



Partner

SOLUCIONES DE SIEMON PARA IA

Arquitecturas de Red Emergentes para Centros de Datos de IA y sus Aplicaciones

Desde el front-end tradicional hasta el back-end potenciado por IA

www.siemon.com

Connecting the World to a Higher Standard

Cómo satisfacer las demandas de la IA y las cargas de trabajo modernas

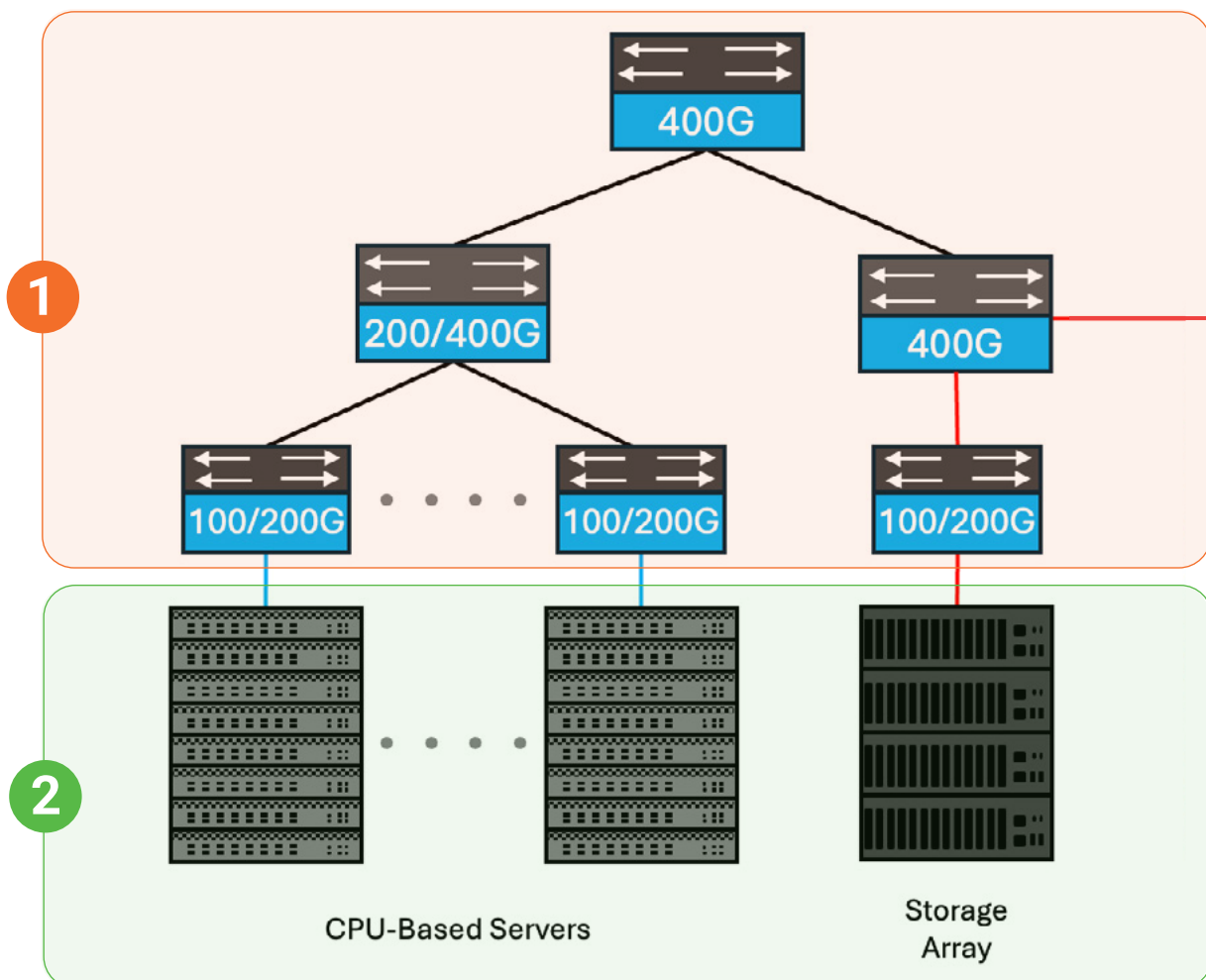
Para respaldar las cargas de trabajo de HPC, como el entrenamiento de IA/ML, las redes de back-end implementan una arquitectura de spine/leaf en la que los switches leaf se conectan a cada switch spine. Dentro de los pods de IA (clústeres) que están diseñados específicamente para realizar tareas específicas, los switches leaf proporcionan interconexiones de alta velocidad y baja latencia entre GPU. Los switches spine y leaf suelen conectarse a 800 Gigabit (G) Ethernet, con posibilidades en el horizonte de incrementar a 1,6 Terabit (T), mientras que las velocidades de switch a servidor son principalmente de 400G a través de los protocolos Ethernet o InfiniBand™. Para garantizar una conectividad sin interrupciones para las cargas de trabajo de uso general y la transferencia de datos hacia y desde la red back-end, el front-end está migrando rápidamente a enlaces de switch a switch Ethernet de 200/400G y enlaces de switch a servidor de 25/50/100G. Una red front-end de alto rendimiento es esencial para acelerar las solicitudes de inferencia de IA de los usuarios, los dispositivos y así obtener una mejor experiencia.

Hay 4 tipos diferentes de áreas de interconexión de red:

1. Front-end, de switch a switch
2. Front-end, switch a servidor
3. Back-end, de switch a switch
4. Back-end, switch a nodo

Veamos cada área de interconexión con más detalle:

Arquitectura front-end tradicional



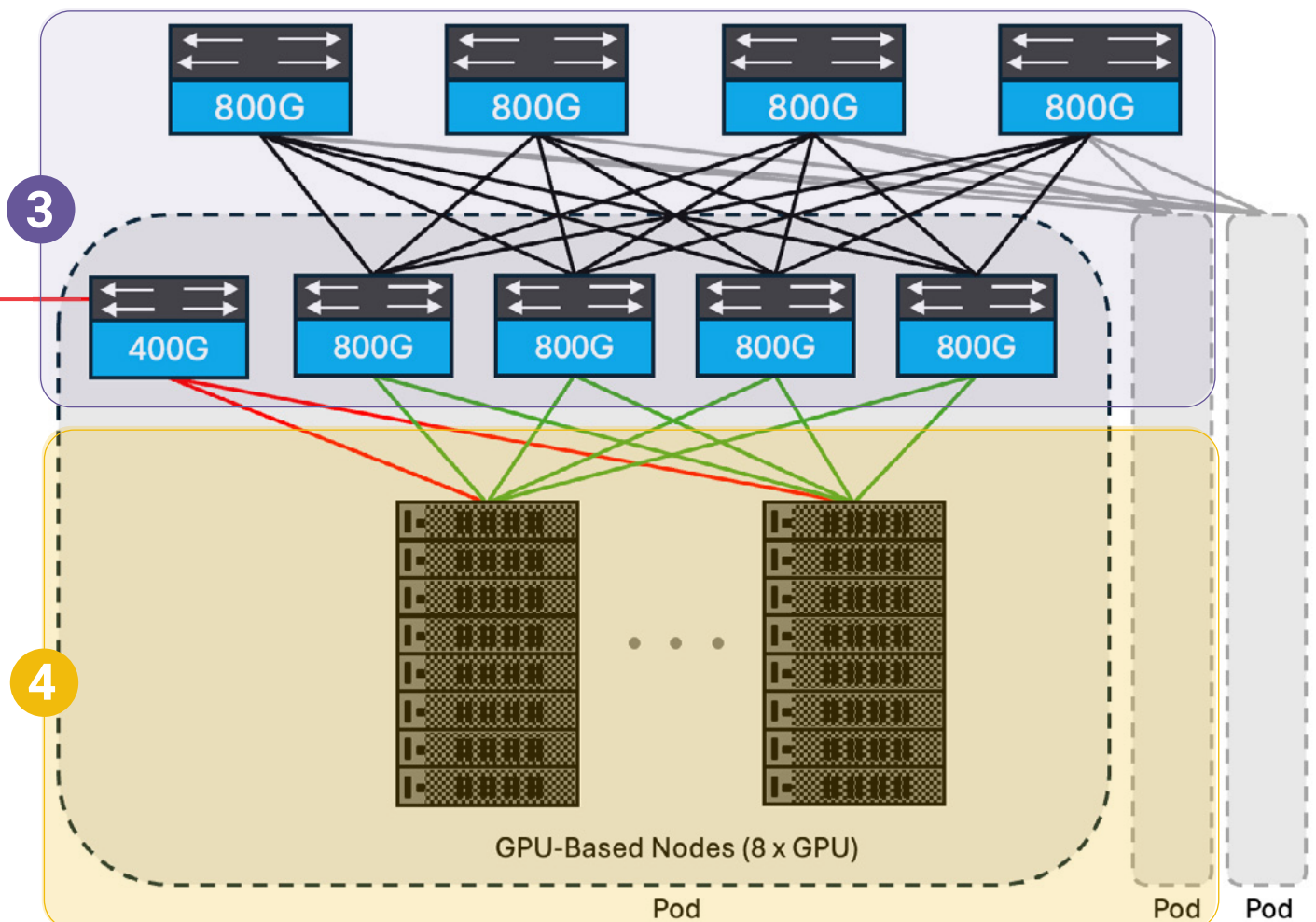
Cerrando la brecha de la IA

El panorama de los Centros de Datos se está transformando gracias a la IA y el ML. Esto requiere una nueva red de back-end en la nube y en los Centros de Datos de grandes empresas, diseñada para cargas de trabajo de HPC como el entrenamiento de IA. Las redes front-end tradicionales están evolucionando para cargas de trabajo de uso general y conexiones externas.

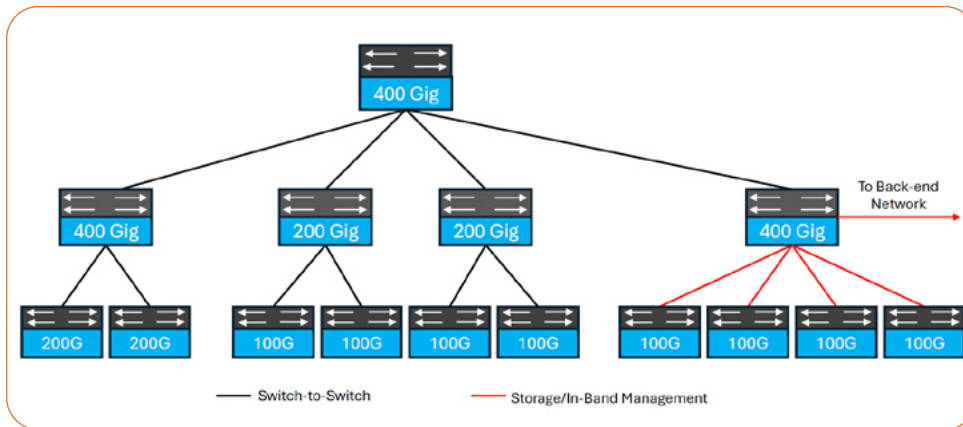
- **Red de back-end:** una red específica **spine - leaf** para la transmisión de datos este-oeste de ultra alta velocidad y baja latencia a través de GPU interconectadas para el entrenamiento de IA.
- **Red front-end:** una red **spine - leaf** tradicional basada en el cliente con conmutación Ethernet, servidores de CPU y sistemas de almacenamiento, que aprovecha la transmisión de datos de norte a sur para la conexión a Internet, el almacenamiento, el movimiento de datos, y las solicitudes de inferencia de IA.

A medida que los Centros de Datos de hoy en día adoptan la revolución de la IA, Siemon comprende la importancia de la infraestructura que conecta a la perfección las redes de back-end de alto rendimiento con el front-end ágil, al tiempo que respalda las necesidades de conectividad únicas de cada entorno. Ya sea que esté alojando redes de back-end y front-end en un solo rack para IA edge, consolidando un modelo básico experimental de IA con solo unos pocos racks, o expandiendo el entrenamiento de IA a escala con pods optimizados de mayor complejidad y proporciones, Siemon tiene todo lo que necesita para garantizar un rendimiento superior de la red de front-end y back-end, sin importar cuál sea su topología.

Arquitectura de back-end de IA



Red front-end | Interconexiones de switch a switch

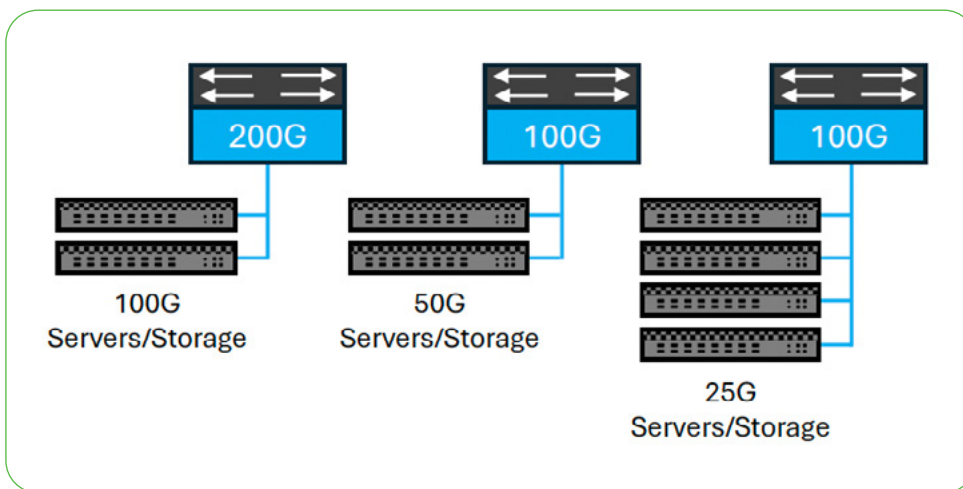


- Las redes front-end tradicionales utilizan Ethernet de 200G y 400G para las conexiones de switch a switch.
- Estas conexiones a menudo se ejecutan de fila a fila, pero pueden usar la topología EoR o MoR en Centros de Datos más pequeños.
- Las configuraciones de conexión (2 x 200G, 2 x 100G, 4 x 100G) pueden maximizar la utilización del puerto y la eficiencia del espacio para las conexiones del switch.

Soluciones de Siemon para Switch-to-Switch de Front-end:

[DAC](#) para uso dentro de un rack (3m típicos) | [AOC](#) para aplicaciones dentro de una fila (hasta 30m típicos)
[BASE-8](#) Cableado MTP, bandejas, paneles y módulos (+10m típicos)

Red front-end | Interconexiones de switch a servidor

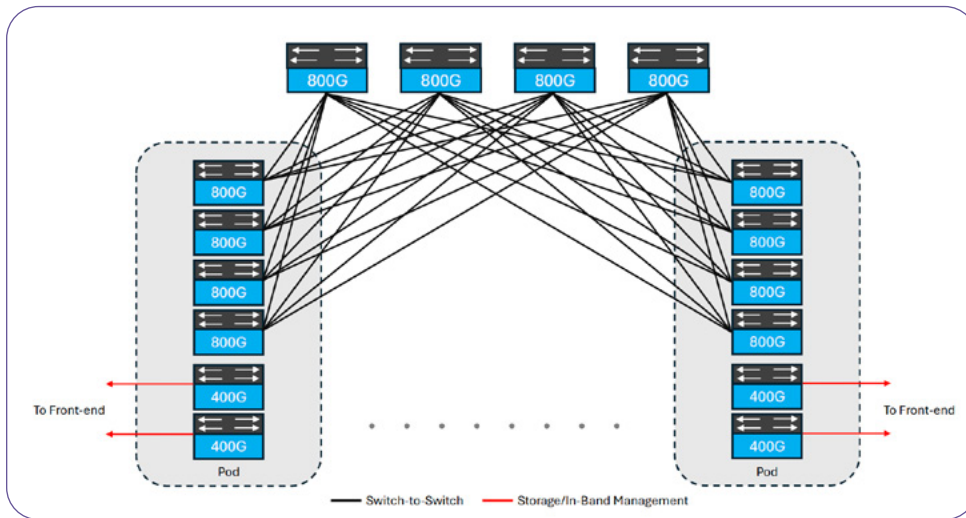


- Los switches de acceso de las redes front-end están migrando a Ethernet de 100G y 200G.
- Esto admite enlaces de servidor de 25, 50 o 100G.
- Las configuraciones de conexión (2 x 50G, 4 x 25G, 2 x 100G) maximizan la utilización del puerto y la eficiencia del espacio.
- Las conexiones de switch front-end a servidor utilizan topologías EoR/MoR o ToR.
- La elección de la topología depende del tamaño y el diseño del Centro de Datos.

Soluciones de Siemon para Front-end switch a servidor:

[DAC](#) para uso dentro de un rack (3m típicos) | [LC y MTP Jumpers](#) dentro de un rack (3m típicos)
[Cat6/6A](#) Cordones de cobre de diámetro para 10G (3m típicos) | [AOC](#) para aplicaciones dentro de una fila/POD (hasta 30m típicos)
[BASE-8](#) Cableado MTP, bandejas, paneles y módulos (+10m típicos)

Back-end Network | Switch-to-Switch Interconnections



- Las redes de back-end HPC utilizan conexiones Ethernet de 800G, con Ethernet de 1,6 T emergente.
- Estas conexiones admiten implementaciones de fila a fila o dentro de fila a través de topologías EoR o MoR.
- La topología ToR se adapta a pequeñas implementaciones de IA de nivel periférico/básico en las que los interruptores de columna y hoja comparten un gabinete.

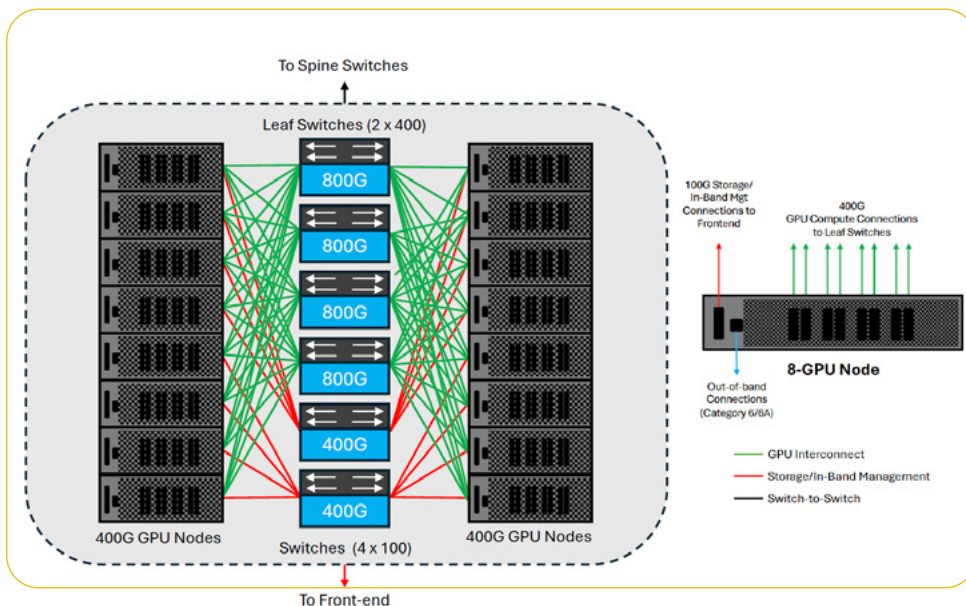
Soluciones de Siemon para back-end de conmutador a conmutador:

DAC para uso dentro de un rack (3m típicos)

ACC/AEC para aplicaciones dentro de una fila/POD (hasta 3-5m típicos) | **AOC** para aplicaciones dentro de una fila/POD (hasta 30m típicos)

BASE-8 Cableado MTP, bandejas, paneles y módulos (+10m típicos)

Red de back-end | Interconexiones de switch a nodo



- Los pods AI cuentan con conexiones de GPU de 400G dentro de los nodos del servidor, cada una vinculada a un switch de leaf distinto para optimizar la comunicación de la GPU.
- Esto se implementa comúnmente con switches de leaf de 800G de alto número de puertos que utilizan configuraciones de conexión de 2 x 400G.
- Cada nodo de servidor también requiere conexiones de red de CPU a front-end (normalmente 100G/200G) para el almacenamiento y la administración.
- Las conexiones de cobre de menor velocidad se utilizan para la gestión fuera de banda.

Soluciones de Siemon para el back-end de switch a nodo:

DAC para uso dentro de un rack (3m típicos) | **ACC/AEC** para uso dentro de un rack/POD (hasta 3-5m típicos)

AOC para aplicaciones dentro de una fila/POD (hasta 30m típicos) | **Cables de cobre Cat6/6A** Skinny Patch para cables punto a punto

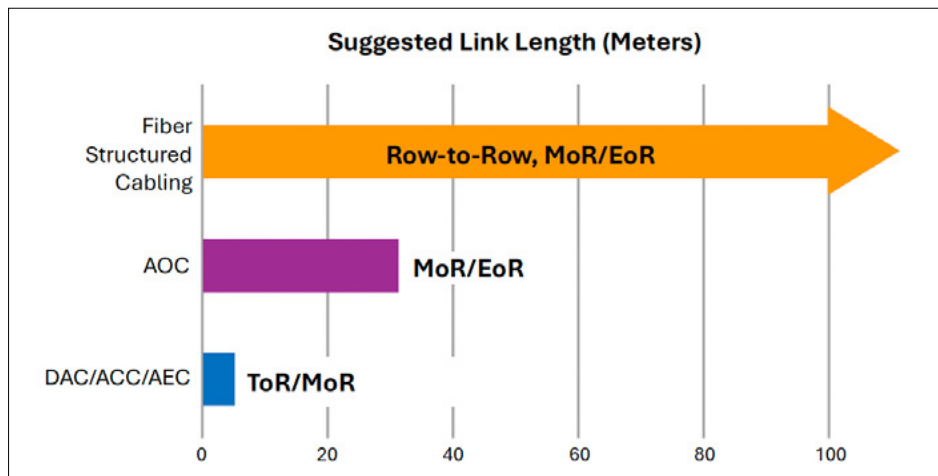
BASE-8 MTP MM/APC de 10 G (3m típicos)

BASE-8 Cableado MTP, bandejas, paneles y módulos (+10m típicos)

Siemon tiene una solución lista para IA que admite todas las topologías

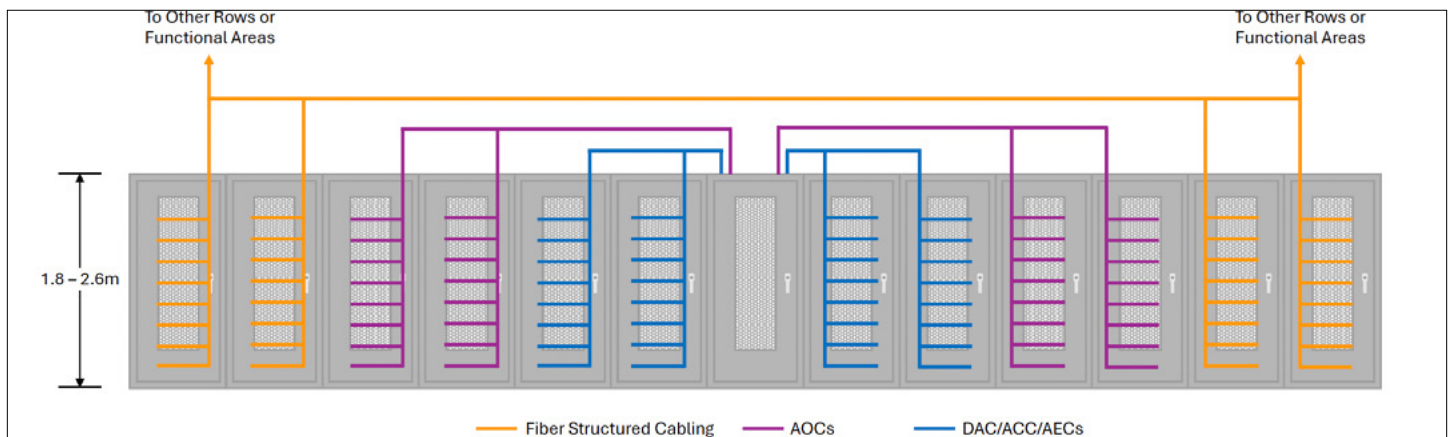
Las topologías de cableado del Centro de Datos varían. Las redes back-end y front-end pueden utilizar cableado punto a punto o sistemas de cableado estructurado a diferentes distancias. Los enlaces se pueden conectar de fila a fila, de gabinete a gabinete (fin de fila/mitad de fila) o dentro del gabinete (Top of the Rack). Los enlaces de switch a switch y de switch a servidor pueden utilizar configuraciones de ruptura, lo que maximiza la utilización y el espacio de los puertos.

Siemon ofrece una gama completa de soluciones de cableado de alto rendimiento para cualquier configuración de distancia, topología y conexión. Esto incluye ensamblajes de cobre de conexión directa (DAC) pasivos y activos, cables ópticos activos (AOC) y sistemas de cableado estructurado de fibra y cobre.



Las consideraciones clave de topología y cableado incluyen:

- **Espacio y diseño** – el diseño del Centro de Datos, el espacio y las distancias entre las áreas influyen en las opciones de topología y cableado.
- **Flexibilidad y escalabilidad** – el cableado estructurado con paneles de conexión permite conexiones cruzadas flexibles, lo que simplifica las operaciones y las actualizaciones. Esto agiliza la integración de nuevos pods en redes de IA en expansión.
- **Alimentación y refrigeración** – la ubicación conjunta de switches y servidores (ToR) aumenta las densidades de alimentación del rack, lo que requiere una infraestructura sólida
- **Latencia** – Las conexiones punto a punto con cables DAC ofrecen una menor latencia y consumo de energía en las topologías ToR y MoR.
- **Análisis de pérdidas** – para tramos de cableado estructurado extendidos (fila a fila y EoR/MoR), el cableado multimodo y monomodo BASE-8 de pérdida ultra baja maximiza el rendimiento.





Cables de cobre de conexión directa

Los DAC, AEC y ACC pasivos y activos de Siemon tienen una construcción de cable flexible y admiten hasta 800G para conexiones de switch a servidor o de switch a switch en longitudes cortas en versiones Ethernet e InfiniBand™. Están disponibles en varios factores de forma, con incrementos de medio metro de 0,5 m a 5 m, y se ofrecen en opciones directas o múltiples de conexión y color.

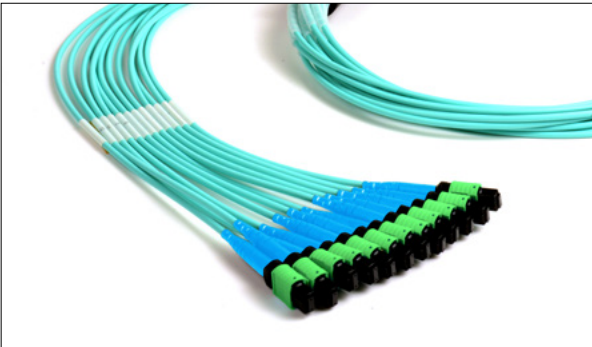
go.siemon.com/DACs



Cables ópticos activos

Los conjuntos de cables AOC de fibra multimodo de Siemon admiten conexiones punto a punto de hasta 30 metros, ideales para topologías EoR/MoR. Hay varias opciones de ruptura disponibles. Su diseño energéticamente eficiente consume menos energía que los conjuntos de transceptores, tienen un diámetro reducido en el cable para mejorar el flujo de aire y reducir los costos de enfriamiento.

go.siemon.com/AOCs



Cableado BASE-8 MTP

La amplia gama de troncales y ensamblajes plug-and-play de Siemon, tanto monomodo como multimodo, incluye troncales MTP a MTP BASE-8 en una variedad de conteos de fibras de hasta 144 fibras, cables de conversión MTP y jumpers MTP.

go.siemon.com/BASE-8



Sistema de cableado de fibra óptica LightStack® de ultra alta densidad

El sistema plug-and-play de fibra de ultra alta densidad LightStack y LightStack 8 de Siemon ofrece una densidad superior de fibra monomodo y multimodo, acceso optimizado a puertos y una gestión de cables eficiente en una plataforma moderna diseñada para centros de datos avanzados. El sistema incluye gabinetes de 1U y 4U con gestión de cables integrada, con capacidad para hasta 144 fibras LC y 864 fibras MTP por unidad de rack. Está disponible en una gama completa de troncales, jumpers y ensamblajes de fibra.

go.siemon.com/LightStackEnclosures



Sistema de cableado de fibra óptica LightVerse® de alta densidad

El sistema de cableado de fibra óptica LightVerse de Siemon ofrece gabinetes de alta densidad, módulos MTP a LC de ultrabaja pérdida, placas adaptadoras MTP y LC, además de jumpers y cables troncales y ensambles MTP y LC. Estos admiten enlaces de fibra monomodo y multimodo en redes front-end y back-end, diseñados para 800G y más allá.

go.siemon.com/LightVerseEnclosures



SIEMON, SU SOCIO EN CADA ETAPA DE SU VIAJE HACIA LA IA



Más información:
www.siemon.com



Encuentre su distribuidor local de
Siemon:
go.siemon.com/distributor



Atención al cliente 24/7:
larena_avila@siemon.com