Teradata Solution Technical Overview

Teradata Pioneered Data Warehousing

Since our first shipment of the Teradata® Database, we've gained more than 30 years of experience building and supporting data warehouses worldwide. Today, Teradata solutions outperform other vendors' data warehouse solutions, from small to very large production warehouses with petabytes of data. With a family of purpose-built platforms, all running the Teradata Database, that lead spans needs from specialized analytical applications to the most general enterprise data management.

Teradata Corporation offers a powerful, complete solution that combines parallel database technology and scalable hardware with the world's most experienced data warehousing consultants, along with the best tools and applications available in the industry today.

> **"In order to meet our goal of 24-hour delivery of custom-built product to our customers, we must have a worldwide presence. Teradata is critical to that effort. Partnering with a company like Teradata is enabling us to be more competitive and to serve our customers well."**
>
> – Doug Hawken, President and COO, PING, Inc.

Making Smarter, Faster Decisions

Most companies have large volumes of detailed operational data, but key business analysts and decision makers still can't get the answers they need to react quickly enough to changing conditions. Why? Because the data are spread across many departments in the organization or are locked in a sluggish technology environment. Today, Teradata solutions are helping companies like yours consolidate those data to present a single view of your business. You can make smarter decisions faster and less expensively, and get answers to questions that previously went unanswered.

Real, Measurable Data Warehouse ROI

In today's economy, companies are striving to maximize the return on their information technology investment. Companies hesitate to implement centralized data warehouses because they're concerned that a single system can't meet the requirements of their multiple constituents, and they worry about the cost of acquiring and maintaining a decision support system outside the scope of their current departmental environments. Now, however, a single database can be used for high-volume strategic and operational queries. The design of Teradata Database is geared toward multiple groups of users and multiple reporting and analytical requirements enabling businesses to meet the needs of their current users and deploy new applications faster than ever, further accelerating their ROI. And, with the Teradata Purpose-Built Platform Family, you can start with a departmental data warehouse and grow into a fully scalable system with complete configuration flexibility or build an enterprise data warehouse (EDW) yet still meet the need of a specialized application for its own system.

Data Warehousing Leader

Teradata Database powers the data warehouses of the world's leading companies in all industries. When our customers entrust terabytes or petabytes of their precious corporate data to a database engine, they expect availability 24 hours per day, 7 days per week, and 52 weeks per year. Teradata Database meets those expectations.

Our services consultants are the most experienced in the industry, and our Teradata platforms have been hailed as

TERADATA®

the most reliable and scalable computing systems for data warehousing on the market today. Teradata also has the most comprehensive data warehousing programs to assist with your current and future initiatives. And, Teradata recognizes that strategic partnerships provide robust, proven, and reliable offerings that extend and strengthen our data warehouse solutions with best-of-breed capabilities. These partnerships enable you to choose from a full suite of tools, applications, and software products – a win-win situation for everyone.

The Database – A Critical Component of Your Data Warehouse

Will the database you're using today support your data warehouse needs in the future? Probably not if you're using the database you also selected for on-line transaction processing (OLTP). Most common relational database management systems (RDBMS) were designed for OLTP environments, which need quick access and updates to single-record requests. These databases, designed for precise and directed access, often perform poorly when faced with entirely different workloads, such as full-table scans, many-table joins, sorting, or aggregating – common data warehousing functions. Their response to unfamiliar needs is to add more tables to the database or add more tuning parameters to the database management system. This adds additional overhead and complexity, to the point where one popular OLTP-based RDBMS has more than 100 different parameters just to control data caches in memory.

What happens when users and applications need to scan large amounts of data to answer complex business questions? Can users afford delays in getting the answers they need because custom aggregate tables must be created to meet their analytical needs? Can you afford to support a department of DBAs to maintain and tune your data warehousing databases? Will the OLTP database support the scalability requirements so necessary in most data warehouse environments?

What happens when you've developed an application on a departmental warehouse, such as the Teradata Data Mart Appliance, or separate specialized system, but it becomes more cost effective later to move it into your EDW?

The Right Blend of Experience and Expertise

There's much more to Teradata than hardware, software, and technology. Teradata services consultants bring you a unique combination of experience, expertise, and proven processes. No one knows more about implementing data warehouses – and no one understands your business needs better than Teradata consultants. Our consultants can deliver industry-specific expertise that focuses on your organization and on making your data warehouse a powerful decision-making tool that will help your business grow.

Years of experience in data warehousing and specific industries allow Teradata consultants to guide you through every aspect of your data warehouse strategy, everything from design and implementation to data modeling, ETL, system architecture, applications, and end-user training. And with more than 30 years implementing the world's most successful data warehouses, Teradata consultants can also deliver best practices that we've incorporated into the Teradata Solutions Methodology. It's a proven, patented approach to data warehousing that blends customized tools and quantifiable metrics to let you quickly capture business value by reducing risks and delays.

"Teradata Professional Services brought their data warehouse architecture leadership experience, strong data models, and strong designs. They worked very closely with us to review those models and designs, hardware, and infrastructure choices."

– Terry Johnson, Technology Manager for Internet Services, Wells Fargo

TERADATA

Teradata Solution Technical Overview

> **"I would love to say that I spend half my time administering the Teradata system. Because, when you consider the EDW's importance to the company, it should be half my time. But, the truth is, that I spend a lot more time maintaining other databases and performing other aspects of my job than I do working on the Teradata system. Teradata runs so well that it really runs on its own. I don't spend a lot of time managing it compared to the other databases that I am responsible for. It's by far, the most important, and yet it takes the least amount of effort to keep running."**
>
> – Sandy Rumble, Data Center and Operations Manager, PING, Inc.

Do you have to redesign the data model and rewrite the application to match the EDW? Not if you've used the same database for all your analytical needs regardless of hardware configuration or query mix characteristics.

These are important questions to consider when choosing your data warehouse partners and vendors. Data warehousing is a dynamic and iterative process, the requirements of which are constantly changing as the demands on your business change. Because a data warehouse is something you build – not something you buy – it's critical to choose data warehouse partners and vendors who will allow you to adapt quickly and flexibly to these ever-changing business requirements.

The Teradata Solution	
Teradata Database	> The world's most powerful data warehousing engine
Teradata Tools and Utilities	> The building blocks to develop, optimize, manage, and integrate your Teradata system
Teradata Consultants	> The most experienced data warehouse consultants in the industry
Teradata Nodes and Enterprise Storage	> The most reliable, scalable hardware platform
Teradata Analytic Applications	> Decision support to differentiate your business in a competitive world
Software and Services Partnerships	> Alliances with other industry leaders

Teradata Database – The Premier Performer

The primary requirement of your database is performance. So why not begin with the database that outperforms all others – Teradata Database? This scalable, high-performance decision-support engine is unmatched in the industry.

Designed from the ground up for decision support, the Teradata Database is free from the limitations that traditionally plague transaction processing databases. Instead, the Teradata Database easily and efficiently handles complex data requirements and simplifies management of the data warehouse environment. How? With a truly parallel implementation that automatically distributes data and balances workloads. The Teradata Database's ease of use, combined with true query and workload parallelism, allows it to achieve performance and throughput levels not found in any other RDBMS.

What is Parallel Processing?

Parallel processing is an efficient method of handling complex tasks. The concept is to break a task into smaller subtasks that can be managed concurrently by multiple work units. Here's an example: Imagine that you were handed a shuffled stack of playing cards and were not allowed to scan the cards beforehand. Then you were asked a simple question, "How many aces are in the stack?" The only way to get the answer would be to scan the entire deck of cards.

TERADATA

Now imagine if the same cards were distributed among four people, each receiving one-fourth of the cards. The time required to answer this same query is now reduced by four times. Each person would simply have to scan their cards, and the four totals would be aggregated for the correct answer. In this simple example, we can refer to these four people as parallelized units of work. As you can see, more available parallelized units of work will result in faster query processing. The larger the data volume and the more complex the queries, the bigger the payoff from using parallel processing. It's also important to note that the most efficient way to distribute the playing cards (or data) is to distribute them evenly among the four people (or parallelized units of work).

> **"We were in need of a highly scalable system – not only in the number of concurrent users or queries, but in the amount of data as well. Teradata gave us the possibility to overcome the limits we had before. The combination of Teradata industry experience with internal business expertise creates solutions that empower us to respond to challenges of the very competitive market."**
>
> – Jaroslaw Bartosik, Leader of the Business Intelligence Competency Center, Polkomtel S.A.

Parallelism allows ad-hoc queries to be performed as efficiently as possible, as opposed to a single person, or large query, becoming a bottleneck because they were given more cards or a particular set of cards where the work must be performed sequentially. (See Figure 1.)

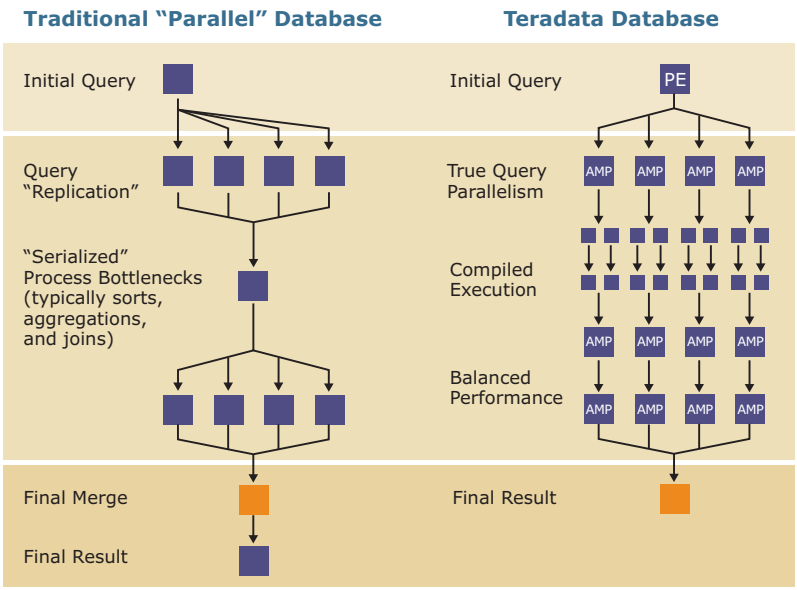


Figure 1. Traditional Parallel Database versus Teradata Database.

Key Definitions

Teradata Database was designed for parallelism. Its patented architecture allows complex decision support workloads to be broken down into small tasks and distributed to multiple parallel software processors, known as virtual processors, within the database. We refer to this virtual processor as an Access Module Processor (AMP).

Each AMP owns a portion of the database. Multiple AMPs reside on a single processor node. Therefore, Teradata Database doesn't rely on the hardware platform for parallelism, scalability, reliability, or availability. These capabilities are inherent in the database architecture and are independent of the hardware configuration.

The AMP is one of two types of virtual processors (VPROC). The second type of VPROC is the Parsing Engine (PE), which breaks up a request or query into manageable pieces and distributes the work to the AMPs for processing. Multiple PEs can also exist on a single node.

TERADATA

Teradata Database's ability to run multiple AMPs and PEs on a single node is enabled by the Parallel Database Extensions (PDE). PDEs provide the infrastructure that enables Teradata Database's parallelism to work within the Linux operating system. The PDE layer allows Teradata Database to perform independently of the operating system. Teradata Database is Teradata Database, regardless of the operating system or hardware platform.

Further ensuring independence of the AMP's database processing from the hardware it runs on is the optional Teradata Virtual Storage. Teradata Virtual Storage enables a variety of storage devices to be used to store data owned by each AMP, and ensures that the data are placed on the most appropriate physical storage. Frequently accessed, or hot, data are automatically stored in fast, quickly accessed locations, such as the outer cylinders of a disk drive or physically separate solid state drives (SSD). Infrequently accessed, or cold, data are stored in locations or storage mediums that take longer to access. And, if some of the data's frequency of use, or temperature, changes, Teradata Virtual Storage can automatically migrate those data to the most appropriate location – all without human intervention or the AMP software's involvement. And, this is accomplished with no changes to the database or the application. As a result, the mix of storage devices connected to a system can be changed after the system is installed, and a system can be configured to economically include data of various values – all adding up to the highest value data warehouse.

> **"I'm continually amazed by what I see in the performance with Teradata. When I look at a load – and in a matter of seconds we're putting thousands – hundreds of thousands of rows out into a table – there's nothing can compete with that. It gives us a great advantage in the warehouse environment as we get more and more data to be able to actually process all that stuff."**
>
> – Bruce Theriault, Systems Analyst, Unum Group

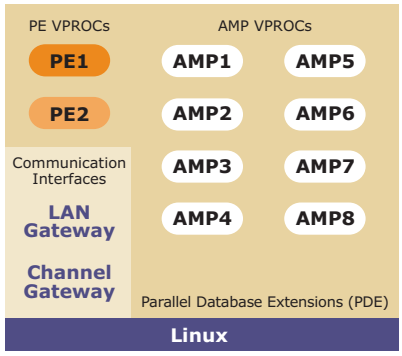


Figure 2. Teradata Node SW Architecture.

Intra-node Parallelism

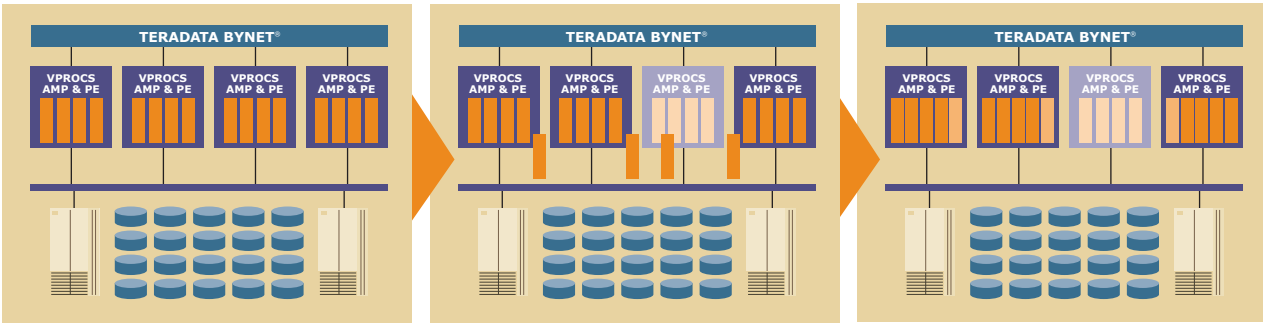
Figure 2 depicts an example of intra-node parallelism. The configuration illustrated is a 2-CPU node with ten VPROCs. There are eight AMP virtual processors in this configuration, as well as two PEs. Each PE has access to each AMP, which allows for complete parallel processing of each request. The data are distributed evenly among the AMPs.

Physical Data Layout

Another way to improve performance is to store each relational table with a row- or column-oriented physical layout. No matter which orientation is used, the rows of each table are spread evenly among the AMPs. But, on each AMP, either the values within each row are stored together, or the values within each column are stored together. Columnar orientation provides extreme performance when a table has many columns, and only a small number of those columns are generally required for a query. In this case, the database can read only data from the required columns off of the disk, reducing the I/O versus reading all of the columns in a row-oriented storage layout. Row orientation is more efficient when a high percentage of the columns are generally required for queries because there is little excess I/O required to read unnecessary data, and there is no overhead to *reassemble* the rows to be returned to the user. Teradata can break the table up into any combination of row- and column-oriented pieces for physical

TERADATA

Teradata Database Enhanced Availability



Teradata software provides high availability beyond other RDBMSs by…

…compensating for hardware failures:

> Automatic failover for dynamic workload rebalancing (migrating VPROCs)

> Online, continuous backup (fallback)

…by being up and available before the operating system completes a reboot (multi-node system)

Teradata Database's unique architectural design was developed to support mission-critical, fault-tolerant decision-support applications.

storage within its parallel architecture. This results in the least possible I/O and fastest performance.

Massively Parallel Processing

Teradata Database takes the architecture even further, providing maximum scalability by adding interconnect capabilities. Running on Linux, each processor node can be connected via our high-speed interconnect, Teradata BYNET®, to form a loosely-coupled, massively parallel processing (MPP) architecture that is managed as a single system. This provides the foundation for Teradata systems' linear scalability, which can start with a single, two-processor node environment and scale to thousands of physical processors and tens of thousands of VPROCs. This makes Teradata Database a perfect fit for entry-level or massive EDWs.

Teradata BYNET – Scalable Interconnect

BYNET is a redundant, fault-tolerant, intelligent, high-speed circuit switching interconnect. BYNET allows the database to coordinate and synchronize the activities of many SMP nodes without increasing network traffic or degrading performance as the system grows. BYNET provides node-to-node data transfer and can linearly scale, supporting up to 4,096 nodes in a single system.

Scalable, Purpose-Built Platform Family

Teradata Database's scalability is enhanced through its tight integration with the Teradata platform family. The Teradata Purpose-Built Platform family, which ranges from the Teradata Data Mart Appliance to the Teradata Active Enterprise Data Warehouse, and Teradata Enterprise Storage, provides seamless, transparent scalability that meets any analytical requirement. Adding more computational power is as simple as adding more platform elements to the current system, which also enables you to leverage technology advances over time while protecting your original investment. The operating system will automatically recognize and adapt to the additional system resources, and Teradata Database will redistribute existing data to take advantage of the hardware. Existing applications will continue to run without

TERADATA®

Teradata Solution Technical Overview

How Industries Use Teradata

Financial
The financial industry uses Teradata solutions for relationship banking and householding where all customer account information is merged for cross-segment marketing. Data are sourced from diverse geographical areas; different lines of business, including checking and savings accounts, auto and home loans, credit cards, and ATMs; and from various on-line systems.

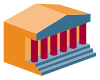
Retail
Retailers use Teradata solutions to compile and analyze months and years of data gathered from checkout scanners in thousands of retail stores worldwide. This information is used to manage purchasing, pricing, stocking, and inventory management, and to make store configuration decisions.

Manufacturing
Manufacturers use Teradata solutions to determine the most efficient means for supplying their retail customers with goods. They can determine how much product will sell at a price point and manufacture goods for just-in-time delivery.

Telecommunications
The telecommunications industry uses Teradata solutions to store data about millions of customers, circuits, monthly bills, volumes, services used, equipment sold, network configurations, and more. Detailed information about revenues, profits, and costs is used for target marketing, revenue accounting, government reporting compliance, inventory, purchasing, and network management.

Travel
The travel industry uses Teradata solutions to gain agility in the marketplace to quickly respond to customer demands or industry changes. Detail-level data are integrated to provide clarity, insight, and decision confidence to drive growth, reduce costs, and optimize revenues. Data from diverse operational systems can be analyzed, queried, and mined to address strategic and tactical decision making in the areas of customer management, financial management, operations management, revenue management, cargo management, and workforce management.

Healthcare
The healthcare industry uses Teradata solutions to pinpoint the sources of medical cost trends, to provide members with more tailored and attentive service, to underwrite and manage risk more precisely, and to encourage evidence-based care by understanding provider behavior and the impact of contract incentives better.



Government
Around the globe, security and intelligence, defense, national, state, local, and civilian agencies are using Teradata solutions to help ferret out waste, fraud, and abuse in government programs; reallocate resources as needs change; overcome budget hurdles; improve customer service, employee productivity, and operational efficiency; track critical parts globally; collect delinquent taxes; and consolidate data from disparate systems to facilitate research.

Transportation, Distribution, and Logistics
Transportation logistics and distribution companies use Teradata solutions to optimize asset utilization, reduce operational expenses, increase revenues, improve margins, and enhance customer service. Through advanced shipment level profitability, leveraging the power of Teradata Database, fact-based, real-time pricing decisions have been automated with more precision and confidence. Visibility across complex transportation networks and supply chains, when combined with applications providing powerful predictive analytics and event-based alerting, allows transportation and logistics companies to focus on the most important aspects of their complex businesses.

Insurance
Insurers use Teradata solutions to provide unprecedented decision support for addressing the short- and long-term health of their businesses. The warehouse uses critical information – every event and concept from policy numbers and product information to claims coding and risk classification – to create a common understanding across all business functions. This eases communication between business people and IT staff and supports everything from regulatory and actuarial reporting to policy administration and underwriting.

Media and Entertainment
Media and Entertainment companies across the content and advertising supply chain use Teradata solutions to understand and analyze multiple dimensions of content and advertising performance, as well as customer engagement with content and advertising. Teradata solutions support robust analytics across Media and Entertainment organizations from digital and physical supply chain, to title and IP performance, audience segmentation, content and advertising targeting, and deep consumer behavioral insight.

TERADATA

modification. No matter which Teradata platform family member fits your needs best, it runs the same Teradata Database.

Application Programming Interfaces

Teradata provides several standardized interfaces to facilitate easy client/server application development. These include the ODBC Driver for Teradata, the Teradata JDBC Driver, the OLE DB Provider for Teradata, the Teradata Call-Level Interface (CLI), the .NET Data Provider for Teradata, and the TS/API, which permits applications that normally access IBM DB2 to access the Teradata Database.

Language Preprocessors

Teradata provides several preprocessors to facilitate application development in languages such as COBOL, C, and PL/1. With the libraries in these preprocessors, developers can create or enhance client- or host-based applications that access Teradata Database.

Data Utilities

Teradata Database includes client-resident and host-based utilities that allow users and administrators to interact with and control the engine. The Teradata Utility Pack provides a core set of tools and utilities for developing, administering, querying, reporting from, and connecting to Teradata Database. Teradata Meta Data Services allows companies to comply with federal and industry regulations by storing business and technical information about the data in their data warehouse from Teradata. In addition to database management tools, Teradata provides a suite of parallel load and extract utilities that addresses diverse data load needs from batch to continuous loading. Teradata Replication Services accommodates your real-time data synchronization needs in a heterogeneous environment.

Database Administration Effort	Other RDBMS	Teradata
Logical Data Modeling	High	High
Physical Data Modeling	High	Low
Data Partitioning Definition	High	Low
Data Placement Definition	High	None
Free Space Management	High	None
Data Balancing Control	High	Low
Data Reorganizing	High	None
Index Reorganizing	High	None
Query Tuning	High	None
Workload Management	Moderate	None
Workspace Management	Moderate	None
Change Management	High	Low

Database Administration Tools

The Teradata Database has a rich collection of tools and facilities to control database operation, administration, and maintenance. These include Teradata Backup, Archive, and Restore utilities that combine the strength of industry-leading partners and Teradata's expertise; Teradata Active System Management for workload management; and Teradata Workload Analyzer for query performance analysis and recommendations about workload groupings and resource allocations. All of these tools, and many others, can be accessed individually or through Teradata Management portlets. Teradata Viewpoint, a web-based systems management portal, provides critical self-service functionality, including query status and system health information, to potentially thousands of end users.

Teradata provides a bundle of powerful tools that enables you to develop and optimize complex queries. The Teradata Analyst Pack includes Teradata Visual Explain, Teradata System Emulation Tool, Teradata Index Wizard, and Teradata Statistics Wizard. These tools simplify the database administrator and query planner's jobs allowing them to analyze complex query plans more efficiently.

Access from Anywhere

As application environments and the latest interface standards change, Teradata

continuously expands and updates Teradata Database access methods. Whether data come in via a message bus architecture or an application customizing offers on a Web page based on data requested from the data warehouse, Teradata provides the latest standard interfaces. Some examples are:

Java

The Teradata JDBC Driver is a Type 4 JDBC driver developed to provide application developers with a simple, easy-to-use API to access Teradata Database from the Internet or an intranet. Any client capable of running a Java applet or application, including web browsers, such as Netscape Navigator or Microsoft Internet Explorer, can access Teradata Database directly.

CGI Access

The CGI Access validates parameters received through the HTTP query string and allows all data manipulation language (DML) constructs, including SELECTS, INSERTS, UPDATES, and DELETES.

.NET Data Provider for Teradata

The Microsoft .NET Data Provider for Teradata provides a complete set of ADO.NET-compliant components that allows a developer to access, manipulate, and administer data in their Teradata Database using any .NET language, such as Visual Basic® or C# (C Sharp). It supports all currently-supported features of Teradata Database, including triggers and stored procedures, user-defined functions, queue tables, BLOBs, and CLOBs.

Need More Reasons to Choose Teradata?

Why select Teradata to be your data warehouse supplier? Teradata is the world's largest company solely focused on data warehousing and integrated marketing management through database software, enterprise data warehousing, data warehouse appliances, and analytics. Teradata provides the best database for analytics with the architectural flexibility to address any technology and business need for companies of all sizes. Supported by active technology for unmatched performance and scalability, Teradata's experienced professionals and analytic solutions empower leaders and innovators to create visibility, cutting through the complexities of business to make smarter, faster decisions. Simply put, Teradata solutions give companies the agility to outperform and outmaneuver for the competitive edge. For more information about how Teradata solutions can help your organization, contact your Teradata representative or visit Teradata.com.

Teradata Solution Top Ten Reasons	
1	Teradata Database
2	Effortless scalability
3	Ease of management
4	Reduced risk
5	Industry-leading data solution expertise
6	Seamless mainframe integration
7	Mission-critical availability
8	Investment protection
9	Quickest time to solution
10	Proof beyond claims

"We've become more sophisticated, implementing components of the data warehouse as the user community has grown. Teradata is able to handle mixed workloads and assist in meeting the performance SLAs of our business community."

– Sullivan B. McConnell, Vice President of Business Intelligence, The Travelers Companies, Inc.

TERADATA®

Teradata Solution Technical Overview

Teradata Solution Technical Specifications

Database Capacity	> 3TB to 186 petabytes (PB) > Up to 4,096 nodes
Data Model	> Relational > ANSI SQL-92 compatible > Completely parallel
Architectures Supported	> Symmetric Multi-Processing (SMP) > Massively Parallel Processing > Cloud
Operating Systems Supported	> Linux
Client Platforms Supported	> Microsoft® Windows® > IBM z/OS MVS > Most UNIX® platforms > Linux
Mainframe Integration	> IBM (or compatible) mainframe
Language Preprocessors	> PL/1, C, COBOL
Tools and Utilities **Teradata Utility Pack**	> BTEQ (query/report writing) > Teradata SQL Assistant (query/DBA) > Teradata SQL Assistant Java Edition (query/DBA) > ODBC Driver for Teradata > Teradata JDBC Driver > .NET Data Provider for Teradata > OLE DB Provider for Teradata > Teradata Call-Level Interface (multiple platforms) > Teradata Meta Data Services > Teradata Plug-in for Eclipse
Teradata Database Management Utilities	> Teradata Management portlets (database system administration) > Teradata Viewpoint > Teradata Active System Management portlets (query workload management) > Teradata Workload Analyzer > Teradata Analyst Pack, which includes Teradata Visual Explain, Teradata System Emulation Tool, Teradata Index Wizard, and Teradata Statistics Wizard (query tuning)
Teradata Parallel Load Utilities	> Teradata TPump (continuous load) > Teradata FastLoad (data loading) > Teradata MultiLoad (multiple table loading) > Teradata FastExport (data extraction) > Teradata Replication Services (real-time data synchronization) > Teradata Unity (multi-system ecosystem management and query routing) > Teradata Parallel Transporter (single architecture for all load protocols)
Other Tools and Utilities	> ARCMAIN (mainframe archival and recovery) > Teradata Aggregate Designer > Teradata OLAP Connector for Microsoft Excel > Teradata Warehouse Miner > Teradata Profiler > Teradata Analytic Data Set Generator > R add-on for Teradata

BYNET, Teradata, and the Teradata logo are registered trademarks of Teradata Corporation and/or its affiliates in the U.S. and worldwide. Microsoft, Windows, Excel, and Visual Basic are registered trademarks of Microsoft Corporation. UNIX is a registered trademark of The Open Group. Intel is a registered trademark of Intel Corporation. Teradata continually improves products as new technologies and components become available. Teradata, therefore, reserves the right to change specifications without prior notice. All features, functions, and operations described herein may not be marketed in all parts of the world. Consult your Teradata representative or Teradata.com for more information.

Copyright © 2006-2011 by Teradata Corporation All Rights Reserved. Printed in U.S.A.

TERADATA®