations overcome their pain points. For example, an automobile manufacturer can track and analyze all components and manufacturing tasks down to an atomic level to identify faulty products, suppliers and processes. The manufacturer can capture and analyze details about a single rivet's supplier, alloy composition, time of day it was made, location, machine and machine pressure to detect early failure rates. This can prevent costly recalls and reduce customer services calls.

In-database analytics eliminates unnecessary data movement and iteration in the exploration and preparation processes by allowing users to access the data where it resides. This lets analysts focus on identifying insights instead of the mechanics of moving and integrating data.

Teradata offers in-database solutions that enable users to more easily analyze data for greater business value. Teradata is also partnering with companies to extend and operationalize analytics across the enterprise.

Vast Analytic Functions

Once data is integrated, explored and prepared for analysis, a variety of tools is needed to perform everything from exploration and discovery to modeling and deployment within the data warehouse. Teradata collaborates with a number of partners to integrate their technologies with the Teradata® Database to provide a greater depth and breadth of analytic capabilities.

Tableau,® TIBCO Spotfire,® SAS,® IBM® SPSS, Alteryx and Information Builders offer SQL pushdown capabilities, while the database provides the heavy lifting. For example, Tableau and TIBCO Spotfire provide a visual interface to enable rapid data exploration, while the other companies push different degrees of data preparation, model development and model scoring into the database via user-defined functions (UDFs), a predictive model markup language (PMML), R and/or SQL generation, all running in parallel.

Teradata has also partnered with Fuzzy Logix™ to provide hundreds of new analytic functions embedded in the Teradata and Teradata Aster Databases. These functions let businesses analyze all their data, not just a subset or limited number of variables. Fuzzy Logix analytics are easily accessible through SQL programs and can be used by any new or existing applications or tools. Adding Teradata Database capabilities to these tools allows users to analyze thousands of variables across millions of rows to solve complex business problems. >>

Extend the Database

As new analytic technologies have been introduced, Teradata has extended the core database engine to address the challenge of how to parallelize non-parallel technologies. For example, R, the open-source analytic language used by most data miners and data scientists, operates in-memory and is a non-parallel language. In-memory processing is commonly used for server-based products and allows for very fast processing; however, data must be transferred to the server and the amount of data processed is limited by memory size.

One method of tackling this problem is to run multiple instances of R in-memory across many servers. Although this achieves concurrent execution, it doesn't provide system-level parallelism for analytics that need to use all the data, such as when calculating the median of a group of numbers. This requires a view of all the data so it can be sorted to select the middle value. Running the median function independently across multiple database nodes produces a response for each node, and the onus of parallel programming is on the analyst.

Teradata and Revolution Analytics have tackled the difficult problem of solving the in-memory and non-parallel R language challenges with the first out-of-the-box parallel, in-database R solution. By bringing parallelized R algorithms into the Teradata Database, analytic performance is greatly enhanced and the latencies previously required to extract data are eliminated.

Revolution R Enterprise for Teradata includes a library of Parallel External Memory Algorithms (PEMAs) that help achieve the highest possible performance. PEMAs are pre-built, extended-memory, parallelized versions of the most common statistical and predictive analytic algorithms that run directly in parallel across the Teradata system.

To facilitate integration, the Teradata Database has opened access to table operators that allow partners to integrate their analytic framework to leverage the database's performance, capacity and workload management capabilities. This expands the database memory, disk and processors transparently to all R programmers.

Growing Capabilities

In addition to new data mining algorithms and languages, Teradata continues to increase in-database capabilities with new data types and analytics. Teradata Database 14.10 enhances XML, geospatial and temporal data support by simplifying processes of these native data types.

XML documents, the most commonly used message format, can be stored and processed in their entirety directly in the database. XML data types, XQuery language, XPath, XML schemas, XML shredding and XML publishing within the Teradata Database give users the ability to better manage, integrate and process queries at high speeds.

Keeping Pace With Business

Teradata now offers the largest, most complete analytic library of more than 1,000 in-database analytics, enhanced access to native XML data, streamlined temporal analytics and capabilities that make geospatial data faster and easier to use. These in-database capabilities allow analysts to have a conversation with their information, regardless of the data type, language or method.

RECONCILE CHANGES IN A SINGLE LOCATION

Allowing XML functions inside the Teradata® Database can simplify processes for businesses, such as capital markets, where deals are struck and data is stored in a dozen or more independent systems. Each disparate system can store data covering different financial instruments, equities, derivatives, bonds and foreign exchanges across various geographies and currencies. Banks currently employ hundreds of employees dedicated to reconciling transactions across multiple systems because they don't have a single source for this data.

However, when the XML data covering these transactions is stored in the Teradata Database as the single master source, changes can be reconciled in one location, then managed, analyzed and shared with the appropriate departments and systems, which drives accuracy and efficiency.

The Teradata Database keeps pace with businesses' analytic requirements by providing new languages, data types and analytics across the Teradata integrated data warehouse and Teradata Aster Discovery Platform. This lets organizations integrate state-of-the-art analytic resources to drive their business. T

Arlene Zaima is a program manager for Teradata Integrated Analytics solutions, including Geospatial and Agile Analytics Data Lab.

This article originally appeared in the Q4 2013 issue of *Teradata Magazine.*

The Perfect Fit

Teradata Cloud offers fast, easy access to a powerful analytics platform in an affordable and flexible pricing model.

by Paul Barsch

Today's agile businesses are seeking to expand analytics fast, gain flexible analytic deployment options and smooth cash flows. Teradata understands these needs and has engineered a cloud solution that meets organizations' analytic needs.

Teradata® Cloud provides enterprises with flexible "as a service" options for data warehousing, discovery platform solutions and open-source Apache™ Hadoop.® These capabilities are available in a simple subscription-based pricing model that helps reduce capital expenditures. Cloud starter kits are also available that allow companies to leverage Teradata's vast industry expertise to load data quickly, create BI reports and gain business value in a rapid fashion.

Organizations can now choose between an on-premises installation or flexible cloud options that are secure and highly available. With Teradata Cloud, it has never been easier to gain instant access to Teradata technologies to meet today's competitive forces head on and capitalize on future opportunities. >>

Key Considerations

Industry analysts enjoy debating the substance of cloud computing. However, squabbles on definitions are mostly counterproductive because the real value of cloud isn't in how it is defined; it's the business value it delivers. In a *Financial Times* publication titled, "The Business Landscape of Cloud Computing," analyst Darryl Plummer writes: "Cloud computing means someone else runs your computers and software while you use what they deliver and focus on delivering value." This simple definition shows the real value of cloud is the ability to focus on customer needs while "someone else" takes care of analytic infrastructure and operations.

In terms of performance, it's important to recognize that not all cloud infrastructures are created equal. That's because analytic workloads are typically CPU- and I/O-intensive. Therefore, it does not make sense to run analytic workloads on cloud services that are provisioned for general purpose computing one day and for data warehousing the next. Business users expect top tier performance, which is why a cloud environment dedicated and engineered specifically for analytic workloads is imperative.

Lastly, when weighing analytic cloud options, considerations such as performance and functionality should override the siren song of "lowest cost per terabyte." Few people would buy a house based on the lowest price per square foot. Likewise, a low price for limited features and lethargic performance might work for a modest dataset and a small number of concurrent queries, but these cloud offerings will usually not meet the needs of demanding business users.

Fast, Flexible and Powerful

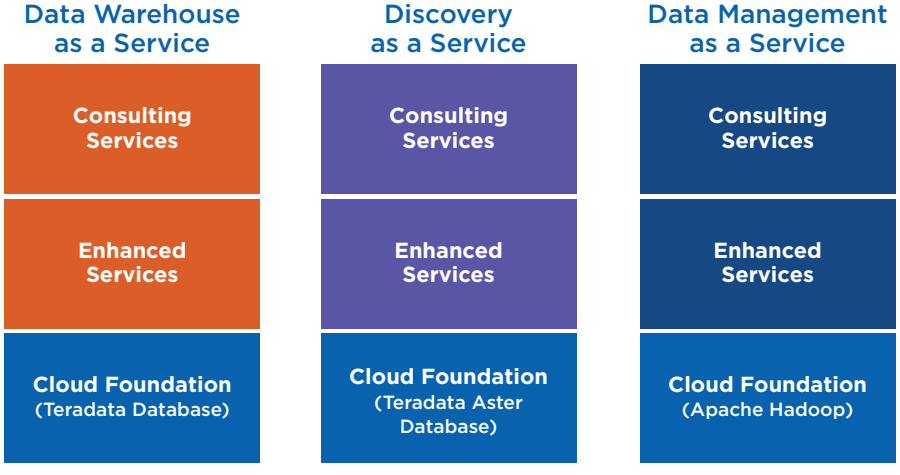
Because performance, service level agreements (SLAs) and availability are important to business users, Teradata is now offering a complete cloud solution that delivers a cloud foundation, enhanced operational services and consulting services to help advise, architect and manage analytics. Teradata Cloud eliminates the need for organizations to procure, maintain and capitalize an analytic infrastructure. Instead, analytic resources can be quickly added to scale up for expanding business needs.

With Teradata Cloud, organizations have the capability to be up and running in days instead of waiting weeks or even months for IT equipment to be shipped, installed and powered on. And cloud capabilities from Teradata can be acquired in a subscription pricing model to help smooth cash flows for asset-light businesses.

Adding to Teradata's success in cloud computing are flexible options to deploy:

- **Teradata Database**, the world's most robust database engine with sophisticated workload management and unparalleled features such as support for geospatial, temporal and columnar-based analytics.

FIGURE **Teradata Cloud Foundation and "As a Service" Options**



Teradata® Cloud offers a full range of analytic services, including data warehousing, discovery capabilities and data management.

6 FACTORS DRIVING CLOUD ANALYTICS

In addition to improved business agility, organizations are increasingly turning to cloud-based analytics for six reasons:

- **"Cloud first" mandates.** Some enterprises institute a "cloud first" policy because C-level leadership no longer wants to support data centers, license software and capitalize IT assets.
- **Departmental use.** Business users will acquire application-specific data marts in the cloud.
- **Analytic pilots.** Data warehousing or discovery pilot projects can take advantage of cloud services for a short duration to show business value quickly.
- **Platform for applications.** Cloud options are often attractive for enterprises that wish to build solutions on top of an existing analytic platform.
- **Development.** Workloads for development are shifting to the cloud to avoid affecting production data warehouse service level agreements.
- **Disaster recovery.** Cloud services provide a convenient, affordable option for IT disaster recovery.

- **Teradata Aster's** pre-built library of more than 100 SQL-MapReduce® functions that allow business users to perform statistical analytics, text analytics, graph, path analysis and more in a single solution.
- **Teradata Portfolio for Hadoop** for capturing, storing and refining multi-structured data.

Teradata's "as a service" options can include everything for end-to-end analytics, such as production DBA support, improved security for compliance purposes, enhanced backup, and management for data integration and business intelligence (BI) servers. Additional consulting services are also available for advising on analytic best practices, migrating from existing databases, implementing new applications and providing analytic access.

To succeed in analytics, customers require more than just a platform and database. Teradata is leveraging its unique vertical industry expertise to provide optional industry starter kits to help enterprises jump-start their analytics projects. These accelerated packages include high-level logical and physical data models, and industry report templates.

Remote Access to Powerful Analytics

Some enterprises have invested heavily in acquiring, training and developing talent in database administration, application development and sustainment, data modeling and other analytic skill sets. For these situations, the Teradata Cloud Foundation is a perfect fit since it offers remote access to the Teradata Database, Teradata Aster's analytic services and Portfolio for Hadoop with maintenance, daily backup and base security all included in one low monthly subscription price. (See figure, previous page.)

With this solution, Teradata handles basic cloud infrastructure functions such as hardware/software monitoring and maintenance, security administration and resource provisioning. Teradata Cloud Foundation includes assistance with:

- Customer onboarding
- Base security and networking
- Data center management
- Backup and recovery
- System availability
- Provisioning
- Daily operational management

Best Solution Possible

With Teradata Cloud, organizations can select flexible cloud options that deliver data warehousing, discovery analytics and data management solutions. All data types, including structured and multi-structured, can be conveniently stored and analyzed in a cost-effective and cash-flow friendly manner. To meet current and future business needs, there's no better option than Teradata Cloud—a powerful analytic solution that's scalable, reliable, secure, and supported by trained and certified associates. T

Paul Barsch is a services marketing director for Teradata and a regular contributor to Teradata Magazine. *He has more than 15 years of IT experience.*

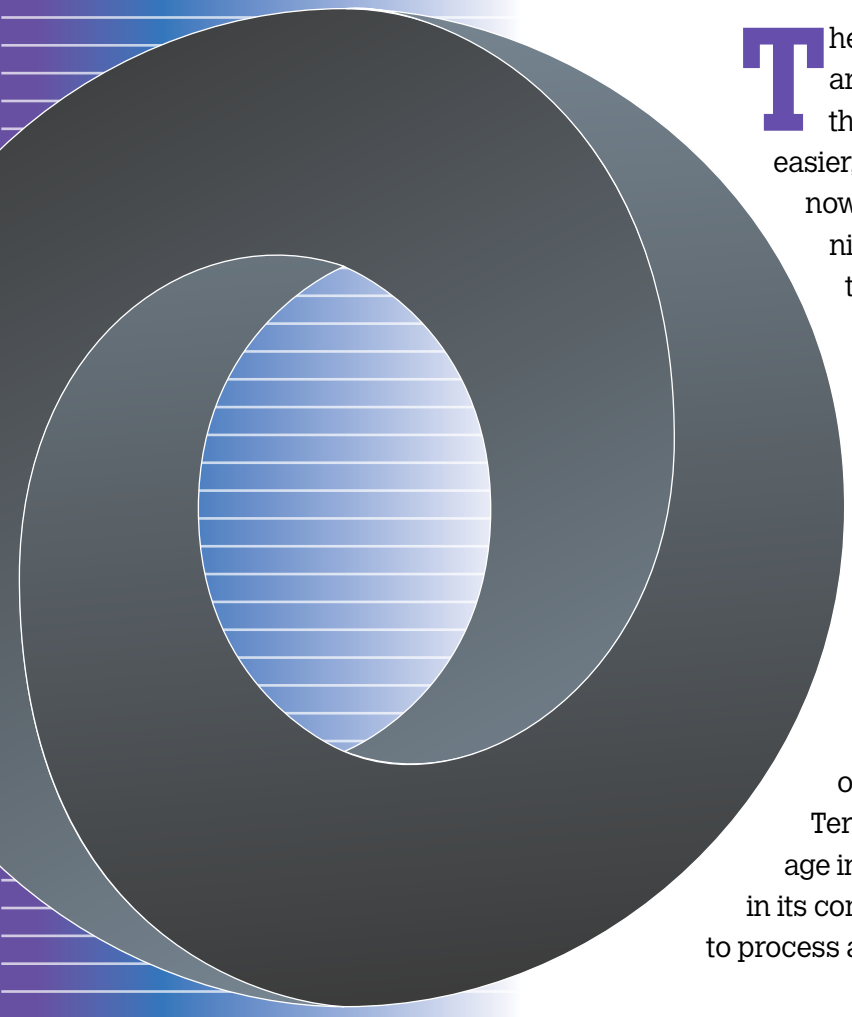
This article originally appeared in the Q4 2013 issue of *Teradata Magazine*.

2014
JSON

Monetize the Internet of Things

JSON turns a flood of data into business actions and results.

by Alan Greenspan and Cameron Lewis



The world around us is humming with data. No longer are everyday devices isolated islands quietly doing their tasks to make our personal and business lives easier, safer and more efficient. All kinds of devices are now interconnected in the Internet of Things, communicating status, environmental facts, and all sorts of transaction and interaction data. The tremendous volume is overwhelming clutter unless it can be easily processed. By combining this data with other sources in an integrated data warehouse (IDW), it can be successfully stored, organized and analyzed to enrich the business with new revenue streams, valuable services and remarkable operational efficiencies.

JSON has become a de facto standard format for exchanging machine-generated data. JSON, or JavaScript Object Notation, is an open-standard format for human-readable text and often consists of key-value pairs in a hierarchical structure. The Teradata® Database can either shred the data for storage in individual structured columns or natively store it in its complete and original JSON structure, allowing users to process and query the data with sophisticated SQL. >>

JSON at Work

Companies can receive supplies orders for a device they manufacture from several channels: the Web, a self-service kiosk in a store or directly from the device in the form of an automatically generated order to refill a supply, like printer toner, that is running low. The device, as well as the other channels, can send all of the information related to the toner order in a JSON file that looks like this:

```
{
     "customer" :
       "CustomerName",
     "customer_number" : 375,
     "order_date" :
       "2014/03/12",
     "items" :
        [
           {"ID" : 123,
            "name" : "black
            toner", "amt" :
            1},
           {"ID" : 234,
            "name" : "blue
            toner", "amt" :
            1}
        ]
}
```

The data can be kept in the JSON format in the IDW by creating an order table with a JSON data type column. Each row contains the information pertinent to one order, but all of the elements are in the JSON structure within the single column rather than in a separate column for each element of the order. The order table can be created with this statement:

```
CREATE TABLE orders (order_
  num INTEGER, order_data
  JSON(250));
```

Once loaded, the incoming order is stored in a single row in the table. All data about the order is in the `order_data` column of the single row. This easily accommodates the array of line items included in the order as described in a standard JSON file.

Using the data is simple with an easily read JSONPath extension to Teradata SQL that mirrors how JSON is used in its native JavaScript environment. A "dot notation" identifies the hierarchy and keys in the JSON structure, much like `table.column` name identifies columns in a table. Several methods on the JSON data type and regular expressions in the JSONPath syntax make powerful queries easy to express. For example, the items ordered by customer 375 within the past 30 days can be found with the following query working directly on the JSON data in the orders table:

```
SELECT order_data.customer_
  number, order_data.
  JSONExtractValue('$.
  items[*].ID')
FROM orders
WHERE order_data.customer_
  number = '375' AND
  order_data.order_date >
  (CURRENT_DATE-30);
```

Analytical and Data Flexibility

Storing data in its original JSON structure provides advantages inherent in using JSON as a data interchange format. Since it uses key-value pairs and is completely self-describing, it offers the flexibility of "late binding" or schema-less processing. This flexibility is maintained in the integration of JSON multi-structured data into the Teradata Database.

With JSON, the data elements and attributes are not defined in the table structure or schema. In the printer toner example, a single 250-character column includes all of the data for an order. The data and structure are defined when a query is run and the database looks at the keys in the key-value pairs within the stored data itself (e.g., order date, items, ID).

Discovering the structure at query time when the database looks at the information is called late binding, schema-less processing or query on read. This enables business flexibility by accommodating dynamic data situations. The information elements included in records can change without a database change. New elements that would require a new column in a traditionally modeled and structured environment could be added simply by inserting a new key-value pair in the next input data loaded into the table. The table itself does not have to change.

If new firmware on installed printers or a new printer model was able to include a "required delivery date" in the order information, for example, that data could immediately flow into the IDW with no physical table change. The business user or order processing application developer would only have to reference the new

EXPECT PERFORMANCE

Many of the standard Teradata® Database performance techniques are just as valid when using JSON data as using other data. For instance, collecting statistics on portions of the JSON data, similar to collecting statistics on a single column within a table, enables effective query planning by the optimizer.

Join indexes (JI) can also be defined with portions of the data in the JSON column, and statistics can be collected on the JI. The JI is used like any other join index. The optimizer will decide whether it is more efficient to access the data in the JI or to scan the JSON text in the base table.

Another physical database design technique is to "shred" a limited amount of the data from the JSON column and duplicate it in individual structured columns. This allows the database to directly access those data elements. For example, if queries on an order often qualify to access data based on `customer_number`, the user can make a `customer_number` column of type `INTEGER` and use it as part of the `WHERE` clause, or even include it in the table's primary index.

Shredding and publishing commands are available to extract data elements into columns or to build a JSON file with data from separate structured columns. Using an ELT data acquisition design, JSON data can be loaded into a staging table, then `MERGEd INTO` the production table while simultaneously shredding key data into structured columns.

Integrating multi-structured JSON data into the Teradata® Database offers new flexibility and enables organizations to monetize the flood of data coming from the Internet of Things and elsewhere.

key in a query to understand how urgent an order is and ensure customer satisfaction through on-time deliveries. The data warehouse management, design and testing necessary for the business process change are eliminated. There are even built-in commands to find the keys contained within the data through a query.

This query retrieves the required delivery date along with order information:

```
SELECT order_data.customer_
  number, order_data.
  JSONExtractValue('$.
  items[*].ID'),
  order_data.
  required_delivery_date
FROM orders
```

Rows without the `required_delivery_date` key-value pair return a `NULL` for that value. Data in a JSON column can be used along with all of the other information in the database. Additional columns with structured types can be in the same table as the JSON column, or data from the column can be joined with data from other tables. The user has analytical freedom to work with all of the data in the IDW, whether it is semi-structured or structured.

Make the Most of All Data

JSON is a very common data interchange format. It can be stored along with structured data in the Teradata Database and used directly in queries along with the other data in the data warehouse. Integrating multi-structured JSON data into the Teradata Database offers new flexibility and enables organizations to monetize the flood of data coming from the Internet of Things and elsewhere. T

Alan Greenspan joined Teradata in 1988 and is the product marketing manager for the Teradata Database and other core software products.

Cameron Lewis has been a Teradata software engineer since 2010. He has extensive experience in Java and JSON.

This article originally appeared in the Q2 2014 issue of *Teradata Magazine.*

2014 Teradata QueryGrid™

Harmonious Orchestration

Teradata® QueryGrid™ lets users benefit from all data in a unified architecture and beyond, regardless of where it's stored.

by Imad Birouty

As the role of analytics within organizations continues to grow, along with the number and types of data sources and processing requirements, companies face increasing IT complexity. Much of the complexity arises from the proliferation of non-integrated systems from different vendors, each of which is designed for a specific analytic task.

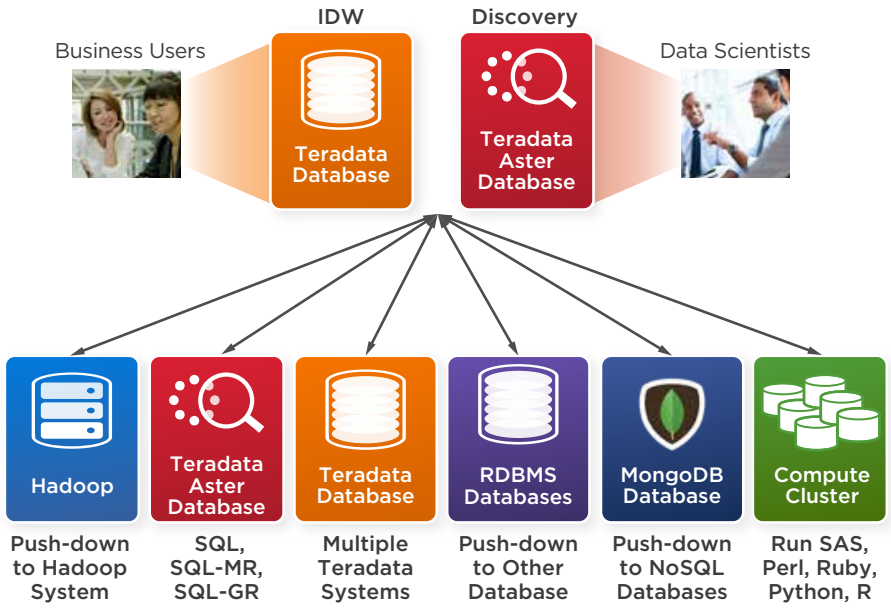
This challenge is best addressed by the Teradata® Unified Data Architecture™, which enables businesses to take advantage of new data sources, data types and processing requirements across the Teradata Database, Teradata Aster Database and open-source Apache™ Hadoop®. Teradata QueryGrid™ optimizes and simplifies access to the systems and data within the Unified Data Architecture and beyond to other source systems; delivering seamless multi-system analytics to end-users.

This enabling solution orchestrates processing to present a unified analytical environment to the business. It also provides fast, intelligent links between the systems to enhance processing and data movement while leveraging the unique capabilities of each platform. Teradata Database 15.0 brings new capabilities to enable this virtual computing, building on existing features and laying the groundwork for future enhancements.

One System With Seamless Operations

Teradata QueryGrid delivers transparent data access and local processing across the systems within the Unified Data Architecture. (See figure, next page.) It takes advantage of each system's specialized engines by sending parts of queries—and even data if necessary—to the other platforms for execution. Data placement and movement are optimized and intelligent query processing is automated

FIGURE **Teradata® QueryGrid™**



When fully implemented, the Teradata Database or the Teradata Aster Database will be able to intelligently use the functionality and data of multiple heterogeneous processing engines.

for best overall results. This allows organizations to reap the most value from their unified architecture by easily and transparently pushing analytics to the data or moving the data to the right analytic platform, all without IT intervention and by allowing users to send their query to only one system; either the Teradata Database or the Teradata Aster Database.

For example, a user will submit a single request to the Teradata Database and receive a completed result back, but some of the work may be performed on one of the other platforms. Other systems get involved because they have the needed data or can perform a function that the Teradata Database cannot.

Once this multi-system analytics vision is fully implemented, Teradata QueryGrid will act as an orchestration layer to utilize the resources of multiple heterogeneous processing engines to complete a query sent to the Teradata Database or the Teradata Aster Database. Each database will be able to push analytics to the data on another platform by requesting a piece of the query be executed there, and it can get data from or send data to the other platform for use in the query. This allows the specialized functions available only in the target system to operate on the data. Each system can be used for its specialized processing capabilities and the data it holds.

Real-World Benefits

Multi-system analytics powered by Teradata QueryGrid delivers functions that help organizations become more effective and efficient. For instance, analysts may want to retrieve data that resides in Hadoop for use in a Teradata Database query. Hadoop offers a useful data platform for holding large volumes of data that have not yet been proven valuable to store in the data warehouse. An analyst can issue a query to the Teradata Database and request to use data stored in Hadoop.

A growing trend in the auto insurance business is the use of telematics in which a driver's behavior is monitored directly and the information is transmitted to an insurance company. The company assesses the risk the driver poses and charges premiums accordingly. This massive collection of raw telematics data can be stored in Hadoop. An analyst can use the Teradata Aster Database to perform a path analysis on the telematics data to determine which driving behaviors are most likely to lead to an accident, such as frequent hard breaking followed by sudden lane changes. Once identified, those patterns can be loaded into the Teradata Database for analysis. The analyst can then write a query to the Teradata Database that moves a filtered set of telematics data from Hadoop into the data warehouse for analysis against the patterns to identify drivers with the highest risk. The risk information is cross-referenced with customer data in the warehouse to determine whether a driver is a high- or low-value customer, and the analyst can then make recommendations on insurance premium adjustments.

As this virtual computing continues to grow in functionality, even more advanced capabilities will be enabled, such as executing advanced processing across systems. For example, customer retention agent may be dealing with an unhappy customer at risk of defecting may already have information about the value of the customer. However, the agent won't know whether this is a >>

TERADATA® QUERYGRID™

The goal and vision of Teradata® QueryGrid™ is to make specialized processing engines, including those in the Teradata Unified Data Architecture™ act as one solution from the user's perspective. Teradata QueryGrid is the core enabling software, engineered to tightly link with these processing engines to provide intelligent, transparent and seamless access to data and processing. This family of intelligent connectors deliver bi-directional data movement and push-down processing to enable the Teradata Database or the Teradata Aster Database systems to work as a powerful orchestration layer and interact with other systems by:

- Retrieving data from a remote system
- Requesting processing of data on a remote system, followed by retrieval
- Sending data to a remote system for processing, followed by retrieval

LAY THE FOUNDATION WITH FABRIC-BASED COMPUTING

An architecture called fabric-based computing provides the infrastructure to enable the high-speed movement of data and deliver the computing required for multi-system analytics at extreme speeds. The Teradata approach to fabric-based computing leverages these elements for seamlessly accessing data across the Teradata® Unified Data Architecture™:

- **Teradata BYNET®** V5 is built on InfiniBand technology. Perfected over 20 years of massively parallel processing experience, BYNET V5 provides low-latency messaging capabilities for maximum data access. This is accomplished by leveraging the inherent scalability and integrity of InfiniBand to load-balance multiple fabrics, seamlessly handling failover in the event of an interconnect failure.
- **InfiniBand technology**, a Teradata fabric, gains much of its resiliency from the Mellanox-supplied InfiniBand switches, adapters and cables that are recognized as industry-leading products for high-quality, fully interoperable enterprise switching systems.

- **Teradata system management and support** provides field-proven capabilities—expanded and enhanced to include all elements of the fabric-based computing system—for full monitoring and proactive support within a single, intuitive interface.
- **Teradata Managed Servers** are fully integrated into the Teradata system management infrastructure. The servers are interconnected on the InfiniBand fabric as the platform for running all Teradata Unity products and data management operations.

highly influential customer who can positively or negatively affect the beliefs and behavior of others within his or her social network.

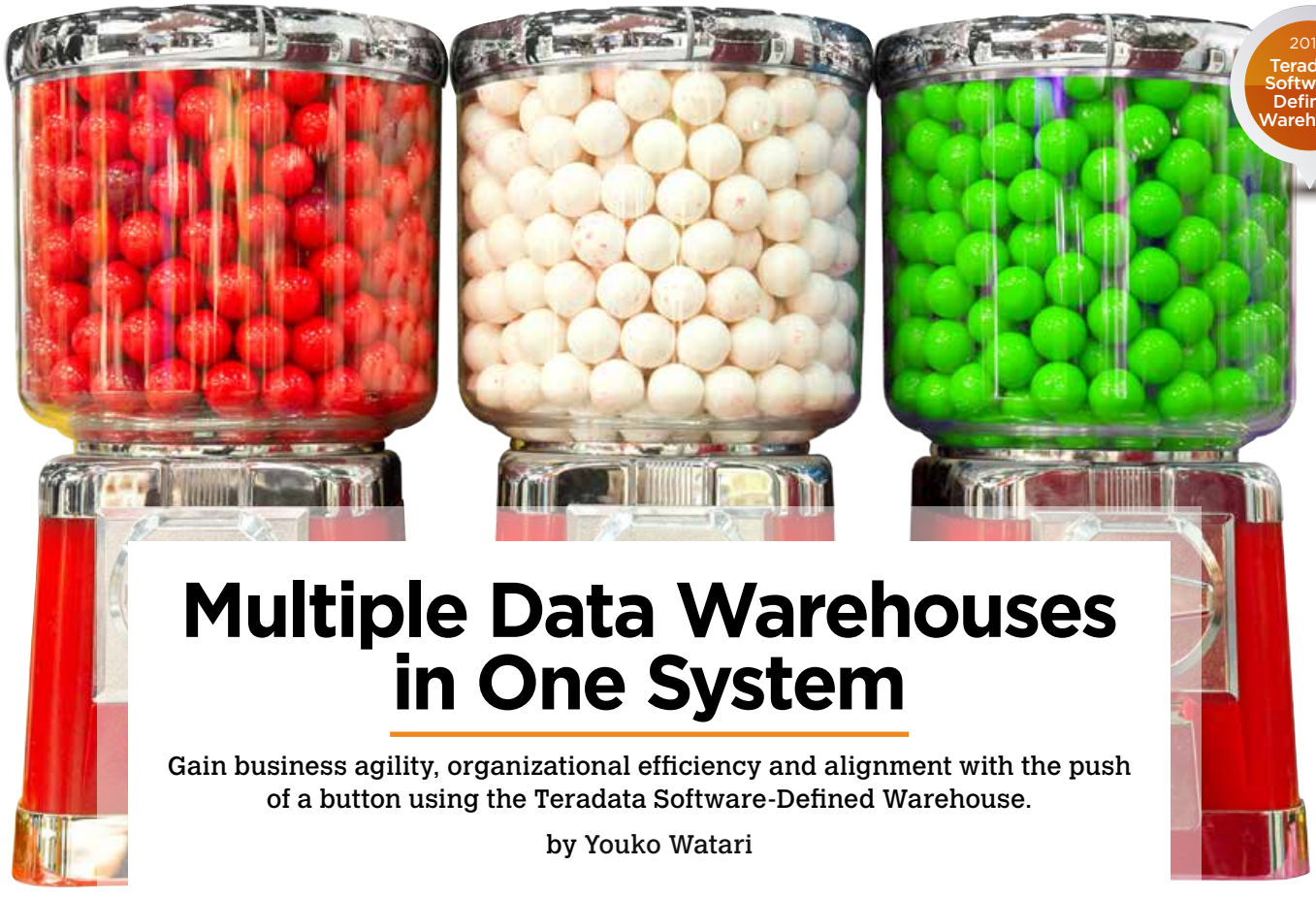
The call center agent can run a query to determine the social influence of the customer. The query will be submitted to the Teradata Database and in the background, customer data is sent to the Teradata Aster Database where a graph analysis determines the user's sphere of influence. Results are returned to the Teradata Database for a quantitative measure of the customer's value based on social influence. The agent then has the data needed to make an informed decision.

The Bar is Raised

Teradata QueryGrid is a powerful enabler of technologies within and beyond the Unified Data Architecture that delivers seamless data access and localized processing. It empowers users to immediately and automatically access and benefit from all their data along with a wide range of processing capabilities, all without IT intervention. This solution raises the bar for enterprise analytics and gives companies a clear competitive advantage. T

Imad Birouty is a marketing manager for Teradata's high-availability solutions and total cost of ownership program.

This article originally appeared in the Q2 2014 issue of *Teradata Magazine*.



Multiple Data Warehouses in One System

Gain business agility, organizational efficiency and alignment with the push of a button using the Teradata Software-Defined Warehouse.

by Youko Watari

You manage your company's Teradata system and receive a phone call: "We are adding a new subsidiary, so get ready!" You are told that the data for this subsidiary needs to be hosted on your Teradata system right away, but the data must be completely secured and isolated from the rest of the company's information. In fact, even *you* are not supposed to be able to view it. Your Teradata system is already scaled to accommodate additional data, but you are not sure how you can tackle the data security requirement. Ordering a new server would be an easy choice, but that is neither timely enough nor cost-effective.

You know that the Teradata® Software-Defined Warehouse with a new Teradata Database feature, Teradata Secure Zones, is an ideal solution for utilizing your existing Teradata system while hosting a completely isolated environment for a set of data and specific users. When you decide to give it a go, you research best practices and come up with a set of guidelines to ensure success. Here are the steps you'll need to follow.

Establish a Game Plan

Before creating zones in the Teradata system, first evaluate your fundamental requirements and plan accordingly with consideration for:

- **Zone modeling.** How many zones are needed and which data should be included in each?
- **Zone administration.** Who should be responsible for the DBA tasks? Should that individual be inside or outside the zones?
- **Zone users and guests.** Who should have what level of access, both inside and outside the zones?
- **Workload management.** What level of business priority or service level expectation do the zone workloads have? What is the impact of the existing I/O, CPU and priority management model? Should virtual partitions be used in conjunction with the zones to manage the system resources?
- **Data loading.** How, when and which data should be loaded into the zone?

Establish a New Zone

Your current system has a `Corporate` database with objects such as tables, views, DBAs, users and roles for your company. The new zone for the recently purchased subsidiary, Acme Company, is created directly

under the Corporate database. (See figure.) The process requires you to:

1. Create a zone administrator.
The DBC user creates a zone administrator for you (using the name Mike), Mike_ZoneAdmin, under the Corporate database with necessary privileges:

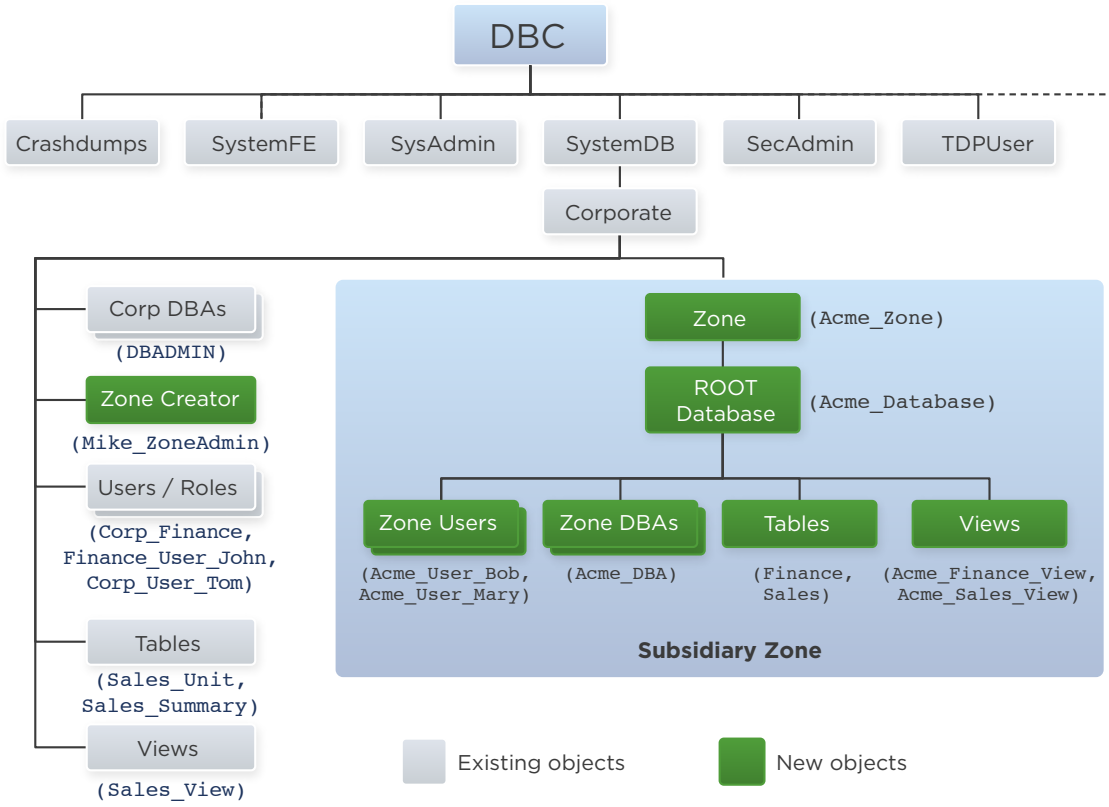
```
CREATE USER Mike_ZoneAdmin
   FROM Corporate
AS
     PERM = 1000000,
     PASSWORD =
       Mike_ZoneAdmin;

GRANT DATABASE ON Corporate
   TO Mike_ZoneAdmin;
GRANT ZONE TO Mike_
   ZoneAdmin WITH GRANT
   OPTION;
```

2. Create a zone, assign a database as a zone root and create a primary zone DBA.
You, as Mike_ZoneAdmin, create a zone, Acme_Zone, and a database, Acme_Database. You then alter the zone to assign the database as the root. You also create a primary zone DBA, Acme_DBA:

```
CREATE ZONE Acme_Zone;

CREATE DATABASE Acme_
   Database FROM Corporate
   AS PERM = 5000000;

ALTER ZONE Acme_Zone ADD
   ROOT Acme_Database;

CREATE USER Acme_DBA FROM
   Acme_Database
AS
     PERM = 1000000,
     PASSWORD = Welcome1;
```

3. Create users, tables and views.
Acme_DBA creates zone users (e.g. Acme_User_Bob) within

FIGURE **Teradata® Database Object Hierarchy**



This figure illustrates the objects and hierarchy within the Teradata system after a zone and zone-related objects are created.

the Acme_Database and grants privileges:

```
CREATE USER Acme_User_Bob
   FROM Acme_Database
AS
     PERM = 1000000,
     PASSWORD = Welcome2;
GRANT ALL ON Acme_Database
   TO Acme_User_Bob WITH
   GRANT OPTION;
```

Zone user Bob can create additional zone users. For example, he can create a user Mary, who has read-only access to Acme data by using GRANT SELECT ON syntax. Since Bob has the privilege to create databases and other objects within the zone, he can create tables and views for Acme's finance (Acme_Database.Finance and Acme_Database.Acme_Finance_View) and sales (Acme_Database.Sales and Acme_Database.Acme_Sales_View) data.

4. Grant privileges to zone guests.

You are notified that the users who are assigned to the corporate finance role (Corp_Finance) should have a SELECT privilege to the Acme_Database.Finance table within the zone.

For the Corp_Finance role, you as the zone administrator grant access to the Acme_Zone. This makes the role a zone guest:

```
GRANT ZONE Acme_Zone TO
   Corp_Finance;
```

Then a zone user with a grant privilege, like Bob, grants a read-only privilege to the zone guest, Corp_Finance role:

```
GRANT SELECT ON Acme_
   Database.Finance TO
   Corp_Finance;
```

Finalize the Implementation

Once all zone-related administration tasks are complete, you update the workload management settings in order to manage the system resource allocations and priorities of your existing and new workloads. You do this through software controls within the Workload Management capabilities in the Teradata Software-Defined Warehouse.

For testing, a series of queries validates the privileges granted to various users. For example:

- Zone user Mary can run SELECT queries against all tables and views within the Acme_Database, but SELECT queries against any tables or views in the Corporate database will fail because she was not granted a privilege outside the zone.
- Users who belong to the Corp_Finance role can run SELECT queries against the Acme_Database.Finance table. But SELECT queries against the Acme_Database.Sales table or other tables/views will fail because they are not granted that privilege.
- You can try to make yourself a zone guest of Acme_Zone. However, it will fail because a zone administrator cannot grant a zone to himself or herself.
- The primary zone DBA (Zone_DBA) can try to grant a SELECT privilege to a Corporate user who is not a zone guest. This will also fail because non-zone guests cannot have privileges within a zone.

Mission Accomplished

Using the Teradata Software-Defined Warehouse, you just created a secure environment for the new subsidiary without going through the lengthy process of procuring and configuring a new server or managing consistent system performance. You achieved timeliness and cost-effectiveness all through software control—your ultimate goal! T

Youko Watari is the technical marketing manager at Teradata who is responsible for the Teradata Database and core software products.

This article originally appeared in the Q3 2015 issue of *Teradata Magazine.*

SEPARATE AND SECURE

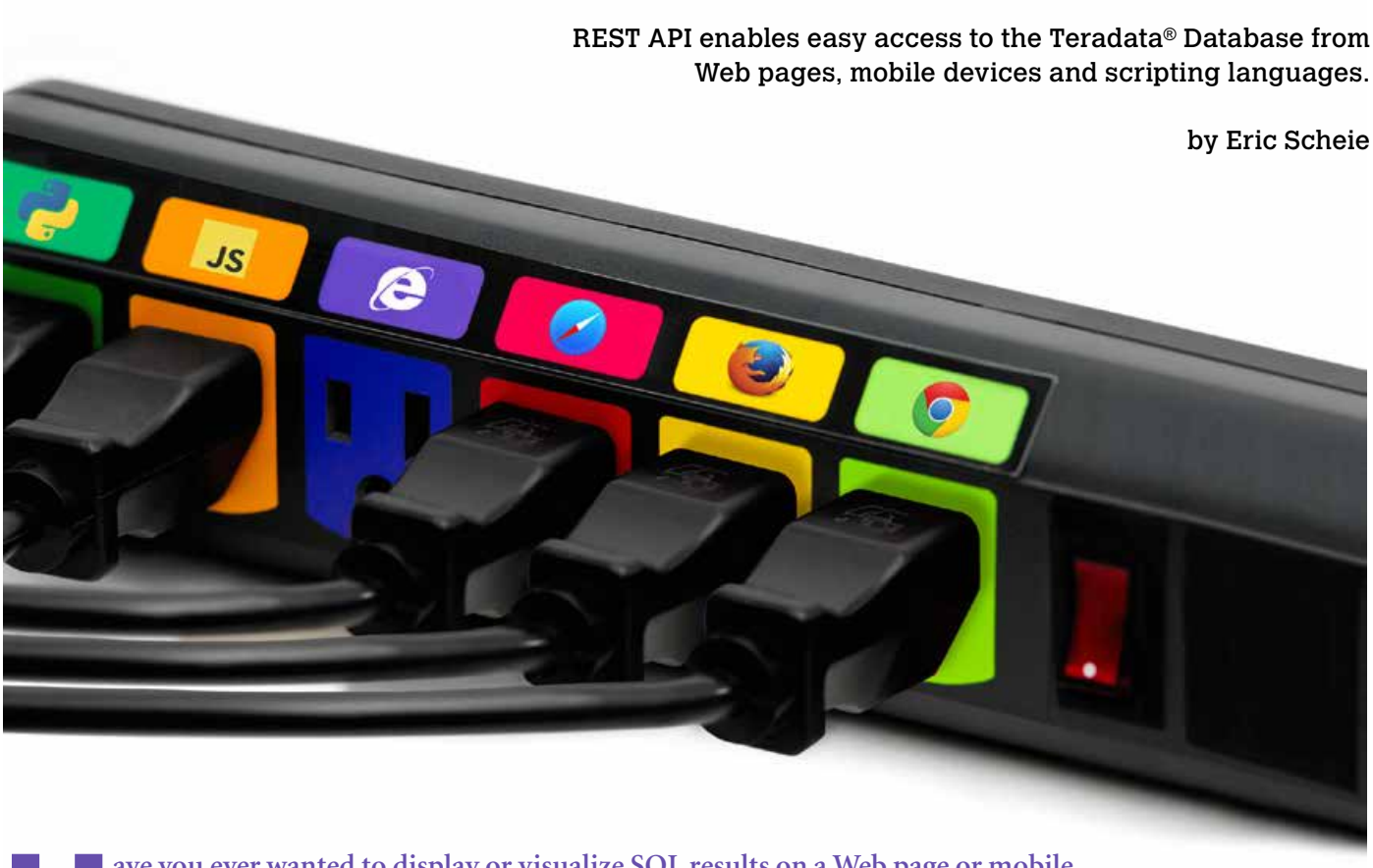
The Teradata® Software-Defined Warehouse, which is an enhancement to the Teradata Database, allows organizations to maintain multiple but separate data warehouses without sacrificing security or service level agreements. Companies can securely isolate data and users by creating "secure zones," each of which appears to users as if it's the only database in the system. Businesses can also manage system resources such as I/O, CPU and priorities for each zone. Plus, all this is possible through software controls.

2015
Teradata
REST

Driverless Connectivity

REST API enables easy access to the Teradata® Database from Web pages, mobile devices and scripting languages.

by Eric Scheie



Have you ever wanted to display or visualize SQL results on a Web page or mobile device? Or use a high-productivity scripting language to implement recurring business logic and deploy it without worrying about installing or maintaining a database driver? Implementing these use cases has not been easy since doing so requires building and supporting a middle-tier application to act as the translation layer between your application and the database.

Imagine, however, if you could use HTTP to query your database and get the results back in a standard format such as JSON. Well, you can. The REST API for Teradata® Database is a middle-tier Web service that accepts SQL requests in JSON through HTTP and forwards them to the Teradata Database using Java database connectivity (JDBC). (See figure.) When the SQL requests are complete, the results are converted into JSON (or alternatively, comma-separated values known as CSV) and streamed back to the requestor in the HTTP response. >>

Submit a Query

Before you can submit your first SQL statement, you need to register the target Teradata system. This can be done with the Service Configuration user interface or the REST API itself. When registering a system, you specify information including:

- Teradata Database hostname
- Database port number
- List of database users allowed to access the service
- Default session settings
- Session limits
- Alias used to refer to the system in future API calls

Once a system is registered, you can start submitting SQL statements. To do this, you send an HTTP request using the `POST` method on the `/tdrest/systems/{name}/queries` resource, replacing `{name}` with the alias of the system. HTTP requests are comprised of a request line, a series of request headers, an empty line and the message body:

```
POST /tdrest/systems/mysystem/
  queries HTTP/1.1
Accept: application/vnd.com.
  teradata.rest-v1.0+json
Authorization: Basic
  ZGJjOmRiYw==
{
      "query":"SELECT * FROM
        Customers",
      "format":"object"
}
```

In this example, you are submitting a `"SELECT * FROM Customers"` query. The first line is the request line and instructs the service that you want to submit a new query to the system named `"mysystem"`. The next two lines are HTTP headers. The Accept header is mandatory since it tells the REST API what version of the API you want to use. In this case, it's version 1.0. This is important because if the API changes in the future, you can continue to work with that version.

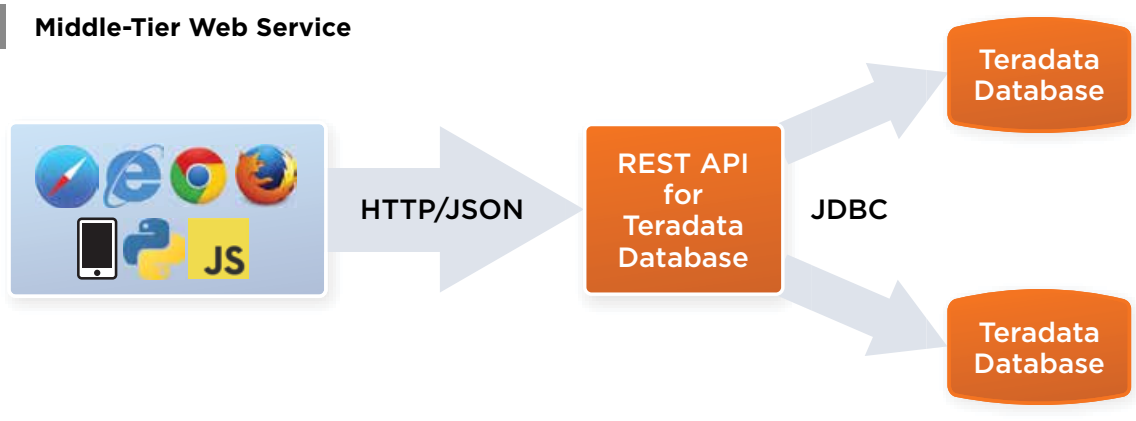
The Authorization header is also mandatory because it specifies the credentials that are used to access the Teradata Database. The REST API leverages HTTP Basic authentication for passing credentials.

The remaining part of the request is the message body. The message body is a JSON object that contains the mandatory `"query"` field with the SQL that will be executed. It can also contain other optional fields such as the ID of the session to use to execute the query, query bands to associate with the query, the desired result set format, and timeout and row limit options.

Of course you won't have to craft HTTP requests by hand since HTTP APIs are available to do it for you. This is an example of a query submitted from a Web page using JQuery, a popular javascript library:

```
$.ajax({
      type: "POST",
      url: "https://myhost:1080/
        tdrest/systems/mysystem/
        queries",
      contentType:
        "application/json",
      headers: {
         'Accept':
           "application/vnd.
           com.teradata.
           rest-v1.0+json",
         'Authorization':
           'Basic ' +
           btoa(username + ':'
           + password)
      },
      data: JSON.stringify({
         query: "SELECT * FROM
           Customers",
         format: 'object'
      })
});                                >>
```

FIGURE **Middle-Tier Web Service**



The REST API for Teradata® Database is a Web service that accepts SQL requests via HTTP and sends them to the Teradata Database using Java database connectivity (JDBC).

Get the Results

Query results for the original request are returned in the HTTP Response. The `"format"` field in the request determines how they are returned. Three formats are supported:

- `"object"` is the default format and causes results to be returned as a JSON object per row in which the column name is the field name and the column value is the field value.
- `"array"` has results returned as a JSON array per row where each column value is an element in the array.
- `"csv"` is for a comma-separated values option in which results are returned as a separate line for each row with commas between the column values.

(See the "`"Object"` Format Query" sidebar for an example of a response to the query in `"object"` format using data from the "Customer Information" table.)

The information returned from the query can be found in the `"results"` field. This field contains an array of results for each query submitted. Since only a single `SELECT` query was submitted, the `"results"` array contains only one element. The `"data"` field contains your query's results and is an array of JSON objects since the `"object"` format was requested. If the `"array"` format had been selected, the data field would contain the following JSON array instead:

```
"data":[[1,"Alice","10795
        Via Del Campo","San
        Diego","USA"],
        [2,"Bob","221B
        Baker Street",
        "London","UK"]]
```

Development and Deployment Made Easy

The REST API for Teradata Database simplifies developing new types of applications that previously may not have been possible. Creating Web and mobile displays can be streamlined, with no time needed for building a middle-tier Web service. Users specializing in front-end Web development can build entire applications without any dependency on server-side development, reducing both the skills and the number of people required to build an application.

The deployment of applications can also be simplified. Since support for HTTP and JSON in programming languages is practically ubiquitous, applications can be written that no longer depend on a database driver, allowing the applications to be deployed anywhere easily. When it comes time for the next database upgrade, there's no need to worry about having to upgrade database drivers either. You benefit now, and you'll benefit later, too. T

Eric Scheie joined Teradata in 2000 and is the chief architect responsible for the REST API for Teradata Database.

This article originally appeared in the Q1 2015 issue of *Teradata Magazine.*

"OBJECT" FORMAT QUERY

```
{
     "queryDuration": 45,
     "queueDuration": 3,
     "results": [
        {
           "data": [
              {
                 "Id":1,
                 "Name":"Alice",
                 "Address":"10795 Via Del Campo"
                 "City":"San Diego"
                 "Country":"USA"
              },
              {
                 "Id":2,
                 "Name":"Bob",
                 "Address":"221B Baker Street"
                 "City":"London"
                 "Country":"UK"
              }
           ],
           "resultSet": true,
           "rowCount": 2,
           "rowLimitExceeded": false
        }
     ]
}
```

TABLE **Customer Information**

Id	Name	Address	City	Country
1	Alice	10795 Via Del Campo	San Diego	USA
2	Bob	221B Baker Street	London	UK

NetApp®

Go further, faster®

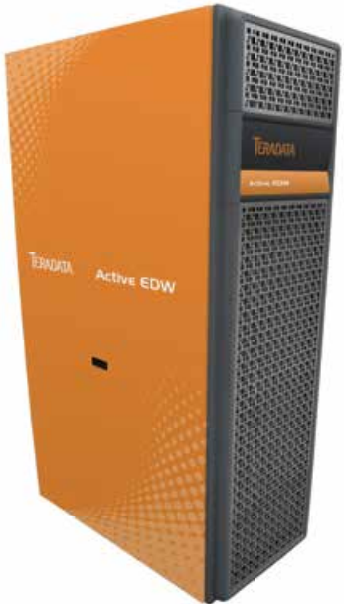
NetApp + Teradata =

The Best Performance for the Most Demanding Customers

Introducing the new Teradata®
Active Enterprise Data Warehouse 6800

- 2X SSD Read Performance
  = Faster access to critical data
- 25% More Computational Power
  = Faster time to market for business decisions
- 44% More Hot Data Capacity
  = Better, faster data analysis capabilities

Click for more information on the Teradata Active Enterprise Data Warehouse 6800.



TERADATA

About Teradata

Teradata helps companies get more value from data than any other company. Our big data analytic solutions, integrated marketing applications, and team of experts can help your company gain a sustainable competitive advantage with data. Teradata helps organizations leverage all of their data so they can know more about their customers and business and do more of what's really important. With more than 10,000 professionals in 43 countries, Teradata serves top companies across consumer goods, financial services, healthcare, automotive, communications, travel, hospitality, and more. A future-focused company, Teradata is recognized by media and industry analysts for technological excellence, sustainability, ethics, and business value. Visit Teradata.com.

NetApp

About NetApp

NetApp creates innovative storage and data management solutions that deliver outstanding cost efficiency and accelerate business breakthroughs. Leading organizations worldwide count on NetApp for software, systems and services to manage and store their data. Customers value our teamwork, expertise and passion for helping them succeed now and into the future. We focus our innovation to meet our customers' most challenging data needs. They recognize the value of our best-in-class integration and commitment to openness. Our knowledge and skills help them envision, deploy and evolve their IT environments. Discover our passion for helping companies around the world at www.netapp.com.

10000 Innovation Drive, Dayton, OH 45342 **Teradata.com**

10.15 EB7125