

# Base de datos Teradata: Introducción Técnica



Nunca antes había vista su negocio así.

# Teradata, pioneros en datawarehousing

Desde nuestra primera instalación de la base de datos relacional Teradata, acumulamos más de 20 años de experiencia en la construcción y soporte de soluciones Data Warehouse en todo el mundo.

Actualmente, nuestra oferta de data warehouse supera a la de cualquier otro proveedor, tanto en entornos pequeños y medios como en los grandes data warehouse corporativos que procesan cientos de terabytes de información.

Teradata Warehouse es una solución completa y potente que integra la tecnología de base de datos paralela Teradata, plataformas hardware que crecen con las necesidades de la empresa, los consultores de datawarehousing más experimentados del mundo y las mejores herramientas y aplicaciones del mercado.

## **Decisiones más acertadas en menos tiempo**

Aunque la mayoría de las empresas disponen de una cantidad ingente y pormenorizada de datos sobre su funcionamiento, a los analistas de negocio y ejecutivos, encargados en definitiva de tomar las decisiones, no les llegan las respuestas que necesitan para reaccionar con la rapidez suficiente ante las variaciones del entorno. ¿Por qué? Porque esa información está diseminada en muchos departamentos o aprisionada en un sistema lento. Hoy en día, gracias a Teradata Warehouse, las empresas están consolidando esa información para tomar decisiones más acertadas en menos tiempo y con un coste menor, obteniendo respuestas a preguntas que antes se quedaban sin contestar.

## **La base de datos: el componente fundamental de su solución data warehouse.**

¿Permitirá la base de datos que utiliza actualmente atender las necesidades y requerimientos futuros de su



entorno de soporte a la toma de decisiones? Probablemente no si utiliza la base de datos que escogió para entornos OLTP. Los sistemas más habituales de gestión de bases de datos relacionales (RDBMS) están diseñados para entornos OLTP, en los que se producen múltiples accesos con pequeños movimientos de información. Estas bases de datos, concebidas para proporcionar un acceso preciso y directo, a menudo dejan que desear al procesar tareas completamente diferentes entre sí, como análisis de tablas completas, unión de varias tablas, clasificación o agregación de datos, todas ellas cotidianas en un entorno data warehouse. Frente a nuevas necesidades lo que hacen es realizar más ajustes a posteriori. Esto complica inmensamente su funcionamiento, hasta tal punto que un conocido sistema RDBMS basado en OLTP emplea más de 100 parámetros distintos sólo para administrar la caché de datos en memoria.

¿Qué ocurre cuando usuarios y aplicaciones tienen que explorar grandes cantidades de datos con objeto de dar respuesta a cuestiones de negocio complejas? ¿Pueden los usuarios permitirse la pérdida de tiempo que supone la creciente necesidad de sintonización requerida para que su base de datos se acomode a los cambios y crecimiento de su entorno data warehouse? ¿Soportará su base de datos los requerimientos de escalabilidad y rendimiento impuestos por el entorno de data warehouse?

Estos son aspectos importantes a tener en cuenta a la hora de seleccionar un proveedor y los partners de soluciones de soporte a la toma de decisiones. Un data warehouse es un proceso dinámico e iterativo, cuyas exigencias cambian continuamente, al ritmo de las de su negocio. Un data warehouse es algo que Ud. construye, no algo que Ud. compra.-Por tanto es vital seleccionar el proveedor y los partners adecuados que le permitan adaptarse rápida y libremente a las nuevas exigencias de su empresa.

**“Nuestro negocio depende cada vez más intensamente de los datos y de la gestión de las relaciones. Para alcanzar nuestro objetivo de crear un servicio de gestión de viajes online personalizado, necesitábamos una base sólida. Teradata nos la ha proporcionado.**

– Mike Stacy, vicepresidente senior de marketing de clientes, Travelocity

## Teradata Warehousing

- Base de datos Teradata** > El motor de data warehouse más potente del mundo
- Servicios profesionales de Teradata** > Los consultores de datawarehousing más experimentados del sector
- Plataformas NCR** > La plataforma hardware más fiable y escalable
- Partners de software y servicios** > Alianzas con empresas líderes en su sector



# Teradata, Líderes en datawarehousing

Nuestros consultores especializados en data warehouse son los más experimentados del sector. Nuestra plataforma hardware NCR ha sido reconocida como la más robusta y escalable del mercado para la implantación de soluciones data warehouse. Además Teradata dispone de metodología, aplicaciones y herramientas capaces de soportar sus proyectos actuales y futuros, y de alianzas con los fabricantes de software y proveedores de servicios mas reconocidos del mercado para ampliar la capacidad de nuestra solución con las mejores opciones.

## El motor Teradata: máximo rendimiento

El requerimiento fundamental a una base de datos es el rendimiento. Entonces, ¿por qué no empezar con la base de datos que supera en rendimiento a todas las demás? La base de datos Teradata.

Este gestor escalable y de alto rendimiento para entornos de toma de decisiones no tiene comparación en el mercado. Es una solución que ofrece un paquete completo de herramientas de gestión y acceso a información, servicios de primera calidad basados en la experiencia

acumulada por Teradata durante años de experiencia pionera en datawarehousing. Es una combinación única de experiencia y conocimientos que nos permite ofrecerle una solución de datawarehousing de probada eficacia, una solución que supone ventajas para usted y para su empresa.

La base de datos Teradata, diseñada desde sus orígenes como soporte a la toma de decisiones con una arquitectura paralela, no está sujeta a las limitaciones que tradicionalmente aquejan al resto de las bases de datos comerciales.

Teradata gestiona de forma fácil y eficaz requerimientos complejos, simplificando al mismo tiempo la administración y gestión del data warehouse. ¿Cómo?, distribuyendo los datos y balanceando la carga de trabajo de forma automática. La facilidad de uso de Teradata, junto con la potencia de su procesamiento realmente paralelo, permite conseguir niveles de rendimiento no alcanzables por el resto de gestores de bases de datos relacionales.

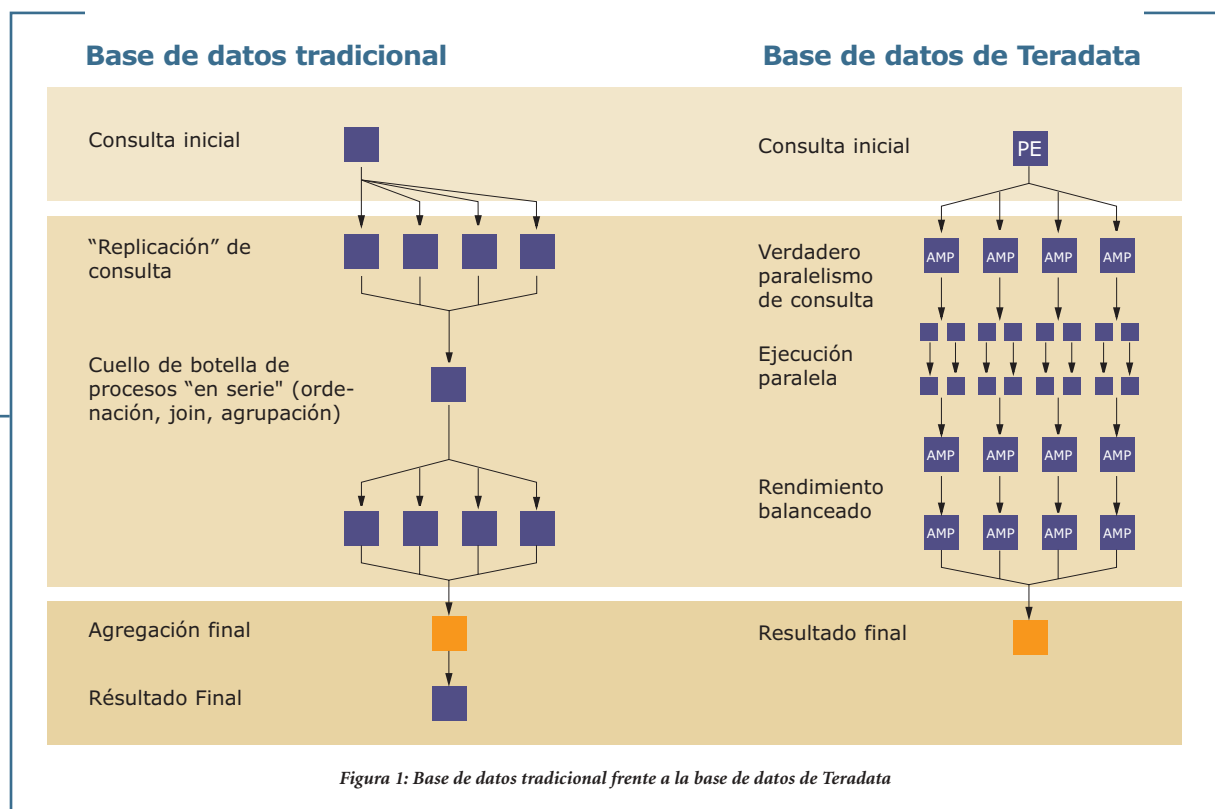


Figura 1: Base de datos tradicional frente a la base de datos de Teradata

### ¿Qué es el procesamiento paralelo?

El procesamiento paralelo es el método más eficiente para gestionar tareas complejas. Consiste en dividir una tarea en subtareas más sencillas que puedan ser gestionadas concurrentemente por múltiples unidades de trabajo. Por ejemplo, imagine que le dan un mazo barajado de cartas y que no le permiten echarles un vistazo. A continuación se le hace una pregunta simple: ¿Cuántos ases hay en la baraja? La única manera de responder sería ir mirando una por una todas las cartas de la baraja. Ahora imagine que las mismas cartas se distribuyen entre cuatro personas, cada una de las cuales recibe una cuarta parte de las cartas. El tiempo necesario para responder a la misma pregunta se reduciría a una cuarta parte. Cada persona sólo tendría que mirar sus cartas, y los cuatro subtotales sumados darían la respuesta correcta a la pregunta realizada.

En este sencillo ejemplo, puede decirse que las cuatro personas serían unidades de procesamiento paralelo. Cuantas más unidades de paralelismo, más rápida será la respuesta. Y cuantas más cartas (datos) haya y más compleja sea la pregunta, más beneficios nos reportará el procesamiento paralelo. Es importante asimismo señalar que la forma más eficiente de distribuir las cartas (datos) es dividir las equitativamente entre las cuatro personas (unidades de paralelismo).

Gracias al procesamiento paralelo, las solicitudes independientes se procesan con la máxima eficiencia posible, evitando que una sola persona o una solicitud grande sea un cuello de botella por tener más cartas o tener un grupo específico. (Ver Figura 1)

**“Lo único seguro en data warehouse es que el volumen de datos crece exponencialmente a medida que los usuarios explotan sus posibilidades. Estos requieren información más detallada, relativa a periodos de tiempo más largos. Por lo tanto una base de datos capaz de escalar hasta múltiples terabytes es un requerimiento crítico.”**

– Patricia Seybold Group



# El motor Teradata

## Definiciones clave

Teradata es un gestor de base de datos relacional específicamente diseñado para soportar paralelismo. Su arquitectura patentada permite descomponer las preguntas complejas entre múltiples unidades de trabajo paralelas en el software de la base de datos, cada una denominada AMP (Access Module Processors).

A cada AMP le corresponde una parte del espacio y de los datos en la base de datos. En cada nodo residen varios AMPs. Por consiguiente Teradata no está condicionada por la plataforma hardware para soportar paralelismo, escalabilidad o alta disponibilidad. Estas características son inherentes a su arquitectura software e independientes del sistema operativo o de la configuración hardware.

El AMP es un tipo de procesador virtual (VPROC). Existe otro VPROC, el PE (Parsing Engine), que descompone las consultas (sentencias SQL) en sentencias más sencillas, distribuyendo a continuación la tarea entre los AMPs. Varios PEs pueden convivir en un único nodo.

Teradata puede ejecutar múltiples AMPs y PEs en un único nodo gracias a la existencia de las PDE (Parallel Database Extensions). Las PDEs conforman la infraestructura necesaria para que la base de datos pueda trabajar en paralelo en entornos NCR UNIX SVR4 MP-RAS o Microsoft Windows. La base de datos funciona independientemente del sistema operativo que tengamos. Teradata Database siempre es Teradata, independientemente de del sistema operativo o plataforma hardware.

## Paralelismo intranodo

En la Figura 2 presentamos un ejemplo de paralelismo intranodo. La configuración representada es un nodo SMP de 2 CPUs con diez VPROCs. En esta configuración hay ocho procesadores virtuales AMP y dos PEs. Cada PE tiene acceso a cada AMP, y esto permite que cada solicitud puede procesarse totalmente en paralelo. Los datos se distribuyen uniformemente entre todos los AMPs.

## Procesamiento paralelo masivo (MPP)

Pero Teradata saca mucho más partido de esta arquitectura en aspectos de escalabilidad y disponibilidad. Es

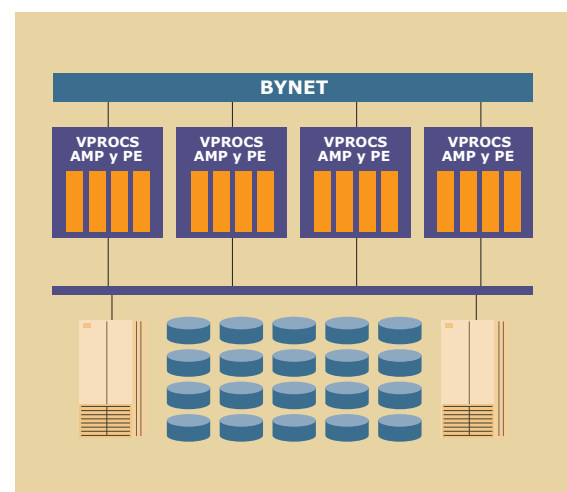
## Disponibilidad ampliada de Teradata

El software de Teradata ofrece una alta disponibilidad, mayor que otros RDBMSs porque:

### ... compensa posibles fallos de hardware:

- > Redistribuye automáticamente la carga de trabajo entre los nodos activos (migrando VPROCs)
- > Realiza copias de seguridad online (fallback)

... permite el acceso a todos los datos, incluso con nodos inactivos (sistema multinodo)



posible interconectar varios nodos SMP, tanto bajo NCR UNIX SVR4 MP-RAS como bajo Windows, mediante la red de alta velocidad Teradata BYNET, para formar un sistema MPP(Massive Parallel Processing) que se gestiona y opera como si de un único sistema se tratara. Esta arquitectura es la clave de la capacidad de crecimiento lineal de Teradata desde un entorno SMP de dos procesadores hasta miles de procesadores físicos y decenas de miles de VPROCs. Ello permite a Teradata operar con éxito tanto en soluciones para pequeñas empresas como en grandes data warehouses corporativos.

**BYNET: interconexión escalable**

BYNET es una interconexión de circuitos de alta velocidad inteligente, redundante y tolerante a fallos. Mediante este elemento, Teradata coordina y sincroniza las actividades de multitud de nodos SMP sin aumentar el tráfico en la red ni afectar al rendimiento cuando el sistema crece. La red BYNET proporciona canales bidireccionales de conexión nodo a nodo con un ancho de banda de 120 MB por segundo, pudiendo crecer linealmente hasta admitir 512 nodos en un solo sistema.

**Plataforma hardware escalable**

La escalabilidad de Teradata se integra en la plataforma. La familia de servidores NCR sobre Intel crece de forma transparente, integrándose con su arquitectura. El crecimiento se logra simplemente añadiendo más nodos de proceso y subsistemas de almacenamiento a la confi-

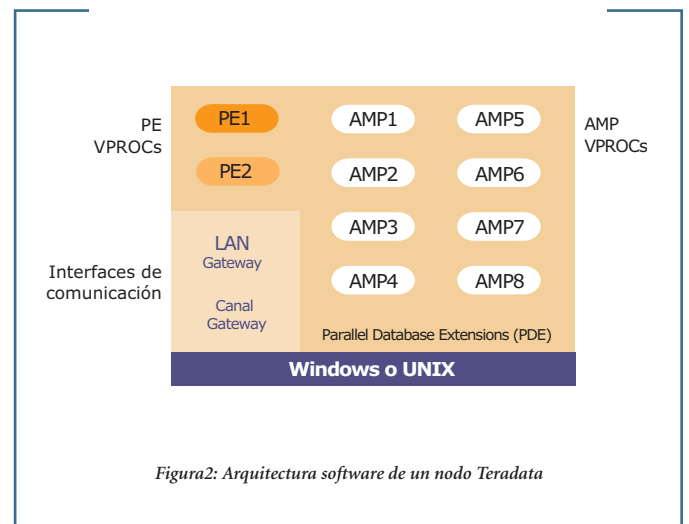
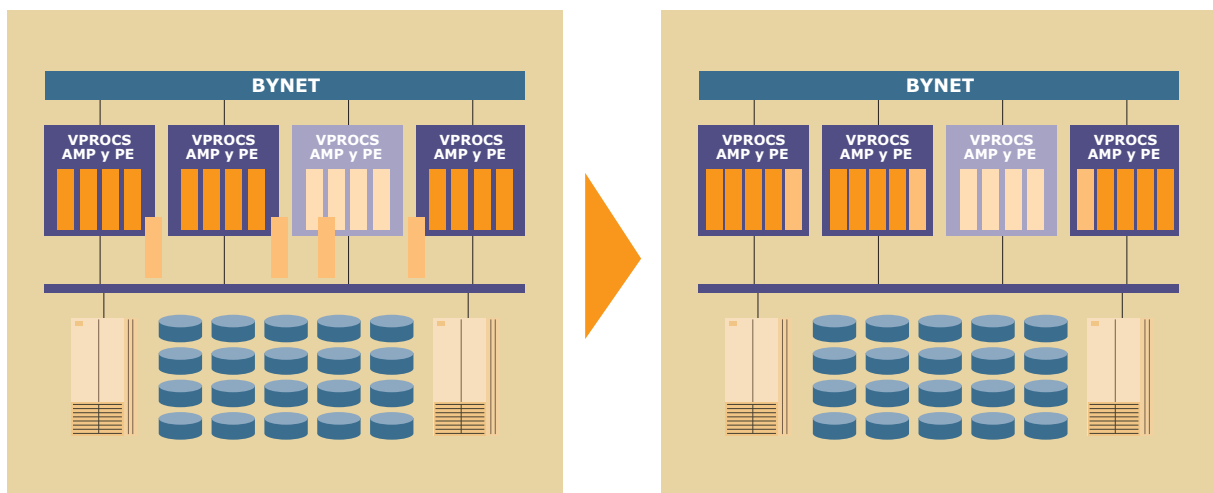


Figura2: Arquitectura software de un nodo Teradata

guración. El sistema operativo los reconocerá automáticamente y Teradata redistribuirá la información existente para aprovechar los nuevos recursos. Las aplicaciones existentes no necesitarán ningún cambio para seguir funcionando como hasta entonces.

**La arquitectura única de Teradata está concebida para soportar aplicaciones redundantes de respaldo a decisiones vitales para su negocio.**



# Teradata: probada, sencilla, potente, eficiente

## Interfaces de programación de aplicaciones (APIs)

Teradata proporciona un conjunto de interfaces estándar para facilitar el desarrollo de aplicaciones cliente/servidor. Entre otras se incluyen: Teradata ODBC Driver, Teradata Call Level Interface (CLI), Teradata JDBC driver, Teradata OLE DB y TS/API, que permite a las aplicaciones desarrolladas para IBM DB2 en el host acceder a Teradata de forma transparente.

## Preprocesadores

Teradata proporciona un conjunto de preprocesadores que facilitan el desarrollo de aplicaciones en lenguajes como COBOL, C y PL/1. Sirviéndose de las librerías de estos preprocesadores se pueden escribir aplicaciones en entorno cliente o servidor/host que accedan a Teradata.

## Utilidades

Teradata proporciona un conjunto de utilidades, tanto en entorno cliente o servidor/host, para que usuarios y administradores accedan o controlen el entorno Teradata. El paquete de Utilidades de Teradata es la caja de herramientas con las que administrar, realizar queries, elaborar informes y conectarse a la base de datos. Además de herramientas de gestión, Teradata incorpora un conjunto de utilidades de carga y extracción en paralelo, que se encargan de cargar la información tanto en batch como en tiempo real.

## Herramientas de administración

Teradata cuenta con una rica colección de herramientas que facilitan el control de las operaciones, administración y mantenimiento de la base de datos. Entre ellas están las soluciones BAR de Teradata (copias de seguridad-archivo-restauración) que combinan la potencia de partners líderes en sus respectivos sectores con la experiencia de Teradata; el Teradata Dashboard para analizar el estado de la base de datos y realizar estadísticas de rendimiento; la Consola de Administración (AWS), punto único desde el que administrar y controlar todo el entorno hardware; y el Teradata Dynamic Query Manager, que gestiona la carga de accesos. Todas estas herramientas, y muchas otras, pueden manejarse directamente o a través de un interfaz común llamado Teradata Manager en entornos Windows.

Teradata ofrece un conjunto de potentes utilidades con las que realizar y optimizar consultas complejas. El Teradata Analyst Pack incluye Teradata Visual Explain, Teradata System Emulation Tool, Teradata Index Wizard y Teradata Statistics Wizard. Estas herramientas simplifican el trabajo del administrador de la base de datos (DBA) y del planificador de consultas, permitiéndoles analizar planes de consultas complejas de forma más eficiente.



## Banca

El sector financiero utiliza Teradata para la banca de relaciones y márketing de venta cruzada. Los datos obtenidos de distintas áreas geográficas, de distintas líneas de negocio, incluidas cuentas de cheques y de ahorro, préstamos para la compra de coches, hipotecas, tarjetas de crédito y cajeros automáticos, así como diversos sistemas electrónicos en línea, se analizan para obtener el correcto perfil del cliente.



## Retail

La industria del retail utiliza Teradata para recolectar y analizar datos detallados procedentes de los scanners de sus miles de tiendas de meses y años. Esta información se utiliza para gestionar las compras, precios, stocks e inventarios, y tomar decisiones relativas a la configuración de las tiendas.



## Acceso a Teradata a través de internet

A través de internet su empresa puede estar presente en mercados de todo el mundo. Teradata le ofrece dos métodos habituales para acceder desde internet a la información almacenada en la base de datos: Java y CGI.

### Java

El Gateway Teradata para Java proporciona a los desarrolladores de aplicaciones un API sencillo de usar para poder acceder a Teradata desde internet o desde una intranet. Cualquier cliente capaz de ejecutar un applet Java o un navegador como Netscape Navigator o Microsoft Internet Explorer puede acceder a la base de datos de Teradata directamente.

### Acceso por CGI

Common Gateway Interface (CGI) es otro estándar utilizado para facilitar el acceso desde aplicaciones basadas en Web a la de base de datos. Nuestra solución CGI para Teradata permite incluir sentencias SQL en páginas HTML, incorporando además un mecanismo para devolver el resultado en el mismo formato. Valida los parámetros recibidos a través de la cadena de consultas en HTTP y permite todas las construcciones propias del lenguaje de manipulación de información (DML), incluidas SELECCIONAR, INSERTAR, ACTUALIZAR y BORRAR.

Esfuerzo de Administración	Otras RDBMS	Teradata
Modelo lógico de datos	Alto	Alto
Modelo físico de datos	Alto	Bajo
Definición de particiones de datos	Alto	Bajo
Definición de ubicación de datos	Alto	Ninguno
Gestión del espacio libre	Alto	Ninguno
Control del balanceo de datos	Alto	Bajo
Reorganización de datos	Alto	Ninguno
Reorganización de índices	Alto	Ninguno
Ajuste de consultas	Alto	Ninguno
Gestión de carga de trabajo	Moderado	Ninguno
Gestión de espacio de trabajo	Moderado	Ninguno
Gestión de cambios	Alto	Bajo



## Fabricación de bienes de consumo

Esta industria utiliza Teradata para determinar la manera más eficiente de suministrar mercancías a los minoristas. Pueden determinar la cantidad de productos a fabricar y a qué precio venderlos, así como el mejor ritmo de fabricación para entregar los productos a tiempo, justo cuando el cliente va a demandarlos.



## Telecomunicaciones

El sector de las telecomunicaciones utiliza Teradata para recolectar y analizar datos de millones de clientes, circuitos, facturas mensuales, servicios prestados, equipos vendidos, tráfico de red, y más. Esta información detallada sobre ingresos, beneficios y costes se utiliza para gestionar las campañas de marketing, análisis financiero, gestión de inventarios, gestión de compras y gestionar adecuadamente la red.

# Teradata: Abierta, escalable, automantenida

Teradata es el gestor utilizado en los data warehouse de las compañías líderes mundiales, en todos los sectores. Cuando nuestros clientes confían terabytes de su preciada información estratégica a un motor de bases de datos, esperan que esté disponible 24 horas al día, 7 días por semana, 52 semanas al año. Sólo Teradata cumple y excede sus expectativas.

## Diez poderosas razones

Diez poderosas razones por las que Teradata es la base de datos para data warehouse

- 1 La base de datos de Teradata**
- 2 Máxima escalabilidad sin esfuerzo**
- 3 Fácil de administrar**
- 4 Menos riesgo**
- 5 Expertos por excelencia en datawarehousing**
- 6 Perfecta integración con el Mainframe**
- 7 Disponibilidad que contribuye a sus objetivos**
- 8 Su inversión, protegida**
- 9 Su solución, lista antes que con nadie**
- 10 Pruebas irrefutables de nuestro buen hacer**

## Apuesta por el líder

¿Por qué seleccionar Teradata para su data warehouse? Porque hemos creado y gestionado los proyectos de data warehouse para cientos de compañías de todo el mundo.

Durante los últimos 20 años, Teradata se ha convertido en el líder en soluciones analíticas. ¿Por qué? Porque ofrecemos una combinación única de experiencia y conocimiento para ayudar a las empresas a tomar decisiones rápidas e inteligentes. Y porque con nuestras soluciones de data warehouse, CRM, e-business y otras aplicaciones, Teradata puede ayudarle a aumentar sus ingresos e incrementar la fidelidad de sus clientes. Nuestras soluciones analíticas le proporcionan una visión única de su empresa, una visión que crece y se adapta al ritmo de su organización.

Y por supuesto, Teradata, una división de NCR, tiene un conocimiento profundo de la industria, experiencia en consultoría, servicios de soporte globales y las mejores plataformas hardware: una combinación sin comparación en el mercado.

El resultado, Teradata le proporciona el mejor arma para luchar en la batalla de la economía global: la información. Información que puede acceder y analizar para encontrar nuevas y mejores formas de acercarse al cliente, sustituir productos y gestionar ciclos de vida: todo lo que necesita para adelantarse a su competencia. Y a medida que su negocio necesite crecer y adaptarse, Teradata protege su inversión en hardware y software con nuestras plataformas NCR escalables, modulares y de alta disponibilidad. En definitiva, Teradata constituye una solución segura con la que adentrarse con paso firme en el futuro.

Para más información acerca de la base de datos de Teradata, visítenos en [Teradata.com](http://Teradata.com), o solicítela a [teradata.spain@ncr.com](mailto:teradata.spain@ncr.com).



## Teradata Warehouse Especificaciones

<b>Capacidad de la base de datos</b>	<ul style="list-style-type: none"> <li>&gt; 1.023 Petabytes (PB)</li> <li>&gt; 512 nodos</li> </ul>
<b>Modelos de datos</b>	<ul style="list-style-type: none"> <li>&gt; Relacional</li> <li>&gt; Compatible con ANSI SQL-92</li> <li>&gt; Completamente paralela</li> </ul>
<b>Arquitecturas soportadas</b>	<ul style="list-style-type: none"> <li>&gt; Multiproceso simétrico (SMP)</li> <li>&gt; Proceso paralelo masivo (MPP)</li> </ul>
<b>Sistemas operativos soportados</b>	<ul style="list-style-type: none"> <li>&gt; UNIX SVR4 MP-RAS de NCR</li> <li>&gt; Microsoft Windows</li> </ul>
<b>Plataformas cliente soportadas</b>	<ul style="list-style-type: none"> <li>&gt; Microsoft MS-DOS</li> <li>&gt; Microsoft Windows</li> <li>&gt; IBM VM, MVS, y OS/2</li> <li>&gt; NCR UNIX SVR4 MP-RAS, y la mayoría de las demás plataformas UNIX</li> </ul>
<b>Integración con Mainframe</b>	<ul style="list-style-type: none"> <li>&gt; IBM</li> <li>&gt; Amdahl</li> <li>&gt; Bull</li> </ul>
<b>Preprocesadores</b>	<ul style="list-style-type: none"> <li>&gt; PL/1, C, COBOL</li> </ul>
<b>Herramientas y utilidades</b> <b>Teradata Utility Pack</b>	<ul style="list-style-type: none"> <li>&gt; BTEQ (consultas/redacción de informes)</li> <li>&gt; Teradata SQL Assistant (consultas/DBA)</li> <li>&gt; Teradata SQL Assistant Web edition (consultas/DBA)</li> <li>&gt; ODBC</li> <li>&gt; JDBC</li> <li>&gt; OLE DB provider</li> <li>&gt; Teradata Call Level Interface (múltiples plataformas)</li> <li>&gt; Teradata MultiTool</li> </ul>
<b>Base de datos de Teradata</b> <b>Utilidades de administración</b>	<ul style="list-style-type: none"> <li>&gt; Teradata Manager (administración del sistema)</li> <li>&gt; Teradata Dynamic Query Manager (gestión de carga de consultas)</li> <li>&gt; Teradata Metadata Services</li> <li>&gt; Teradata Analyst Pack, compuesto por Teradata Visual Explain, Teradata System Emulation Tool, Teradata Index Wizard y Teradata Statistics Wizard</li> </ul>
<b>Utilidades de carga paralela</b>	<ul style="list-style-type: none"> <li>&gt; TPump (actualización de datos)</li> <li>&gt; FastLoad (carga de datos)</li> <li>&gt; MultiLoad (carga/actualización de datos)</li> <li>&gt; FastExport (extracción de datos)</li> </ul>
<b>Otras herramientas y utilidades</b>	<ul style="list-style-type: none"> <li>&gt; ARCMAN (back-up en Mainframe)</li> <li>&gt; TS/API</li> </ul>

Teradata y NCR son marcas registradas. BYNET es marca registrada propiedad de NCR Corporation. UNIX es una marca registrada propiedad de X/OPEN Limited. Microsoft, Windows y Windows NT son marcas registradas propiedad de Microsoft Corporation. Intel es una marca registrada propiedad de Intel Corporation. NCR mejora constantemente sus productos a medida que surgen nuevas tecnologías y componentes. Por consiguiente, NCR, se reserva el derecho de modificar sus especificaciones sin previo aviso. Es posible que las características, capacidades y operaciones descritas en el presente folleto no se comercialicen en todo el mundo. Consulte al respecto con su representante de Teradata o visite la web [Teradata.com](http://Teradata.com).

(c) 2003 NCR Corporation Dayton, OH U.S.A. Impreso en EEUU. Todos los derechos reservados.

Nunca antes había vista su negocio así.

**Teradata**  
una división de  NCR