BIG DATA:
TERADATA UNIFIED DATA
ARCHITECTURE™ IN ACTION

TERADATA

TABLE OF CONTENTS

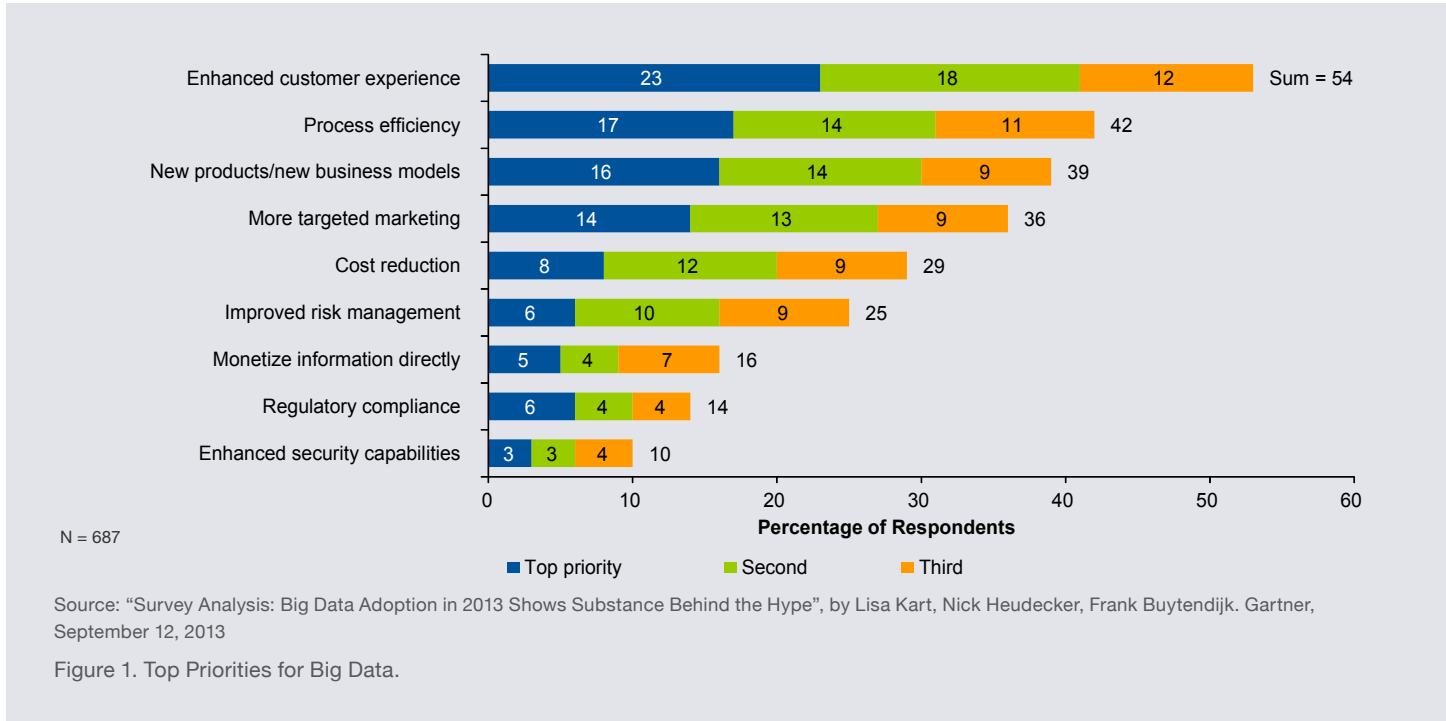## WHAT'S TERADATA'S PERSPECTIVE ON BIG DATA AND ITS VALUE?

Teradata believes big data is real, has value, and will change the speed and quality of business insights. We also believe it is an evolution, not a revolution, and that incorporating big data into your existing IT framework is fairly straightforward, but there are pitfalls we can help you easily avoid.

Is Big Data Real?

Yes. From a Teradata perspective, big data is a phenomenon we addressed with our customers as part of our ongoing product innovation process over the past two years. From a Strengths, Weaknesses, Opportunities, Threats (SWOT) perspective, we were able to turn big data from a threat into an opportunity. The opportunity is for our customers to beat their competitors by leveraging Teradata expertise to stay ahead of technology trends. It meant making some product acquisitions (such as Aster Data), creating new partnerships (Hortonworks, for instance) and doing a considerable amount of product and services integration work. The big data trend fits with the multi-year analytical ecosystem evolution of our "centralized data warehouse" approach to data warehousing. We recognize that there are a variety of cost-effective best-of-breed data and computation platforms, and that our primary value to our customers is to provide flexible, well-integrated options for capturing data, generating insight, and taking action.

Does Big Data have Value?

Yes, but so far it's mostly "at the margin," meaning the analysis of new data sources and types can improve what you are already doing. Based on Teradata customer uses, we see three broad categories—customer experience/service, operational efficiency, and corporate marketing[1]—that can be made 5 to 20 percent better by exploiting big data to glean deeper insights. Gartner[2] surveyed 720 IT professionals across industries in early 2013 and asked what top three business areas their big data projects targeted. Their results are shown in Figure 1.

TERADATA®

N = 687

Source: "Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype", by Lisa Kart, Nick Heudecker, Frank Buytendijk. Gartner, September 12, 2013

Figure 1. Top Priorities for Big Data.

Is Big Data an Evolution or a Revolution?

While some vendors go out of their way to over-hype this technology as revolutionary, we believe big data is a natural evolution of several trends already happening with traditional data. Specifically, three trends with big data stay the same:

~ **Same #1. The goal of using data and analytics as a competitive weapon.**

It has been proven by Tom Davenport and MIT[3,4] that companies with a data-decision culture are 5 to 6 percent more productive, with associated improvements in asset utilization, ROE, and market value. But is more data the goal? No. While size isn't really a concern since Teradata was architected for massive amounts of data and parallel computation, we recognized years ago that all data is not created equal. The placement of data on the most cost-effective layers of a storage hierarchy is critical because it contributes to healthy ROI.*

~ **Same #2. The need for people using analytical and visualization tools to glean insights from data and accelerate the rate of generating insights.**

You will always need skilled users of the tools, whether we call them data scientists or business analysts. Bottom line: the data has zero value unless you have the people, processes, and tools to transform it and access it in timely and relevant ways. The good news is that the tools are getting easier to use, and the number of pre-built analytic modules that plug-and-play to create new insights is growing rapidly. But that has been the trend for years.†

~ **Same #3. The substantial effort to drive insights into production.**

While insights are great, they are not enough. You can't have a science experiment in the basement, in isolation from the rest of the company. This "operationalization problem" is the Achilles heel for big data projects

For technical readers, we'll add occasional technical footnotes that can be skipped by business readers.
* Teradata's novel technical approach is called "Teradata Virtual Storage" and automates the intelligent storage of data based on usage. Our fast interconnect (high throughput Teradata BYNET™ V5 on InfiniBand) ensures that combining and moving data across systems is also cost-effective in the architectures we build.
† Teradata has built our libraries of functions in two ways. The Aster Data acquisition brought us 80 new analytic and visualization functions specifically for big data. We have invested in partnerships with BI vendors like SAS, IBM SPSS, Revolution Analytics, Zementis and Information Builders to ensure we have the best-of-breed analytics on top of the data platforms. Our recent partnership with Fuzzy Logix added another 600 functions. Finally, we have also partnered with visualization vendors like Tableau and Tibco Spotfire.

3

because most people grossly underestimate the costs of taking insights into production, which can be orders of magnitude larger than initial big data project insight generation costs.*

Bottom line: there is a certain constancy to the problems being addressed with big data. The age-old questions of how to increase profits, decrease costs, improve manufacturing efficiency, and curtail fraud have been around for a long time and are still being asked.

So what's new? We can address these questions in newer, faster ways that would previously have been too expensive or take too much effort. So that leads to the four things with big data that are different:

~ **Different #1. The amount of data you can economically collect is going up rapidly.**

Often called "data exhaust," things left in the data center corner—like digitized phone calls to customer care, or Web server detailed logs, or even external data like tweets—can now be captured and analyzed to create new value-added insights. With ever-decreasing storage prices, this changes the economics but requires more sophistication about how much data to capture, where to transform and store it, how long to retain the data, and more.

~ **Different #2. New analytics on these new, non-relational data types (coupled with existing relational data) open up exciting new insight possibilities.**

Example: you can create new customer attrition factors for people who are angry and shouting, or using code words like "rip off" when speaking to Customer Service reps. Further, you can combine analytics on various data types. For example, are those angry customers all located in one geography? Subject to new contract terms and conditions? Experiencing recent customer service issues? Discovering root cases may require combinatorial analytics on voice, text, geospatial, temporal, and traditional customer data—as well as the ability to do this quickly and seamlessly.

Other examples: Teradata customers are exploiting hail damage and weather reports to find agriculture fraud; grabbing Twitter feeds to discover unhappy customers in real time in order to react quickly; and monitoring email investment promises by brokers to clients.

~ **Different #3. The process for generating insights is changing.**

The "old way" of doing analytics assumed that new data would be stored in the data warehouse. Hence many processes that precede the discovery step required rigor in data modeling, data integration, and ETL. The "new way" separates discovery from integration and use, freeing up some time until after insights have been generated and explored. In short, it frees up more time for analysis and exploration.

Example: a fraud team thinks that perhaps a fraudster who has stolen someone's identity spends a different amount of time—and clicks in different ways—on their banking web site from the true owner of an account. Capturing web clicks and time durations, and the sequence of clicks, they create a "normal pattern," then watch for deviations to that pattern in end-end experiments. They grab all the web click data they require for experimentation but do not need to model the data rigorously or design production fraud patterns until they are certain they have found an interesting and worthwhile case of fraud.

~ **Different #4. One kind of processing engine (e.g., just an EDW) is not going to be enough—the architecture must evolve to accommodate specialized platforms, integrated into a comprehensive system.**

Just like in your garage, you will need a workbench that includes a variety of tools. In the data analytics world this means workload-specific engines/platforms for storage and analysis, new type-specific analytic functions and engines, and integration technologies to copy and move data or insights into an easy-to-use environment for your business and tech analysts. Gartner calls this the "Logical Data Warehouse," and we call this the Teradata Unified Data Architecture.™ Teradata customers like Intel® agree and have headed in this direction.[5]

---

*It's an area where Teradata has traditionally excelled: rock-solid architecture and engineering, all the technologies and services that you need (with straightforward extensions for big data projects), including things like data synchronization, data governance, workload management, backup and recovery, system management, metadata management, configuration planning—many "real-world" requirements that are NOT easy to put in place on your own. Yes, an architecture for big data looks complicated but the value from Teradata is the integration of our (and partner technology) components to reduce the time to operationalize insights. And it's engineered with flexibility in mind so new components, new partners, can be included while maintaining control of IT and ensuring ROI.

TERADATA®

TIP: ASK PROPONENTS OF BIG DATA PROJECTS THREE QUESTIONS:

1. What business areas will improve because of the collection, analysis, and use of big data?

2. What's the incremental Return on Investment (ROI) from exploiting more data? How do you plan to calculate this?

3. Will our environment support the generation of more useful insights and actions faster than our competitors?
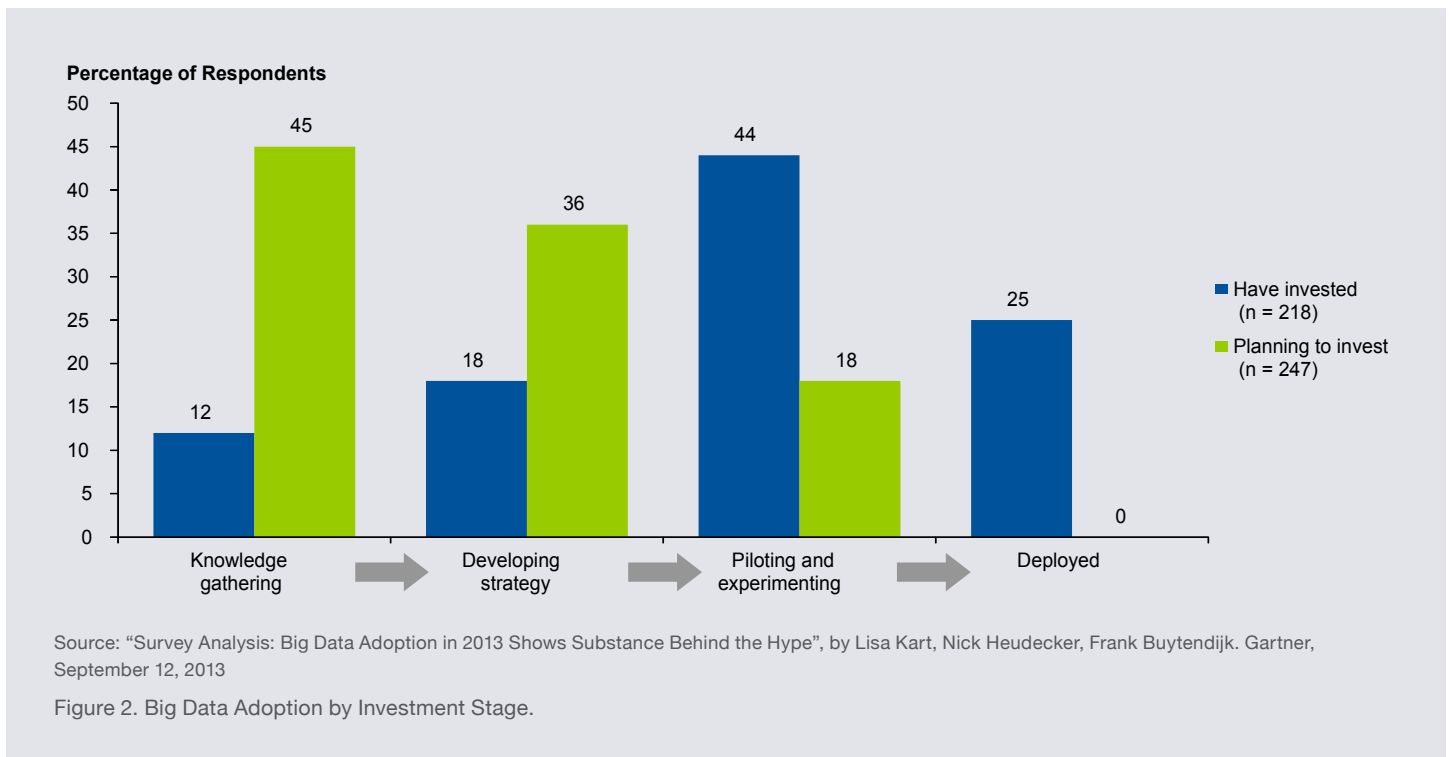
Bottom line: the economics of how to capture data and where to store it, where to compute on it, and how to use it, are constantly changing. The discovery process is accelerating with the help of new toolkits to generate insights without the usual rigor required by an EDW.

And as a consequence, Teradata has extended its architecture to accommodate the needs of big data.

Who is adopting big data, and am I late to the party?

The same Gartner survey[2] also tried to gauge how far along respondents' companies were in planning and executing big data projects. Were they gathering knowledge about big data? Developing strategy? Doing pilots? Already in production? Figure 2 shows the results for both those companies who have already invested, and those planning to invest.

A quick estimate of Teradata customers is that about 15 percent have a big data project underway (mostly using Teradata Aster® and/or Hadoop with Teradata Database), and another 20 to 30 percent are considering projects. Early adopters who have finished projects, believe they have achieved a one- to two-year competitive advantage. Some of these have five or more big data projects under their belts. Most leading companies are operationalizing their first projects now, while the laggards are still in learning mode. However, it's never too late to catch up.

Source: "Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype", by Lisa Kart, Nick Heudecker, Frank Buytendijk. Gartner, September 12, 2013

Figure 2. Big Data Adoption by Investment Stage.

WHAT IS TERADATA UNIFIED DATA ARCHITECTURE™ AND ITS VALUE?

Just like hybrid cars get superior gas mileage from coupling traditional gas engines with electric batteries, the Teradata Unified Data Architecture™ integrates three types of engines to provide the best fit-for-purpose analytical capabilities for any kind of data.

Teradata has been in the big data market longer than anyone, so we've leveraged our expertise to tackle the "what's new" part of the big data phenomenon with five parallel engineering activities. Beginning with our core product, the Teradata Integrated Data Warehouse, we:

1. **Defined a new architecture** called the Teradata Unified Data Architecture™ that adds a discovery platform and a data platform to complement the Teradata Integrated Data Warehouse. In the Teradata advocated solution, the discovery platform is Teradata Aster, while the data platform can either be Hadoop or a Teradata Integrated Big Data Platform for large, cost-effective storage and processing of big data.

2. **Engineered new access connectors** between the platforms and to external sources of big data, and extended our hardware platform portfolio and inter-connect options to provide even faster data transfers.

3. **Created a library of pre-built Teradata Aster analytic modules to speed up and simplify discovery**, all in an easy-to-use Teradata Aster SQL-MapReduce® programming paradigm. We also continue to add more technical partner analytics to the Teradata Aster and Teradata engines.

4. **Extended our traditional systems management products** (Viewpoint, Studio™) for similar use with Hadoop and Teradata Aster.

5. **Developed new Professional and Customer Service offers** to help customers quickly design and deploy enterprise data architecture and analytics projects.

Teradata Unified Data Architecture™

We'll divide the explanation of the engineering work into two parts: logical and physical architectures.

The *logical* view of Teradata Unified Data Architecture™ in Figure 3 shows the flow from left to right of data to
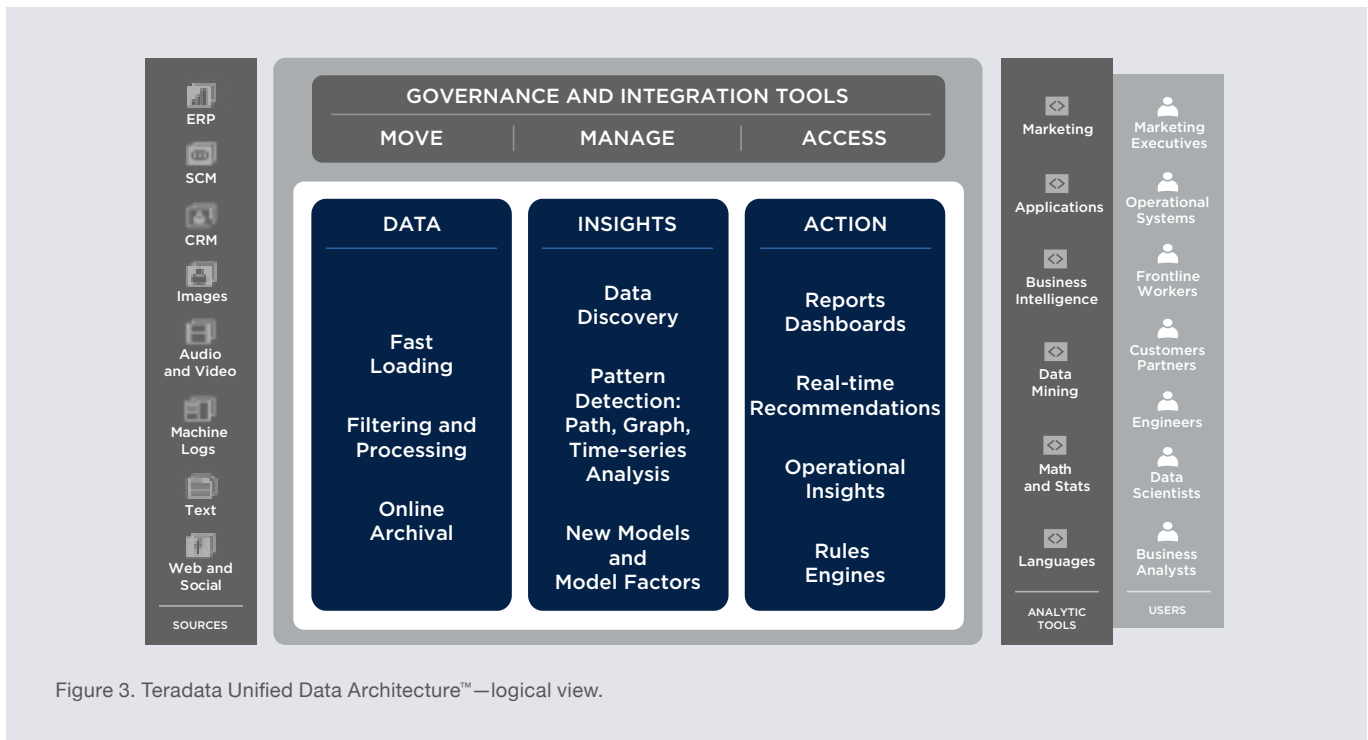

Figure 3. Teradata Unified Data Architecture™—logical view.

insights to action that are required to process big data. Data sources appear on the left (both traditional and new sources), and consumers of data and insights appear with their end-user tools on the right.

This picture is meant to be vendor-agnostic because Teradata Unified Data Architecture™ is a toolkit. The logical flow also shows a block for the governance and integration functions needed for operationalization, including policies for moving, managing, and accessing data and insights.

Figure 4 shows the advocated *physical* implementation of Teradata Unified Data Architecture™ with three engine categories—data platform, integrated data warehouse and discovery platform. While any of these platforms could be used for loading and storing raw data from a wide variety of sources using many different ETL tools,

new economics drive toward the use of Hadoop for land-ing, transforming, and storing big data (as opposed to traditional data, which continues to flow into Teradata).[6] Some people call this a "data lake". Another option if you want the power and simplicity of Teradata is to use the new high-capacity Teradata Integrated Big Data Platform (IBDP).

The Teradata Unified Data Architecture™ also provides three options for where to do discovery. Again, doing the analysis depends on the economics of computation, staff skill sets, and big data discovery and visualization tools, as well as access to information that may have been stored in other systems (or in an external source). New economics and discovery process changes drive us to advocate the Teradata Aster Discovery Platform as the big data discov-ery platform.
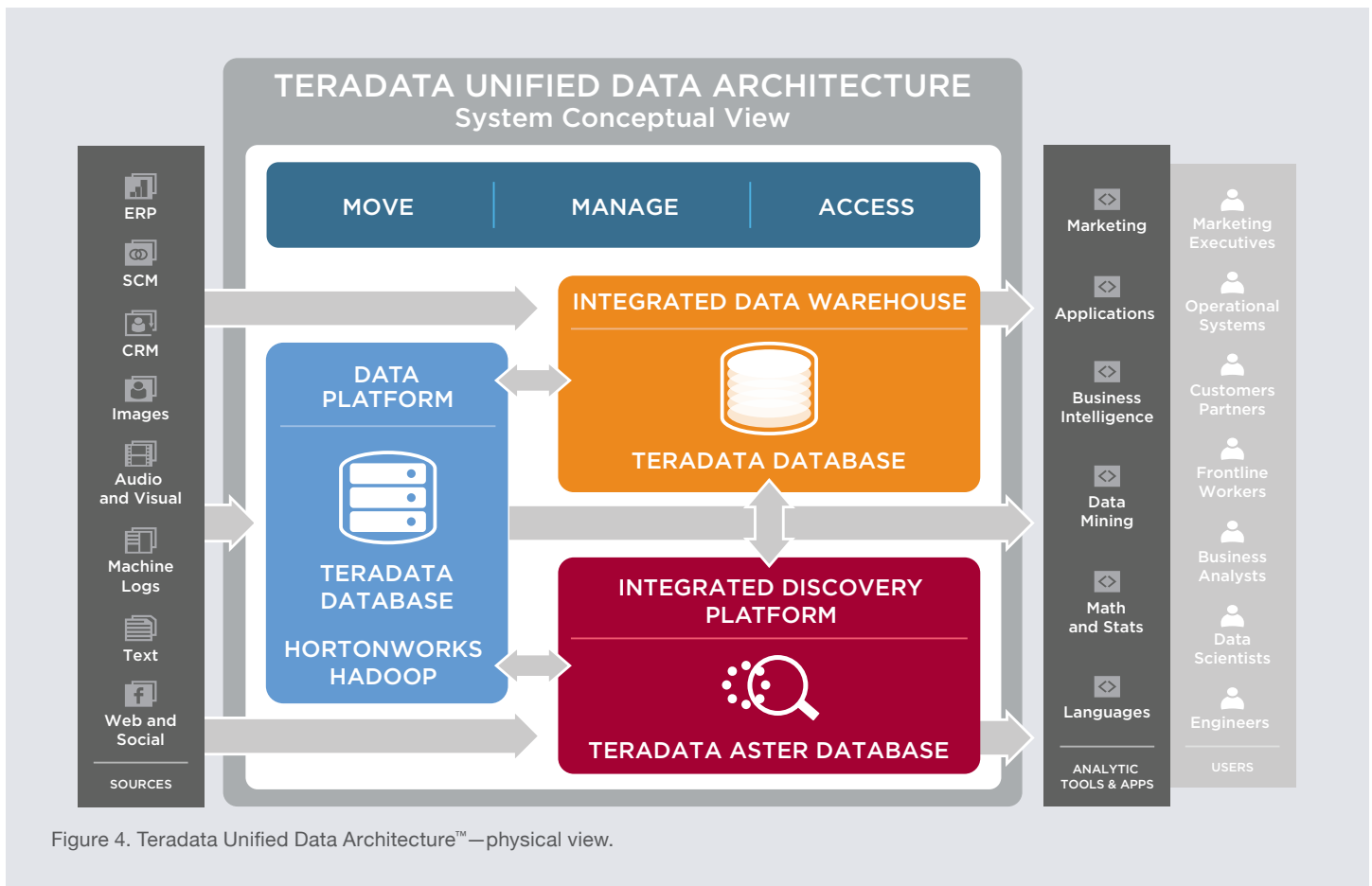

TERADATA UNIFIED DATA ARCHITECTURE
System Conceptual View

MOVE | MANAGE | ACCESS

INTEGRATED DATA WAREHOUSE

DATA PLATFORM

TERADATA DATABASE

TERADATA DATABASE

HORTONWORKS HADOOP

INTEGRATED DISCOVERY PLATFORM

TERADATA ASTER DATABASE

SOURCES: ERP, SCM, CRM, Images, Audio and Visual, Machine Logs, Text, Web and Social

ANALYTIC TOOLS & APPS: Marketing, Applications, Business Intelligence, Data Mining, Math and Stats, Languages

USERS: Marketing Executives, Operational Systems, Customers Partners, Frontline Workers, Business Analysts, Data Scientists, Engineers

Figure 4. Teradata Unified Data Architecture™—physical view.

TERADATA®

Once insights are created, they need to be used to have value. Again, there are numerous choices of how to connect the data and insight generation engines shown here to action-taking systems, like a CRM system, a contact center, a website rendering engine in real-time, dependent data marts driving other apps, or even an executive dashboarding system. The advocated solution here is the Teradata Integrated Data Warehouse, well known for its ability to scale to thousands of BI analysts (and millions of consumers), running a variety of applications as well as BI tools for the end consumers of big data (and traditional data). With this understanding of the advocated pieces of Teradata Unified Data Architecture™ let's drill a level deeper into the components.

Teradata Aster

In 2011, Teradata acquired Aster Data with its patented SQL-MapReduce®-based discovery platform, useful

to data scientists and business analysts who want to get faster insights on combinations of traditional (i.e., relational) and non-traditional data (essentially any type of data), also known as "multi-structured data." A typical use of the Teradata Aster Discovery Platform is to pull information from a variety of sources (e.g., raw web logs from web page rendering systems, machine data from factory floor systems, and traditional data from Teradata) to do investigative analytics, to find new patterns of customer behavior, or new aberrations in manufacturing data that might indicate quality control issues.

Teradata Aster is based on a patented SQL-MapReduce® implementation which makes it easy for BI analysts who know SQL to write extensions that use four classes of built-in analytical functions (see Figure 5) to very quickly capture, prepare, analyze, and visualize big data sets.
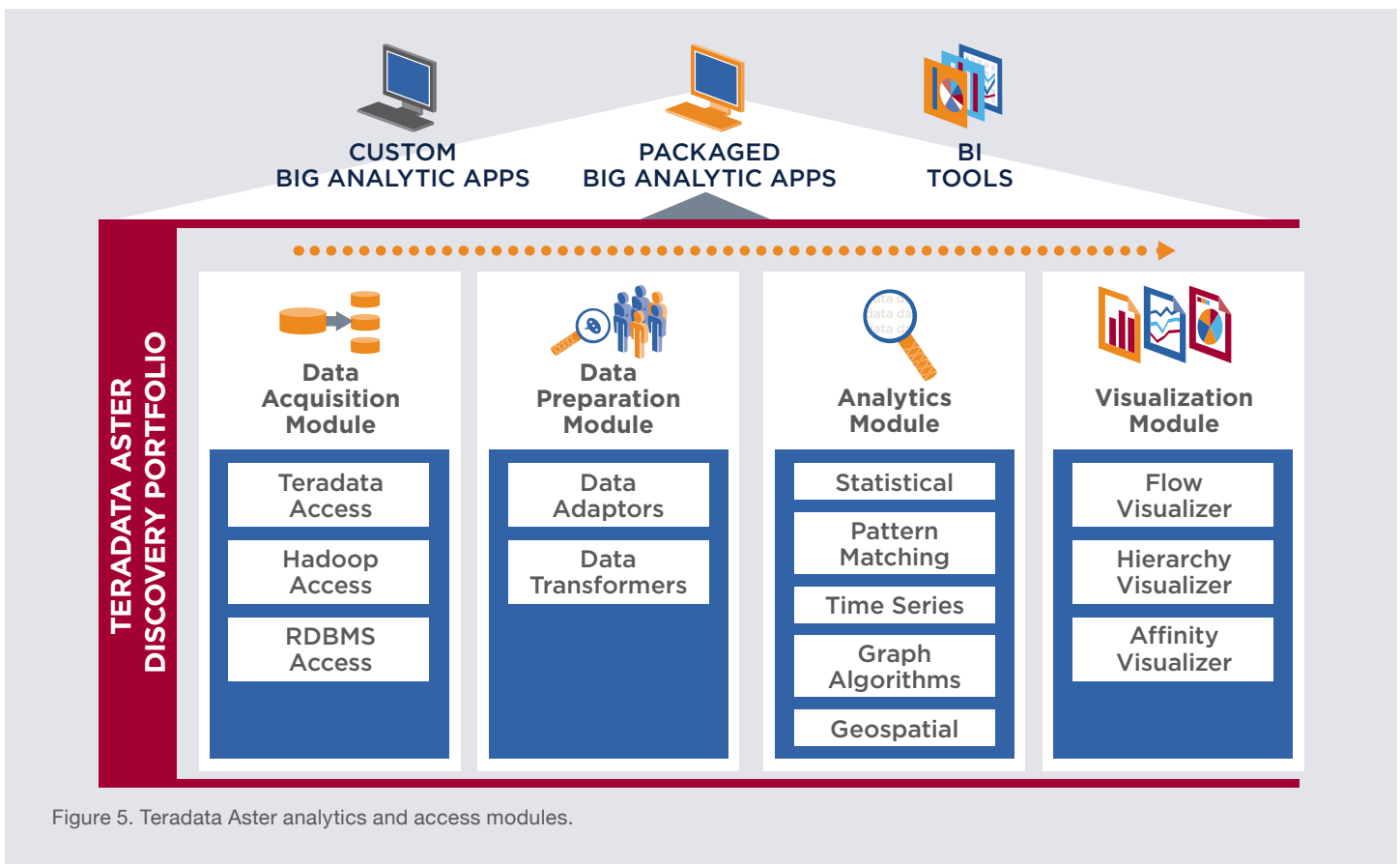

Figure 5. Teradata Aster analytics and access modules.

Hadoop

In 2013, Teradata announced the Teradata Portfolio for Hadoop, providing another data capture/storage/refining platform (leveraging our partners Hortonworks and Cloudera). The initial focus is on wide, cheap capture of "data exhaust" for possible stand-alone use, but more likely as a source or ETL system to the Teradata Aster Discovery Platform or Teradata Integrated Data Warehouse, augmenting the volume and types of data that can be included in analysis.

Hardware Platforms

From a hardware perspective, adding Teradata Aster and Hadoop augmented the number of options from which to choose, as shown in Figure 6. Customers can even buy workload-specific appliances that are configured with combinations of Teradata Aster and Hadoop nodes.

Teradata also added a fabric-based high throughput computing interconnect called Teradata BYNET™ V5 on InfiniBand, for very high-speed data movement and fault tolerance among the various platforms. This enables fast interchange of data to the most optimized platform for specific storage, processing, or analysis requirements.

Data Connectors

Connecting these three systems required that Teradata build bulk data connectors as well as specialized connectors for on-the-fly data access, e.g., Teradata SQL-H™ and Teradata Aster SQL-H™ for accessing data in Hadoop, and an Aster-Teradata Connector.

Pre-built Analytic Models

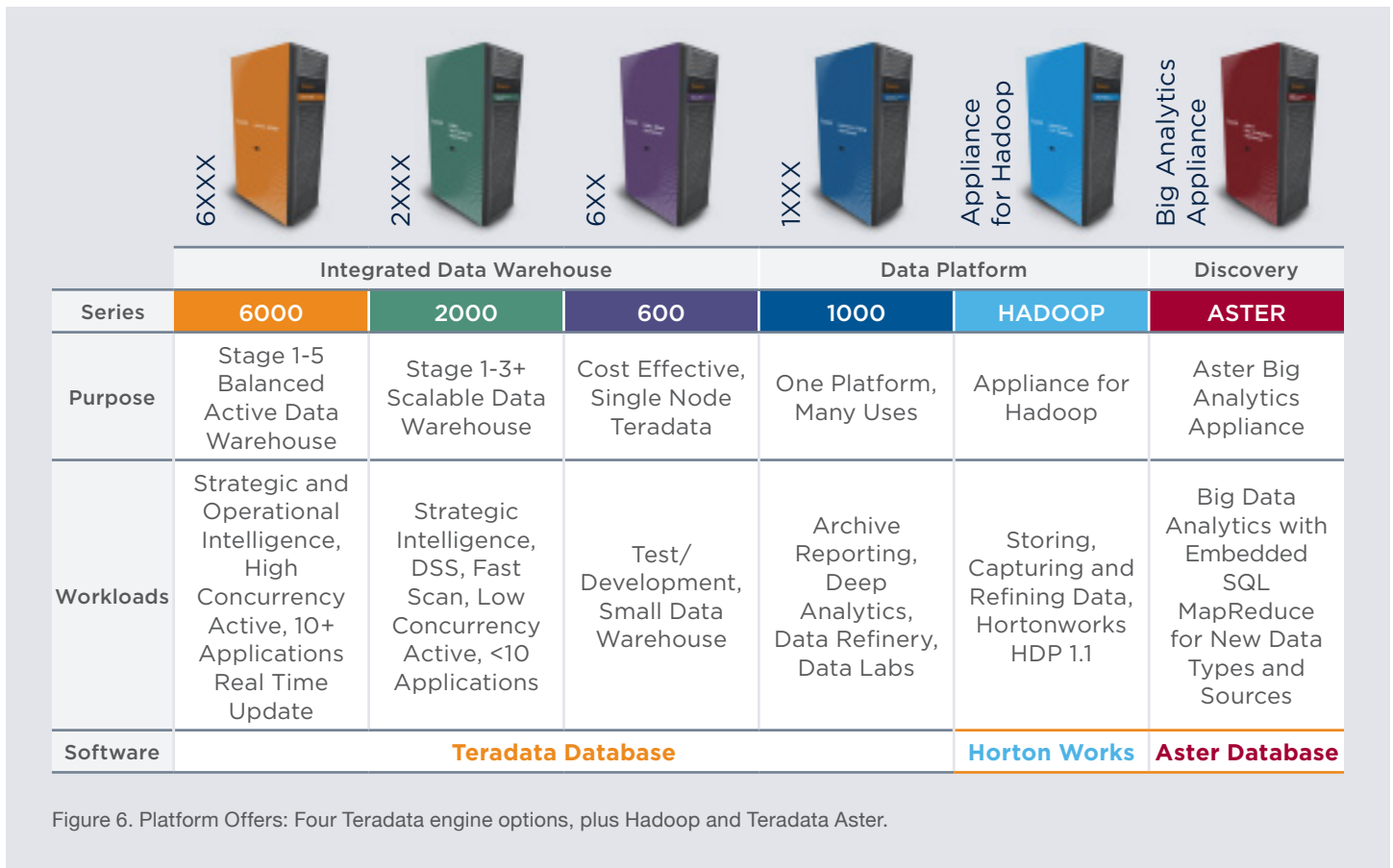Our goal is to make insight generation fast, despite the increasing amounts and types of data. To do that, we

	6XXX	2XXX	6XX	1XXX	Appliance for Hadoop	Big Analytics Appliance
	Integrated Data Warehouse			Data Platform		Discovery
Series	6000	2000	600	1000	HADOOP	ASTER
Purpose	Stage 1-5 Balanced Active Data Warehouse	Stage 1-3+ Scalable Data Warehouse	Cost Effective, Single Node Teradata	One Platform, Many Uses	Appliance for Hadoop	Aster Big Analytics Appliance
Workloads	Strategic and Operational Intelligence, High Concurrency Active, 10+ Applications Real Time Update	Strategic Intelligence, DSS, Fast Scan, Low Concurrency Active, <10 Applications	Test/ Development, Small Data Warehouse	Archive Reporting, Deep Analytics, Data Refinery, Data Labs	Storing, Capturing and Refining Data, Hortonworks HDP 1.1	Big Data Analytics with Embedded SQL MapReduce for New Data Types and Sources
Software	Teradata Database				Horton Works	Aster Database

Figure 6. Platform Offers: Four Teradata engine options, plus Hadoop and Teradata Aster.

TERADATA®

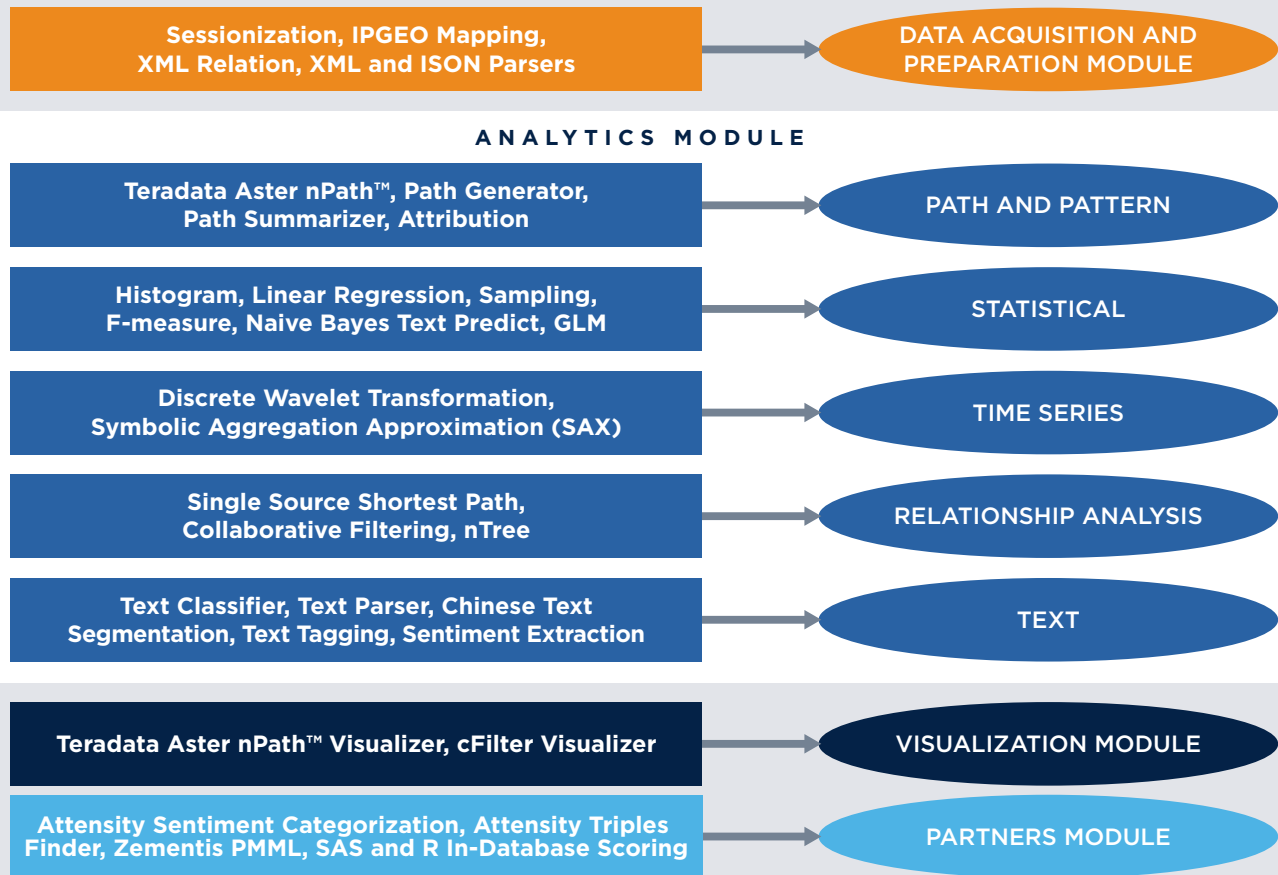## SAMPLE TERADATA ASTER SQL-MAPREDUCE® FUNCTIONS BY CATEGORY

Sessionization, IPGEO Mapping, XML Relation, XML and ISON Parsers → **DATA ACQUISITION AND PREPARATION MODULE**

ANALYTICS MODULE

Teradata Aster nPath™, Path Generator, Path Summarizer, Attribution → **PATH AND PATTERN**

Histogram, Linear Regression, Sampling, F-measure, Naive Bayes Text Predict, GLM → **STATISTICAL**

Discrete Wavelet Transformation, Symbolic Aggregation Approximation (SAX) → **TIME SERIES**

Single Source Shortest Path, Collaborative Filtering, nTree → **RELATIONSHIP ANALYSIS**

Text Classifier, Text Parser, Chinese Text Segmentation, Text Tagging, Sentiment Extraction → **TEXT**

Teradata Aster nPath™ Visualizer, cFilter Visualizer → **VISUALIZATION MODULE**

Attensity Sentiment Categorization, Attensity Triples Finder, Zementis PMML, SAS and R In-Database Scoring → **PARTNERS MODULE**

Figure 7. Teradata Aster pre-built analytics functions.

have rapidly grown the number of Teradata Aster pre-built functions to more than 80 (see Figure 7; for more details on each function, see the Teradata Aster Discovery Portfolio white papers[7] listed in the Suggested Reading section). These can be used in isolation, or in various combinations, to enable "multi-genre analytics" and look at data through different perspectives to improve accuracy of models and algorithms for churn, fraud, customer segmentation, and more.

In addition, we have worked with partners like SAS, IBM SPSS, Revolution Analytics, Zementis, and Fuzzy Logix,™[8] leading text vendors like Attensity, and geospatial vendors like ESRI, to increase the number of Teradata-based analytic modules that often use in-memory (or in-database) capabilities to speed insight development on Teradata. The traditional BI approach of using a Teradata Data Lab on structured data has been extended to work in conjunction with the Teradata Aster Discovery Platform to provide a rich collection of opportunities for insight generation and integration.

TERADATA®

Management

As you add more components to the enterprise data architecture—and in particular, as you move from discovery to operationalization—a danger is that many more people might be required to monitor and manage the system. In our traditional Teradata systems, we drive lower total cost of ownership (TCO) by automating, and in many cases eliminating, many systems management functions, building in self-monitoring and tuning capabilities, and making DBA functions easier than the competition's. We hold true to that corporate goal with the Teradata Unified Data Architecture,™ bringing many of the rich Teradata management tools and world-class support infrastructure to the Teradata Aster Discovery Platform and the Teradata Portfolio for Hadoop.

Key technologies include Teradata Viewpoint, Teradata Vital Infrastructure, and Teradata Studio™ to name a few. The Teradata Unity portfolio for managing multiple Teradata systems is being enhanced to support Teradata Aster and Hadoop environments where the technologies fit, cost-effectively taking management of the Teradata Unified Data Architecture™ to the next level. Teradata also has comprehensive data governance offers, extending metadata and master data services to leverage the value and reliability of traditional data management across all three platforms. This gives organizations a trusted foundation for a data-driven culture. Bottom line: Teradata recognizes that handling big data requires many of the same production quality tools that traditional Teradata has provided for years, and this is a big part of our TCO value to our customers.

Training and Services

Our final piece of the Teradata Unified Data Architecture™ is our training curriculum along with Professional and Customer Services. We have hired experts with new skill sets in Hadoop, as well as training a considerable number of our Professional Services people to be data scientists. In engagements, they sit side-by-side with our customers, training them on the Teradata Unified Data Architecture™ technologies and providing leave-behind templates for projects and data discovery insight methods that can be easily replicated. Our Industry Consultants collect the latest and greatest customer case studies and can help evaluate project plans for big data efforts using our Industry Frameworks. And "taking action" on new insights—which is where the value lies—often depends on

discussions and data sharing with other groups, or integration projects. For example, Teradata customers are leveraging big data insights by connecting to any of a host of automated front-end systems like the Contact Center, the web group, the factory floor, or mobile dashboards. Planning, designing, deploying, and maintaining the environment—all these activities have corresponding Professional Service offers that have been extended to cover big data needs. This comprehensive view of big data and how it is operationalized in your business sets Teradata apart from other technology vendors—and gives you "one throat to choke" if you have problems and "one hand to shake" with success.

HOW DOES TERADATA UNIFIED DATA ARCHITECTURE™ WORK?

The primary value of Teradata Unified Data Architecture™ is to convert data—big and small, and all combinations—into useful, actionable insights. This involves analytical approaches designed to uncover previously unknown patterns, or the identification of key events that trigger customer behaviors like decisions to buy products or cancel contracts. In this section, we'll bring the Teradata Unified Data Architecture™ to life by showing "A Day in the Life" of two customer projects from various industries, highlighting various implementation details.

SCENARIO 1: RETAIL
MULTI-CHANNEL INSIGHTS DRIVE BETTER MARKETING AND CATEGORY MANAGEMENT

Several large retailers use Teradata's Unified Data Architecture™ to help their marketing and category managers better understand consumer behavior on their web sites, on mobile devices, and in the stores. In this composite case illustrating several customers' best uses, these retailers already had Teradata systems that contained details about millions of customers, sales, and products. Their big data interest was in adding web and mobile data, specifically trying to understand the impact of marketing campaigns, and couponing efforts to drive larger market baskets or improve overall margins.

Data

One retailer began their proof of concept (POC) by loading very detailed web server log records into a Hortonworks Hadoop cluster. This raw data transfer was fast; they found they could load two months of web log

TERADATA®

in one hour (about 60 times faster than their traditional approach). One day's worth of data included about 30 million web click entries, each involving up to hundreds of attributes that needed to be parsed.

Sessionization

Hadoop was further used to sessionize the data from multiple log sources to create integrated customer visits. Since materializing each page hit might cause multiple log records to be written to several servers, there are often from 5 to 30 records that need to be spliced together to show what's happened on any particular page. Once this is done, then the temporal sequence of pages can be reconstructed for any particular customer. Splicing this together with customer store data on market baskets provided a more useful data set of each customer's behavior.

Analytics

These sequences were then accessed using HCatalog by Aster for further analysis based on path matching (using the pre-built Teradata Aster nPath™ function). Their engineers built a tool so users could easily write SQL-MapReduce queries to do path matching at a higher semantic level. For example: "find all customers who looked at products in the category 'Lawn Furniture' and clicked three times in that category but never purchased anything," then drill down into what those customers were looking at, specifically, to get deeper insights about product interest. They could also cross-correlate purchasers versus non-purchasers from referral sites (web sites with ads driving customers to the web site), so they could see which ad placements drove the more profitable customers to their web site. They could also compare the web pages that purchasers and non-purchasers were browsing, to see again the differences—all providing levels of insight the category managers could never get before. Finally, by accessing the Teradata data warehouse, they could tie this customer analysis together with previous store visits, purchases, lifetime value, response to promotions of various types, and other customer scores already in their existing database to better calibrate future marketing campaigns.

Another Teradata customer focused on omni-channel behaviors, specifically email campaigns that drove (or did not drive) conversions using coupons in the stores. They drilled into finding which people bought/searched/clipped coupons for item A, and then ran Aster's affinity and clustering analytics to reveal deeper insights about what combinations of items are attracting or repelling coupon-oriented shoppers. Clustering customer groups by behavior then provided the clues they needed, for a particular shopper, to match "her historical baskets" and "other coupon shoppers like her" to come up with better personalized recommendations and coupons for future emails (see Figure 8).

Yet another company also used the Teradata Unified Data Architecture to analyze which emails drove people to "un-subscribe" altogether from emails. For example, are there particular words or phrases that seem to cause this (discovered using text analytics), or are time-of-day, day-of-week (temporal), or number of recent emails driving un-subscription events?

Product A: BOUGHT
Product B: LOOKED
Product C: BOUGHT

Product B
COUPON for Product E

A Customer's "Historical Basket" **+** Shopping Patterns of Similar Shoppers **=** Personalized Recommendations

Figure 8. Matching Customer Behavior with Clusters to Drive Recommendations.

TERADATA®

Value to Users?

For the first time, category managers could easily "see" exactly how customers were behaving on the web site, and what were the "hot" and "cold" parts of the web site. Not just the "score" of a customer segment, but the actual browsing behavior of discrete individuals and micro-segments. They could see dwell times on various pages, and understand better which pages contributed to the pathway to purchases, as opposed to pathways that ended in dropouts (non-purchases), which helped drive rethinking of unproductive web pages. They could see which coupon offers worked and didn't.

Action

The faster insights using detailed web behaviors were used by this retailer to change web design layouts and content. They also help product category managers derive better insights about customers and better personalized offers. They could also recalibrate their spending on search engine keywords on other sites, and focus on unproductive email text and bad email timing.

The Business Value?

These are expected to increase market basket size and improve conversion rates of existing customers by five to ten percent and drive another eight to ten percent more browsers into the sales funnel through better targeting. The bottom line impact is tens of millions of dollars.

Project Time

These projects were completed typically in three to six weeks with one to two full-time analysts, usually one Teradata data scientist and one Teradata client BI expert, plus the sponsor and a little help from an ETL engineer.

SCENARIO 2: BANKING
IMPROVING THE CUSTOMER SUPPORT EXPERIENCE

All large banks and credit card companies receive millions of customer support requests per year. Like most consumer companies, they have tried to shunt inquiries to self-service mode on their web site to reduce the volume of phone calls coming into the customer contact center. The goal of these big data projects was to see if they could figure out why people were calling, especially when they had been on the web site just prior to calling. They had never had the ability to "see" the customer

data this way because it was on various systems—web data and analytics on one system (with outsourcing of some analytics), contact center data on another, interactive voice response (IVR) data on yet another, call center agent notes in yet another place, and customer transaction information on one more system.

Data*

Four to six months of data were loaded from five to seven source systems. 500M to 1.5B uncompressed web log records from 10-30M customer sessions on the web took 40-70 minutes to load into Teradata Aster Database. 800M to 1.3 billion IVR and call records from 5-15M contact center interactions took another 40-60 minutes, and 50-75 million call center text notes and complaint database records took typically less than 30 minutes.

Sessionization

Using time stamps, as shown in Figure 9, the IVR and call records and memos had to be matched up to create a complete record for each call, then merged with web sessions and sorted. They then looked for web sessions that were immediately followed by calls to customer support.
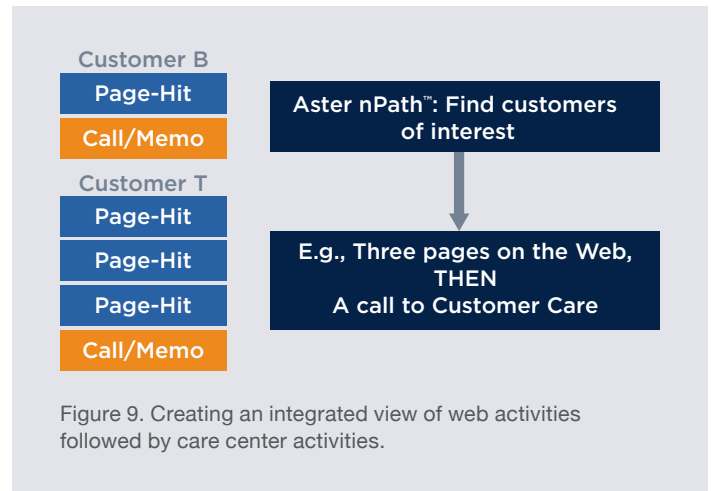

Customer B
Page-Hit
Call/Memo

Customer T
Page-Hit
Page-Hit
Page-Hit
Call/Memo

Aster nPath™: Find customers of interest

E.g., Three pages on the Web, THEN A call to Customer Care

Figure 9. Creating an integrated view of web activities followed by care center activities.

Analytics

Using Teradata Aster, they found that about 2 percent of the millions of web sessions drove subsequent calls to customer care, providing a base of several hundred thousand cases that they could then further investigate.

* Teradata has done multiple big data projects for this use case; numbers below are a composite or range of figures representative of "typical implementations."

TERADATA®

They created Teradata Aster nPath™ visuals like Figure 10 to show the top web pages people had clicked on right before calling, and mapped those to the CSR's "reason for call" codes.

Value to Users

The customer saw new insights about paths, specifically which website pages and IVR sequences seem to cause confusion and dropouts resulting in excessive calls to customer care.
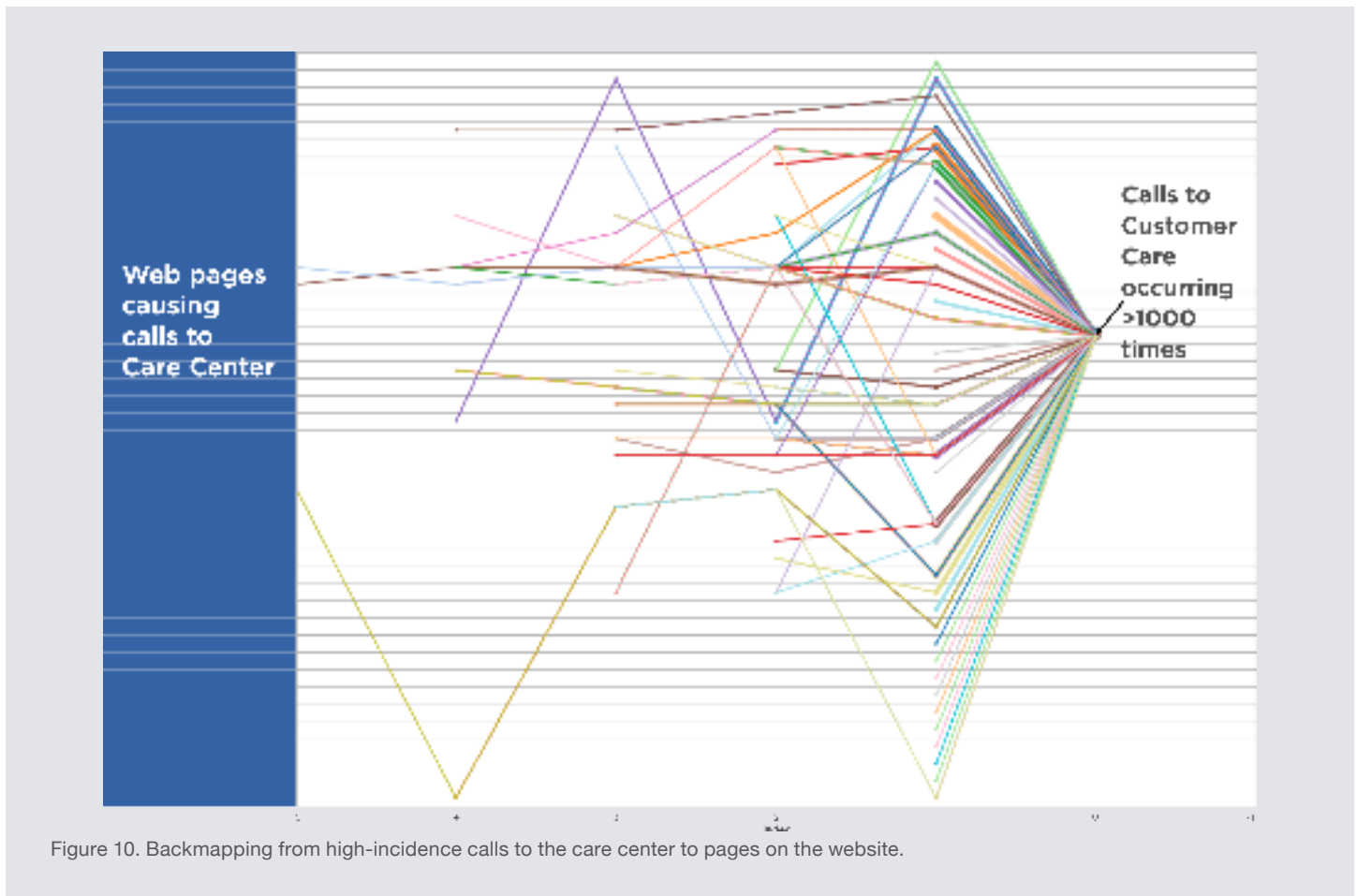
Action

The findings led to the redesign of several website pages and IVR button push sequences so consumers can resolve issues on their own in self-service mode.

Business Value?

Transfers between channels cost from five to fifteen million dollars annually, and repeat calls for unresolved issues cost almost the same amount. The big data project provided increased levels of self-service while decreasing cost of operations, better customer experience and satisfaction, and provided early identification of valuable customers with care issues who might defect. A rough projection of savings is cost takeout of 10 percent or roughly $1.5M-$3M from this big data project.

Project Time

These projects were completed typically in five to eight weeks with three to five full time analysts and a handful of part-time employees to help with data sourcing.

Figure 10. Backmapping from high-incidence calls to the care center to pages on the website.

TERADATA®

HOW CAN MY COMPANY GET STARTED?

This is an evolution not a revolution, so the same best practices apply to the project management of big data projects. It's about ROI, appropriate leadership and mentoring, staffing and training, as well as linkages between business and IT to focus on key aspects of the project.

The key area to focus on is the economics of opportunity gained or lost. Think like an economist. What business problems could be resolved with more/new/different data? What is the economic value of incremental data? What is the cost? Or more importantly, what is the economic value of new insights? Ask your business people questions like "what would you do if you knew," "what immediate actions could we take if we knew," and "how much value would that provide?" If you are managing a call center, what could you do better if you got an automatic hourly list of the people who are so angry in calls that this could be THE key factor in radically improving customer retention? ROI is about value versus costs. The costs are usually easy to obtain—costs of storage, software, tools, consulting services, and ongoing licenses. It's the value that's difficult to compute—and something that has to come from the business groups, not IT.

Surveying Teradata customers, it's interesting to see who launched the big data project organizationally. The champion at our first big data customer came from the business strategy group, not the line of business, and not the IT group. That's rare. Often, initiatives come from IT people supporting marketing groups—which is understandable since many Teradata implementations are marketing-support systems. In all the successful cases though, they had a business sponsor, picked a problem that was big enough to make the results interesting/useful/immediately of value, but not so large as to overwhelm the group doing the work. By and large, we've been delighted with how quickly the POCs have gone (often only four to six weeks) and it's an opportunity for our people to help train yours. The power of technology is easily proven—it may well be that the bottleneck for big data projects using Teradata Unified Data Architecture™ will be the limits of your analysts' imaginations, not the technology.

A quick set of questions to get started:

1. What is your current pain point or opportunity that will give you a competitive edge using big data?
2. Who understands and has the power to drive a project?
3. What is the likely benefit in quantifiable terms for you? ROI? Do you need help with a Business Value Discovery?
4. Does Teradata Unified Data Architecture™ fit the problem?
5. Does it make sense to proceed to a POC on a sample data set, with Teradata's help, so you can get your feet wet, get some early experience with the tools, and see what the technology can do for you?
6. If you already have Teradata, can you plan an Architecture Assessment so your IT group is linked in and you achieve buy-in for operationalization?
7. Is the timing right? Is your culture ready?

Three things to watch for, and avoid, based on our experience:

First, the "shiny new toy" phenomenon. While it's laudable for people to want to learn new technologies, it's important to make sure they are doing so in a business context. There are always shiny new things competing for your attention. Eighty percent of them (and 90 percent of the associated vendors) flame out, so make sure that you place your bets carefully. For an entertaining perspective on this by Dan Graham, one of Teradata's most insightful curmudgeons, see 'Throwing Your Hat in the Ring' or 'Big Elephant Eats Data Warehouse' blog posts.[9]

Second, the "science project in the basement" phenomenon. In cases where big data projects have run into problems, they usually occur near the end, when they struggle to take the results into productization without the appropriate planning and consideration of many of the operationalization aspects of real projects. It's often important to insulate the risk takers in your company who want to be change agents so they can work quickly in a high-energy environment. However, it's equally important to make sure that you have a good game-plan for folding their insights back into the other insights you've created with Teradata, and that you can take action. We call this the "Insight Integration" step and the "Taking Action" step. Usually, the "Taking Action" aspects of data and insights depend on Teradata and its connections to front-line systems like websites, dashboards, and mobile apps. This is important, because it's very unlikely that Hadoop (given its batch orientation) or even

Teradata Aster is going to be able to handle the volume of requests from operational (as opposed to strategic) BI workloads, an area that Teradata calls "Active Data Warehousing™." (You can find much more information on **Teradata.com** about Teradata's prowess in mixed workloads as well as case studies of customers connecting hundreds of thousands, even millions of customers on web or mobile devices directly to the same system that handles their traditional BI workloads.)

Finally, monitor for the "culture clash." There's often a culture clash between the people who did the "new project" and those who have to run the business—and that's the area where you need to spend some extra energy. It really is about people, process, and technology. While the big data technology is proven and the architecture is robust, people and process issues are often the stumbling blocks to putting Teradata Unified Data Architecture™ into action.

SUMMARY

Big data is an evolution, not a revolution—of technology, analytical capabilities, and skills for people responsible for using data to create insights and discoveries that lead to a competitive edge. Our expertise with Teradata Unified Data Architecture™ helps companies navigate the options, understand the economics, and learn from best practices to put big data programs in place. For more information visit us at **Teradata.com**.

Suggested Readings

1. Lisa Arthur, *Big Data Marketing: Engage Your Customers More Effectively and Drive Value*, Wiley, 2013.

2. Lisa Kart, Nick Heudecker, Frank Buytendijk. "Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype," Gartner Report GG0255160, September 12, 2013.

3. Thomas H. Davenport and Jeanne G. Harris, *Competing on Analytics, The New Science of Winning*, Harvard Business School Publishing, 2007.

4. Erik Brynjolfsson, Lorin M. Hitt, Heekyung Hellen Kim—MIT Sloan, "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?", Social Science Research Network, April 2011. **http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1819486**

5. Intel, "Using a Multiple Data Warehouse Strategy to Improve BI Analytics", IT@Intel White Paper, March 2013. **http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/using-a-multiple-data-warehouse-strategy-to-improve-bi-analytics.pdf**

6. Richard Winter. "Big Data—What Does It Really Cost," June 2013. **http://www.asterdata.com/big-data-cost**

7. To drill into more information about the Teradata Aster Discovery Platform, read:
   ~ **http://www.asterdata.com/resources/whitepapers.php**
   ~ **http://www.asterdata.com/resources/analyst.php**

8. Teradata Website. "Teradata and Fuzzy Logix Deliver Game-Changing In-Database Analytics," Teradata Brochure, July 2013, Available on **Teradata.com**.

9. Dan Graham Blog Post, **http://blogs.teradata.com/data-points/throwing-your-hat-in-the-ring**

10. Teradata Website. "Teradata Unified Data Architecture™", TDWI E-Book, May 2013. Provides an 11-page compilation of interviews of noted analysts Mark Madsen and David Loshin, as well as interviews with Teradata's Imad Birouty and Steve Wooledge. Available on **Teradata.com**.

11. Teradata Website. "Teradata Unified Data Architecture™: A Visionary Framework", EB6705, April 2013. Provides a six-page quick overview of the Teradata Unified Data Architecture™ and three customer use cases—two in telecomm and one in banking. Available on **Teradata.com**.

12. Teradata Website. "Turning Data into Profit with Teradata Unified Data Architecture.™" is a two-page summary from Q1-2013. **http://www.teradata.com/white-papers/Turning-data-into-profit-with-Teradata-Unified-Data-Architecture**

10000 Innovation Drive Dayton, OH 45342   teradata.com