

# Teradata (Teórica)



# Introducción

Bienvenidos a la unidad de Teradata.

En este módulo abordaremos una introducción a los Data Warehouses como puerta de entrada para avanzar sobre la arquitectura y usos de Teradata para la administración de bases de datos.

A lo largo de la unidad comprenderemos en detalle el funcionamiento del motor de búsqueda de teradata, su arquitectura de almacenamiento y recuperación, sus componentes principales (AMP, VPROC, PE) y su estrategia de paralelismo por la que se destaca.

Una vez finalizado este módulo serás capaz de:

- Comprender distintos casos de uso en el cual se puede aplicar Teradata como solución a situaciones de grandes volúmenes de datos y compatibilidad de integración con distintos sistemas.



# Tema 1. Data Warehouse

## Objetivos

El objetivo de este primer tema es conocer a qué nos referimos cuando hablamos de Data Warehouses y cuales son sus ventajas.

## ¿Qué es un Data Warehouse?

Un data warehouse es un repositorio unificado para todos los datos que recogen los diversos sistemas de una organización. El repositorio puede ser físico o lógico y hace hincapié en la captura de datos de diversas fuentes sobre todo para fines analíticos y de acceso conjunto. Normalmente, un data warehouse se aloja en un servidor corporativo o bien, en la nube siendo una arquitectura de almacenamiento de datos que permite a los ejecutivos de negocios organizar, comprender y utilizar sus datos para tomar decisiones estratégicas. Un data warehouse es una arquitectura utilizada en muchas empresas modernas.

Los datos de diferentes aplicaciones de procesamiento de transacciones Online (OLTP) y otras fuentes se extraen selectivamente para su uso por aplicaciones analíticas y de consultas por usuarios.

## Estructuras de un Data Warehouse

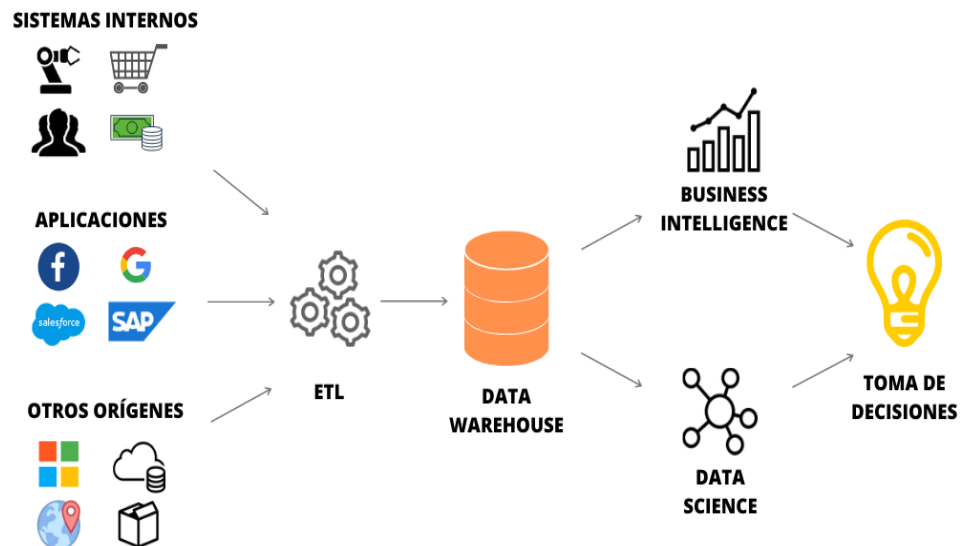
La arquitectura de un data warehouse puede ser dividida en tres estructuras simplificadas: básica, básica con un área de ensayo y básica con área de ensayo y data marts.

Con una estructura básica, sistemas operativos y archivos planos proporcionan datos en bruto que se almacenan junto con metadatos. Los usuarios finales pueden acceder a ellos para su análisis, generación de informes y minería.

Al añadir un área de ensayo que se puede colocar entre las fuentes de datos y el almacén, ésta proporciona un lugar donde los datos se pueden limpiar antes de entrar en el almacén. Es posible personalizar la arquitectura del almacén para diferentes grupos dentro de la organización.

Se puede hacer agregando data marts, que son sistemas diseñados para una línea de negocio en particular. Se pueden tener data marts separados para ventas, inventario y compras, por ejemplo, y los usuarios finales pueden acceder a datos de uno o de todos los data marts del departamento.

Un data mart lleva a cabo las mismas funciones que un almacén de datos, pero con un alcance mucho más limitado (por lo general, para un solo departamento o línea de negocio). Es por eso que es más fácil establecer un data mart que un data warehouse. Sin embargo, suelen introducir incoherencias, ya que es difícil administrar y controlar los datos de manera uniforme en muchos data marts.



Un almacén de datos comúnmente incluye los siguientes elementos:

- Una base de datos relacional para almacenar y gestionar los datos.
- Una solución de extracción, carga y transformación (ELT) para preparar los datos para el análisis.
- Análisis estadísticos, informes y funciones de extracción de datos.
- Herramientas de análisis de clientes para visualizar y presentar datos a usuarios de negocio.
- Otras aplicaciones analíticas más sofisticadas generan información procesable mediante la aplicación de algoritmos de ciencia de datos e inteligencia artificial (IA)

El ETL es un tipo de integración de datos que hace referencia a una secuencia de tres pasos: extraer, transformar, cargar que se utilizan para combinar datos provenientes de múltiples fuentes y construir un almacén de datos (data warehouse). Durante este proceso, los datos se toman (extraen) de un sistema de origen, se convierten (transforman) en un formato que se puede almacenar y se almacenan (cargan) en un data warehouse

## Data Warehouse vs Data Lake

Cuando las organizaciones tienen que gestionar grandes volúmenes de datos de diversas fuentes, utilizan tanto data lakes como data warehouses. La opción de usar un sistema u otro depende de lo que la organización quiera hacer con los datos. A continuación, describimos cómo utilizar uno y otro:

- Los data lakes almacenan una gran cantidad de datos sin filtrar y de todo tipo, a fin de usarlos más adelante para un fin determinado. En un data lake, los datos de aplicaciones de áreas de negocio, aplicaciones móviles, redes sociales, dispositivos de IoT, etc. se obtienen como datos sin procesar. La estructura, la integridad, la selección y el formato de los diversos conjuntos de datos se adquiere en el momento del análisis por parte de la persona que lo realiza. Cuando las organizaciones necesitan un almacenamiento de bajo costo para datos sin estructura ni formato y de múltiples fuentes (a fin de usarlos en el futuro para algún propósito en particular), los data lakes podrían ser la opción ideal.
- Los data warehouses están diseñados específicamente para analizar datos. Dentro de un almacén de datos, el procesamiento analítico se lleva a cabo con datos previamente preparados para su análisis —recopilados, contextualizados y transformados— con el fin de generar conocimientos basados en análisis. Los data warehouses también son perfectos para gestionar grandes cantidades de datos de diversas fuentes. Cuando las organizaciones necesitan una analítica de datos avanzada o un análisis que se base en datos históricos de múltiples fuentes de toda la empresa, es probable que el almacén de datos sea la mejor opción.



## Tema 2. Arquitectura de Teradata

### Objetivos

El objetivo del siguiente tema es conocer la infraestructura de Teradata, sus alcances y componentes.

### El motor de teradata

Teradata es un gestor de base de datos relacional específicamente diseñado para soportar paralelismo. Su arquitectura patentada permite descomponer las consultas 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.

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. La base de datos funciona independientemente del sistema operativo que tengamos ya que Teradata Database siempre es Teradata, independientemente del sistema operativo o plataforma hardware.

En base a la principal función de Teradata, la arquitectura se puede categorizar en dos secciones:

- Arquitectura de almacenamiento

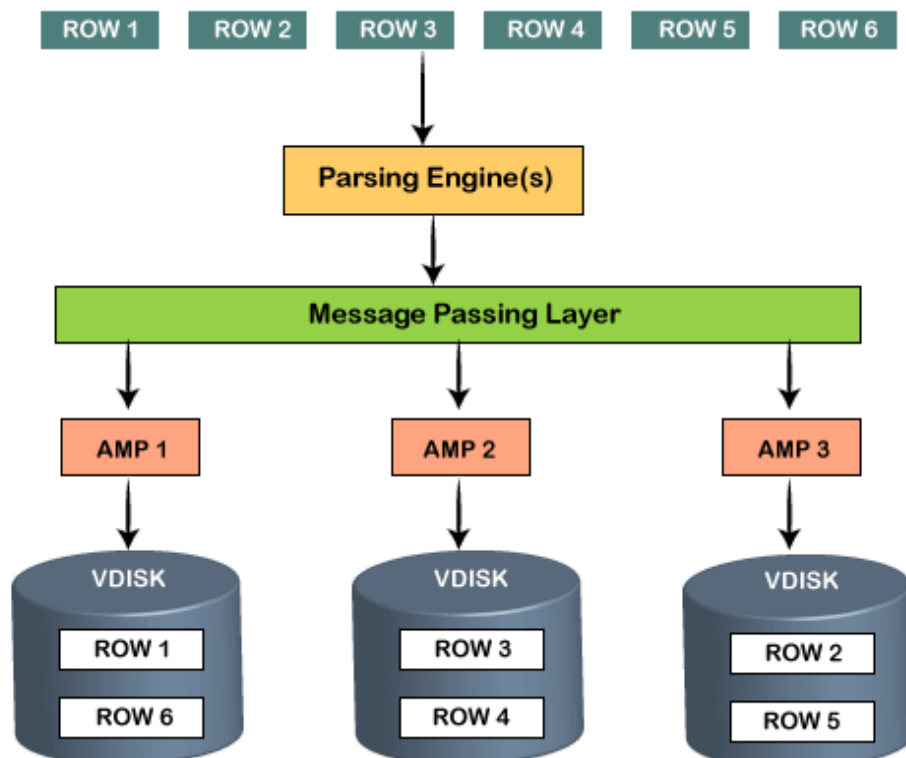
- Arquitectura de recuperación

### Arquitectura de almacenamiento

La arquitectura de almacenamiento consta de los dos componentes anteriores de la arquitectura de Teradata. Cuando el cliente ejecuta consultas para insertar registros, el motor de análisis envía los archivos a BYNET. Allí BYNET recupera los archivos y envía la fila al AMP de destino. El AMP almacena estos registros en sus discos.

### Arquitectura de recuperación

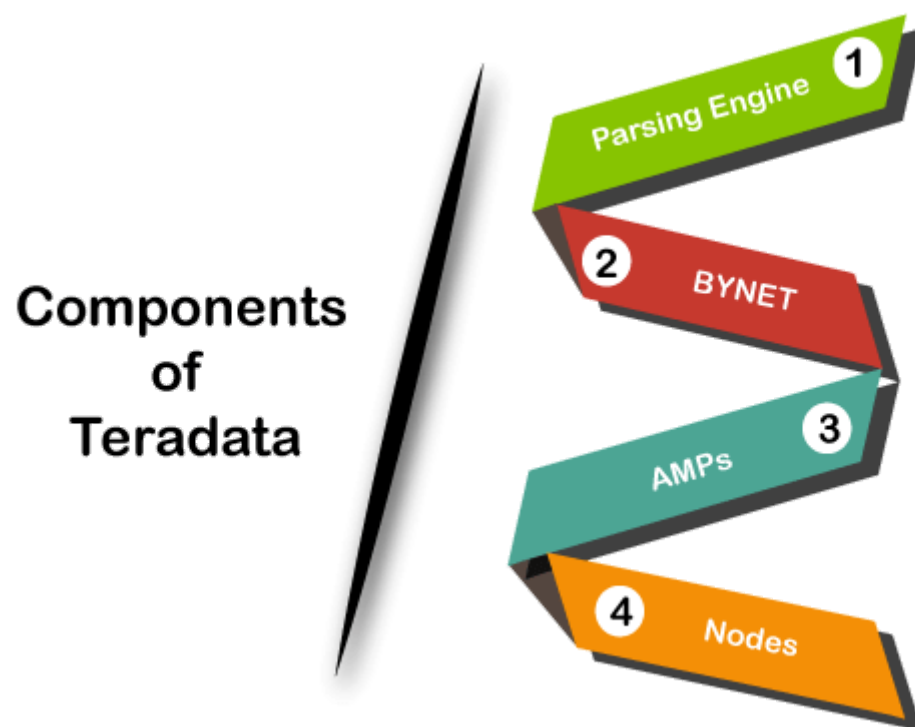
Esta sección de la arquitectura consiste en dos partes. Cuando el cliente ejecuta consultas para obtener registros, el motor de análisis envía una solicitud a BYNET. Allí BYNET envía la solicitud de recuperación a los AMP apropiados. Luego, los AMP buscan sus discos en paralelo e identifican los registros requeridos y los reenvían a BYNET. Luego BYNET envía los registros a Parsing Engine, que a su vez los devolverá al cliente.



## Componentes de teradata

Teradata es un sistema de administración de bases de datos relacionales muy eficiente, menos costoso y de alta calidad. Teradata se basa en la arquitectura Massively Parallel Processing (MPP). Está compuesto por un motor de análisis (PE), BYNET, procesadores de módulo de acceso (AMP) y otros componentes como nodos.

Teradata tiene cuatro componentes principales como figura en la siguiente imagen



A continuación haremos una breve descripción y luego veremos en más detalle cada uno de los puntos

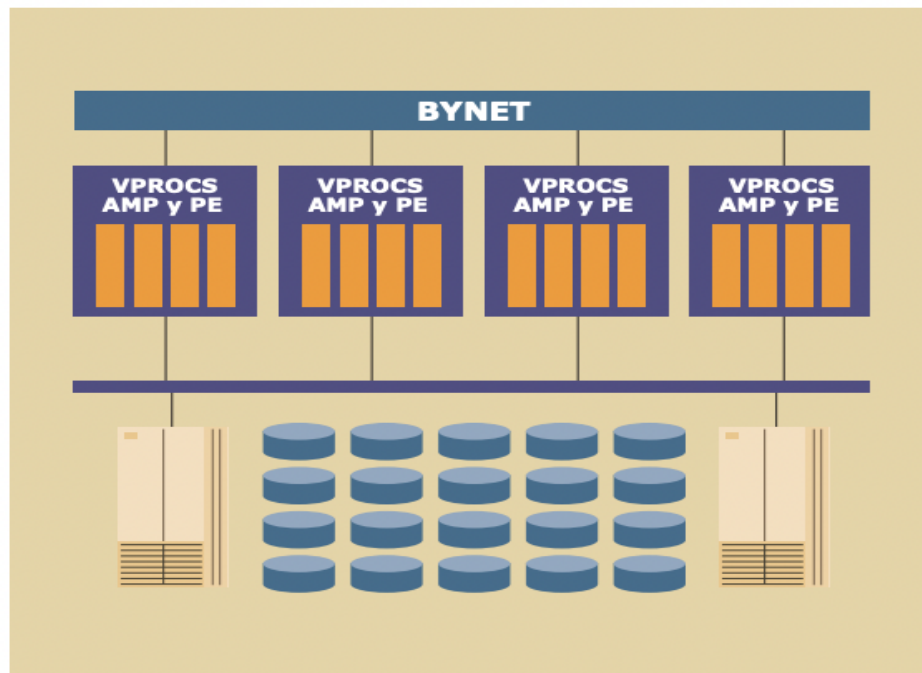
**Parsing Engine:** Parsing Engine es el componente base que recibe las consultas y prepara el plan de ejecución.

**Access Module Processors (AMPs):** Estos son los procesadores virtuales de Teradata. Reciben el plan de ejecución y los datos del motor de análisis. Los datos se someterán a la conversión, el filtrado, la agregación, la clasificación, etc. necesarios, y se enviarán posteriormente a los discos correspondientes para su almacenamiento. Los registros de la tabla se distribuirán a cada AMP para el almacenamiento de datos. Solo ese AMP puede leer o escribir datos en los discos que tienen permiso de acceso

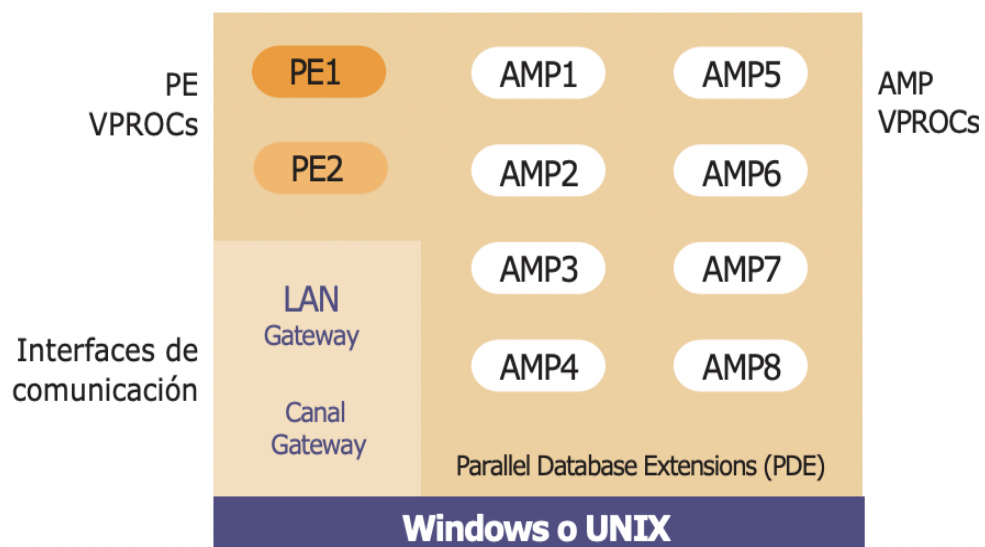


**Nodes:** La unidad básica de un sistema Teradata se llama Nodo. Cada nodo tiene su sistema operativo, memoria de CPU, una copia del software RDBMS y algo de espacio en disco. Un solo gabinete puede tener uno o más nodos.

### Esquema: Arquitectura de Teradata



### Esquema: Arquitectura de software de un nodo de teradata



## **BYNET: interconexión escalable**

Esta es la capa de paso de mensajes o simplemente la capa de red en Teradata. Recibe el plan de ejecución del motor de análisis y lo pasa a los AMP y los nodos. Después de eso, obtiene la salida procesada de los AMP y la envía de regreso al motor de análisis. Para mantener una disponibilidad adecuada, están disponibles los dos tipos de BYNET 0 y BYNET 1. Esto asegura que un BYNET secundario esté disponible en caso de que falle el BYNET principal.

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.

## **Procesamiento paralelo masivo (MPP)**

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

## **Paralelismo intra-nodo**

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 pueda procesarse totalmente en paralelo. Los datos se distribuyen uniformemente entre todos los AMPs.

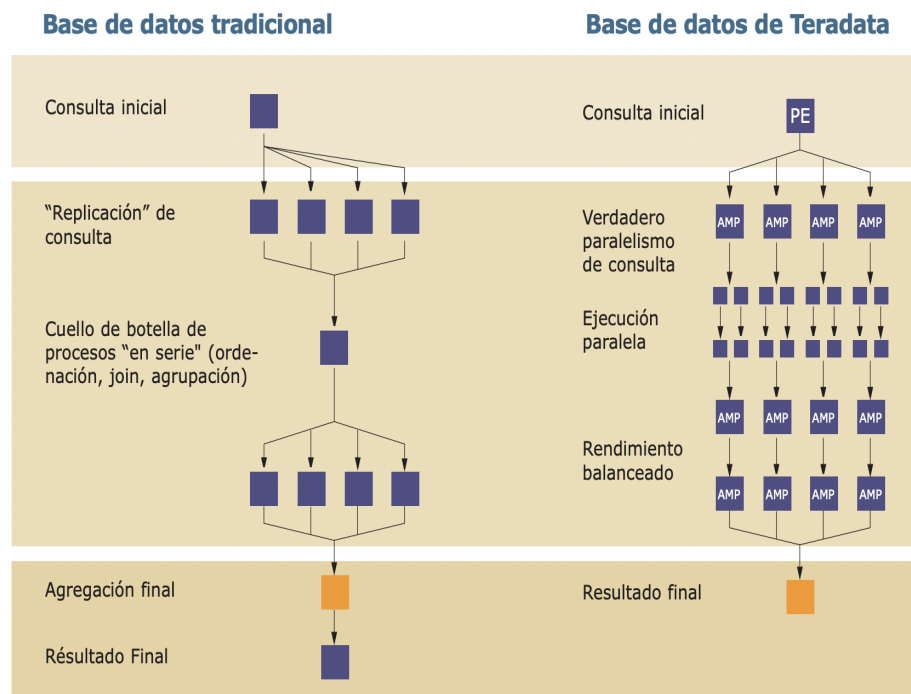
## **Plataforma de hardware escalable**

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 configuración. 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.

## Comparativa de base de datos tradicional y base de datos de Teradata

A continuación se presenta un esquema comparativo entre las bases de datos tradicionales y Teradata



# Tema 3. Access Module Processor (AMP)

## Objetivos

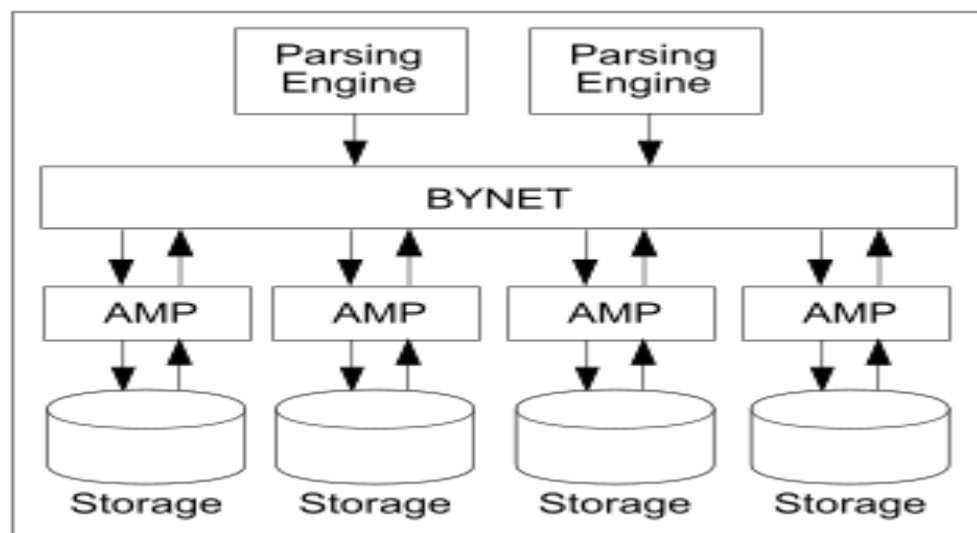
Detallar en profundidad qué son los AMP, cuales son sus funciones y rol que cumplen en Teradata.

## AMP

### ¿Qué es un AMP en Teradata?

El vproc de AMP administra las interacciones de Teradata con el subsistema del disco. Cada AMP administra una parte del almacenamiento del disco.

Cada AMP, como se representa en la siguiente figura, administra una parte del espacio del disco físico. Cada AMP almacena su porción de cada tabla de la base de datos dentro de ese espacio.



AMP es la abreviatura de Access Module Processor. Cada Teradata AMP es un proceso de Linux responsable de manejar su porción individual de datos. La asignación de filas físicas a los AMP se realiza mediante un algoritmo hash que garantiza que todas las filas de una tabla se distribuyan uniformemente entre todos los AMP. La distribución de datos a procesos virtuales es típica de los sistemas de shared-nothing como Teradata. Cada AMP tiene su propia memoria principal y un disco lógico asignado. Cada nodo de Teradata normalmente ejecuta cientos de AMP simultáneamente.

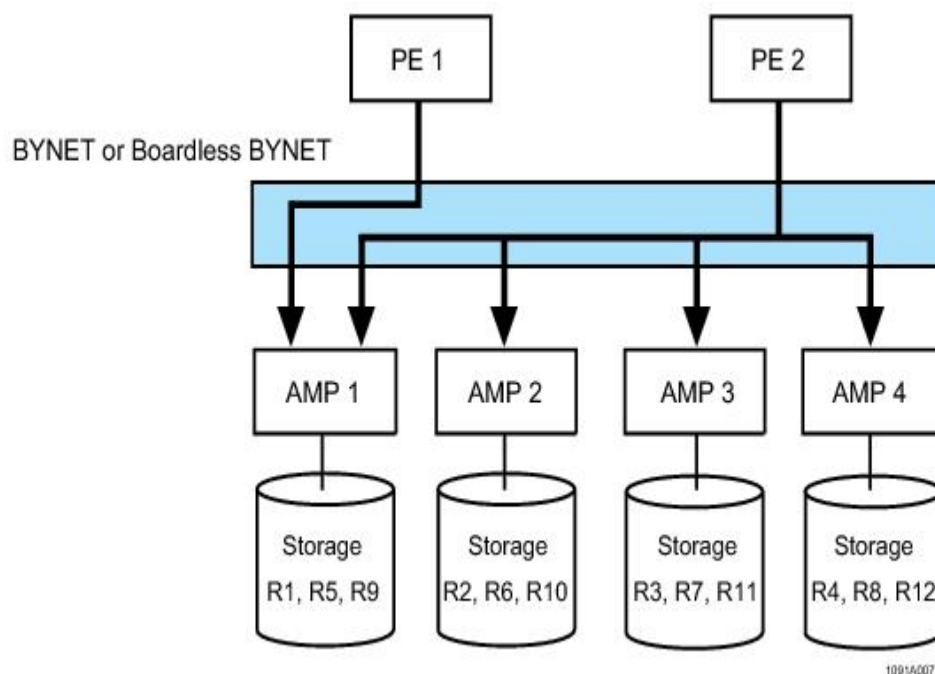
## ¿Cómo funcionan los AMP?

Los AMP obtienen las filas necesarias para procesar las solicitudes (suponiendo que los AMP están procesando una instrucción SELECT). BYNET transmite mensajes hacia y desde AMPs y PE. Un paso AMP se puede enviar a uno de los siguientes:

- Un AMP
- A un grupo selecto de AMPs, denominado grupo BYNET dinámico
- Todos los AMPs en el sistema

En la siguiente figura se observan 2 situaciones:

- Si el acceso es a través de un índice principal y la solicitud es para una sola fila, el PE transmite los pasos a un solo AMP, como se muestra en PE1.
- Si la solicitud es para muchas filas (una solicitud a todos los AMP), el PE hace que BYNET transmita los pasos a todos los AMP, como se muestra en PE2. Para minimizar la sobrecarga del sistema, el PE puede enviar un paso a un subconjunto de AMP, cuando sea apropiado.



## ¿Cuáles son las tareas de un AMP?

Cada AMP de Teradata es capaz de realizar todas las tareas relacionadas con sus filas. Esto incluye leer y almacenar filas, agregar filas, realizar uniones, bloquear, ordenar filas, administrar el espacio en disco y convertir fuentes cuando se crea un conjunto de resultados.

## ¿Qué es un AMP Worker Task (AWT)?

Cada AMP puede asignar un plan de pasos recibido a una de las 80 tareas de trabajador de AMP disponibles; el paralelismo no termina en el nivel de AMP, pero cada AMP puede ejecutar estas 80 tareas en paralelo.

Las AMP Worker Tasks son especializadas y pueden realizar diferentes tipos de tareas, según su asignación. Algunos asumen nuevos trabajos; otros AWT están reservados para continuar el trabajo que se ha iniciado. Finalmente, algunos AWT se reservan exclusivamente para cargas de trabajo.

Los diferentes tipos de carga de trabajo requieren diferentes números de tareas de trabajo de AMP. Normalmente una instrucción SQL necesita 1 o 2 tareas de trabajo por AMP, mientras que Teradata Fastload necesita 3 tareas de trabajo de AMP en la primera fase (fase de adquisición) y 1 tarea de trabajo de AMP en la fase de aplicación (todas por AMP). Esta es la razón por la cual la cantidad de cargas rápidas y cargas múltiples que se ejecutan en paralelo está limitada a un pequeño rango de dos dígitos (la configuración predeterminada se puede ajustar según sea necesario)

Dado que las tareas de trabajador de AMP son un recurso crucial, cada AMP reserva 24 tareas de trabajador para trabajo interno. Esto ayuda a evitar situaciones de bloqueo y estancamiento.

Funciones de AMP	Por ejemplo
tareas de administración de base de datos	Ordenar Joinear Agregación Bloqueo de tablas, filas Conversión de datos de salida
administración del sistema de archivos	administración del espacio en disco

### **¿Qué ocurre cuando todas las tareas de un AMP están en uso?**

Si un AMP no tiene disponible una tarea de trabajo del tipo requerido, las tareas se recopilan en una cola de mensajes. Dado que la cola de mensajes solo tiene una memoria limitada, el AMP envía un mensaje al motor de análisis (PE) cuando la cola de mensajes está llena. Luego, el motor de análisis aborta la tarea en todos los AMP y vuelve a intentarlo después de un breve período de tiempo (en el rango de los milisegundos). Esto sucede en un bucle hasta que el AMP tiene tareas de trabajador de AMP libres para esta tarea específica. En cada intento, aumenta el tiempo de espera hasta el siguiente intento. Esto también se denomina modo de control de flujo y, por lo general, significa que el rendimiento de todo el sistema Teradata se ve afectado.

### **¿Están todos los AMP de Teradata siempre involucrados en una solicitud (request) ?**

Los AMP se pueden usar individualmente, en grupos o en conjunto para completar una tarea. Esto se aplica tanto a pasos de recuperación como a pasos de unión y otros. Las operaciones de un AMP son, por ejemplo, acceder a las filas a través del índice principal. Se requieren todos los AMP para los escaneos de tablas completos y la mayoría de las uniones

### **¿Qué sucede si falla un AMP?**

El acceso a las filas de un AMP es redundante. La protección FALLBACK (en la que un AMP de respaldo reemplaza al AMP fallado) siempre está activada en los sistemas Teradata Intelliflex modernos. Los AMP también pueden migrar a otro nodo si falla un nodo completo. Sin embargo, es una práctica común tener un nodo de espera “en caliente” que esté inmediatamente listo para usar en caso de un problema.

# Tema 4. Virtual Processors (VPROC)

## Objetivos

Conocer los diferentes tipos de Vprocs y su función dentro de la infraestructura de Teradata

## VPROC: Virtual Processors

La versatilidad de Teradata se basa en procesadores virtuales (vprocs) que eliminan la dependencia de procesadores físicos especializados. Vprocs son un conjunto de procesos de software que se ejecutan en un nodo bajo Teradata Parallel Database Extensions (PDE) dentro del entorno multitarea del sistema operativo.

La siguiente tabla contiene información sobre los diferentes tipos de Vprocs:

Tipo de Vproc	Descripción
<b>AMP</b>	Los procesadores del módulo de acceso realizan funciones de base de datos, como ejecutar consultas. Cada AMP posee una parte del almacenamiento total de la base de datos.
<b>GTW</b>	Gateway vprocs proporciona una interfaz de conexión para Teradata.
<b>Node</b>	El nodo vproc maneja el PDE y las funciones del sistema operativo que no están directamente relacionadas con el trabajo de AMP y PE. Node vprocs no se pueden manipular externamente y no aparecen en la salida del Vproc Manager.
<b>PE</b>	Los motores de análisis realizan control de sesión, análisis de consultas, validación de seguridad, optimización de consultas y envío de consultas.



<b>RSG</b>	Relay Services Gateway proporciona una interfaz de conexión para la comunicación con Teradata Data Stream Architecture( DSA).
<b>TVS</b>	Administra el almacenamiento de Teradata. Los AMP adquieren sus porciones de almacenamiento de base de datos a través de TVS vproc.

Un solo sistema puede admitir un máximo de 16 384 vprocs. El número máximo de vprocs por nodo puede ser de hasta 128, pero normalmente está entre 6 y 12.

Cada vproc es una copia separada e independiente del software del procesador, aislada de otros vprocs, pero que comparte algunos de los recursos físicos del nodo, como la memoria y las CPU. Se pueden ejecutar múltiples vprocs en una plataforma SMP o un nodo.

Los Vprocs y las tareas que se ejecutan bajo ellos se comunican mediante mensajes de dirección única, como si estuvieran físicamente aislados entre sí. Esta comunicación de mensajes se realiza mediante el software Boardless BYNET Driver en plataformas de un solo nodo o el hardware BYNET y el software BYNET Driver en plataformas multinodo.

# Tema 5. Parsing Engine (PE)

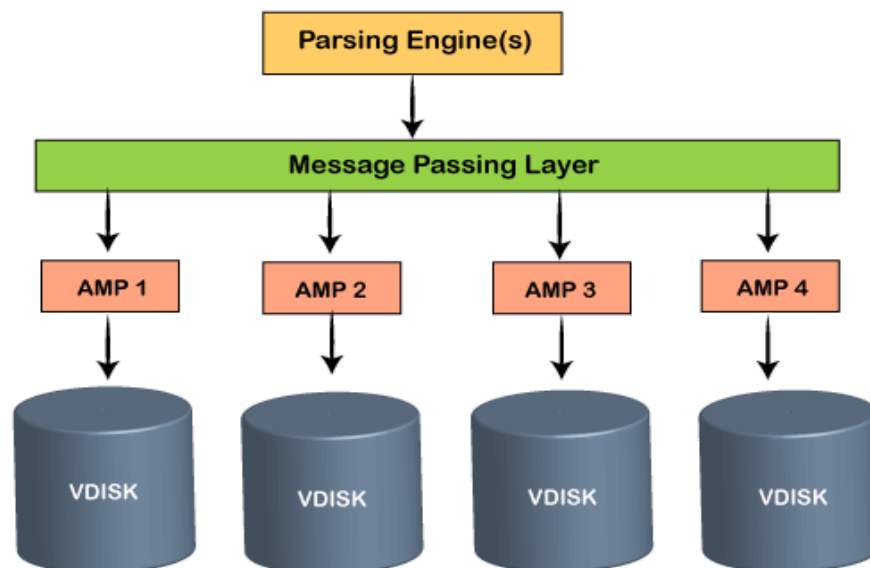
## Objetivos

Conocer las funciones del PE (parsing engine) en la infraestructura de Teradata

## PE: Parsing Engine

El PE (Parsing Engine) es el vproc que se comunica con el sistema cliente por un lado y con los AMP (a través de BYNET) por el otro lado.

Cada PE ejecuta el software de la base de datos que administra las sesiones, descompone las declaraciones SQL en pasos, posiblemente en paralelo, y devuelve las filas de respuesta al cliente que las solicitó.



Cuando un usuario inicia una consulta SQL, primero se conecta al motor de análisis. Los procesos, como la planificación y distribución de los datos a AMPS, se realizan aquí. Es el encargado de encontrar el mejor plan óptimo para la ejecución de consultas.

El software PE consta de los siguientes elementos:

Elementos del PE	Procesos
Parser (analizador)	Descompone SQL en pasos de procesamiento de gestión de datos relacionales
Optimizer (optimizador)	Determina la ruta más eficiente para acceder a los datos
Generator (generador)	Genera y empaqueta pasos
Dispatcher (despachante)	Recibe los pasos de procesamiento del analizador y los envía a los AMP apropiados a través de BYNET.  Supervisa la finalización de los pasos y maneja los errores encontrados durante el procesamiento
Session Control (control de sesión)	Administra las actividades de la sesión, como el inicio de sesión, la validación de contraseña y el cierre de sesión. Recupera sesiones después de fallas del cliente o del servidor.

# Tema 6. Consultas SQL en Teradata

## Objetivos

La presentación del siguiente tema tiene por objetivo comprender la posibilidad del uso del lenguaje SQL dentro de Teradata.

## Consultas SQL en Teradata

Teradata admite el uso de ANSI SQL para consultar datos alojados en el sistema de base de datos. Junto con ANSI SQL, también admite extensiones de Teradata que se pueden usar

SQL tiene la ventaja de ser el lenguaje más utilizado para los sistemas de gestión de bases de datos relacionales. Debido a esto, tanto las estructuras de datos en Teradata como los comandos para manipular esas estructuras se controlan mediante SQL. Además, todas las aplicaciones, incluidas las escritas en un lenguaje de cliente con SQL integrado, macros y consultas SQL ad hoc, se escriben y ejecutan utilizando el mismo conjunto de instrucciones y sintaxis.

Otros sistemas de administración de bases de datos utilizan diferentes lenguajes para la definición y manipulación de datos y es posible que no permitan consultas ad-hoc de la base de datos. Teradata permite usar un único idioma para definir, consultar y actualizar sus datos

## Sentencias SQL

El lenguaje SQL permite, mediante instrucciones SQL, definir objetos de la base de datos, definir el acceso de los usuarios a esos objetos y manipular los datos almacenados.

Estas funciones forman las principales familias de sentencias SQL:

- Data Definition Language (DDL)
- Data Control Language (DCL)
- Data Manipulation Language (DML)

Además, SQL proporciona declaraciones HELP y SHOW que brindan ayuda sobre las definiciones de objetos de la base de datos, las sesiones y las estadísticas, el modificador de solicitud EXPLAIN, la sintaxis de las declaraciones SQL y la visualización del código SQL utilizado para crear tablas.

Teradata ofrece las siguientes posibilidades de invocar una sentencia ejecutable:

- Interactuando desde una terminal.
- Embebido en un programa de aplicación.
- Dinámicamente creado con una aplicación embebida.
- Embebida en un stored procedure o un stored procedure externo.
- Dinámicamente creado con un stored procedure.
- A través de un trigger.
- Embebido en una macro.

# Tema 7. ¿Por qué usar Teradata?

## Objetivos

El objetivo del tema es comprender las razones por las cuales puede resultar conveniente utilizar Teradata.

## ¿Por qué usar Teradata?

Existen varias razones para elegir Teradata sobre cualquier otra base de datos, a saber:

- Teradata proporciona un conjunto completo de servicios que se centran en el almacén de datos (data warehouse)
- El sistema se basa en una arquitectura abierta. Es por eso que cada vez que proporcionamos un dispositivo más rápido, ya se puede incorporar a la arquitectura de construcción.
- Teradata admite más de 50 petabytes de datos.
- Utiliza servicio de Workstation para proporcionar una vista de operación única para un amplio sistema de múltiples nodos de Teradata.
- Es compatible con una amplia gama de herramientas de BI para obtener los datos.
- Puede actuar como un único punto de control para que el DBA administre la base de datos.
- Admite alto rendimiento, varias consultas, análisis en la base de datos y gestión sofisticada de la carga de trabajo..
- Teradata nos permite obtener los mismos datos en múltiples opciones de implementación.
- Soporta la distribución automática de los datos a través de múltiples procesadores (AMP). Divide las tareas en partes iguales para que todas las partes del sistema se encuentren ocupadas para completar las tareas más rápido.
- Proporciona un optimizador compatible con paralelo que ejecuta las tareas de manera eficiente.
- Es compatible con la función de programación analítica en línea (OLAP) para realizar una función analítica compleja en los datos.

## Cierre

A lo largo de este documento hemos visto una variedad de conceptos importantes. Para empezar se comenzó viendo la definición de lo que es un datawarehouse, su diferencia con una base de datos tradicional y el concepto de data lake con sus respectivas diferencias..

Luego nos introducimos en teradata describiendo su arquitectura general para luego ir en detalle en cada uno de sus componentes, destacando principalmente AMP (Access Module Processor), VPROC (Virtual Processors), PE (Parsing Engine).

Posteriormente vimos las ventajas que posee la arquitectura de Teradata y por qué razones puede ser conveniente utilizarlas.

# Referencias

**¿What is a data warehouse?**

<https://www.oracle.com/ar/database/what-is-a-data-warehouse>

**Teradata tech overview**

[https://www.businessintelligence.info/resources/assets/teradata/teradata\\_tech\\_overview\\_sp.pdf](https://www.businessintelligence.info/resources/assets/teradata/teradata_tech_overview_sp.pdf)

**AMP - Teradata Database**

<https://docs.teradata.com/r/xrmOxQIXa0ETFi5IPxHJaQ/uG5LMkRGEXLztxOk9dJBNw>

**VPROC - Teradata Database**

<https://docs.teradata.com/r/Teradata-VantageTM-Database-Introduction/March-2019/Teradata-Database-Hardware-and-Software-Architecture/Virtual-Processors>

**¿What is Teradata?**

<https://www.javatpoint.com/teradata#:~:text=Teradata%20is%20a%20parallel%20open,by%20a%20company%20named%20Teradata.>



tiiiiiit by 