

# La disrupción de la IA: Desafíos y orientación para el diseño de centros de datos

## Informe técnico n.º 110

Versión 1.1

### Centro de investigación de administración energética

por Victor Avelar

Patrick Donovan

Paul Lin

Wendy Torell

María A. Torres Arango

### Resumen ejecutivo

Desde los grandes clústeres de entrenamiento hasta los pequeños servidores Edge de inferencia, la IA (Inteligencia Artificial) se está convirtiendo en un porcentaje cada vez mayor de las cargas de trabajo de los centros de datos. Esto representa un cambio a densidades de energía de rack más altas. Las empresas nuevas de IA, las empresas establecidas, los proveedores de colocación y los gigantes de Internet ahora deben considerar el impacto de estas densidades en el diseño y la administración de la infraestructura física del centro de datos. En este documento se explican los atributos y tendencias relevantes de las cargas de trabajo de la IA y se describen los desafíos resultantes del centro de datos. Se proporciona orientación para enfrentar estos desafíos para cada categoría de infraestructura física, incluida la administración de energía, enfriamiento, racks y software.

**CALIFICA ESTE INFORME ★★★★★**

## Introducción

En los últimos años, hemos presenciado una extraordinaria aceleración en el crecimiento de la inteligencia artificial (IA), al transformar la forma en que vivimos, trabajamos e interactuamos con la tecnología. La IA generativa (por ejemplo, ChatGPT) es un catalizador para este crecimiento. Los algoritmos predictivos están teniendo un impacto en los sectores de la industria que van desde la atención médica<sup>1</sup> y las finanzas a la manufactura<sup>2</sup>, el transporte<sup>3</sup> y el entretenimiento. Los requisitos de datos asociados con la IA están impulsando nuevas tecnologías de chips y servidores que dan como resultado densidades de alimentación de rack extremas. Al mismo tiempo, existe una demanda masiva de IA. En conjunto, estos presentan nuevos desafíos en el diseño y la operación de centros de datos para admitir esta demanda.

### Proyección de crecimiento de la IA

Estimamos que la IA representa 4.3 GW de demanda de electricidad en la actualidad y proyectamos que crecerá a una CAGR del 26 % al 36 %, lo que da como resultado una demanda total de 13.5 GW a 20 GW para el año 2028. Este crecimiento es de dos a tres veces superior al de la demanda total de electricidad del centro de datos (CAGR) del 11 %. Consulte la **tabla 1** para obtener más detalles. Una idea clave es que las cargas de inferencia<sup>4</sup> aumentarán con el tiempo a medida que más modelos recién entrenados pasen a producción. La demanda real de energía dependerá en gran medida de los factores tecnológicos, incluidas las generaciones sucesivas de servidores, conjuntos de instrucciones más eficientes, mayor rendimiento de los chips y una investigación continua sobre la IA.

**Tabla 1**

*Visión general de las cargas de trabajo de IA en centros de datos.*

Estimación de Schneider Electric	2023	2028
Carga total del centro de datos	54 GW	90 GW
Carga de trabajo de IA	4.3 GW	13.5-20 GW
Carga de trabajo de IA (% del total)	8 %	15-20 %
Carga de trabajo de IA (entrenamiento vs inferencia)	20 % de entrenamiento, 80 % de inferencia	15 % de entrenamiento, 85 % de inferencia
Carga de trabajo de IA (Central vs edge)	95 % central, 5 % perimetral	50 % central, 50 % perimetral

En este documento se explican importantes atributos y tendencias de la IA que crean desafíos para cada categoría de infraestructura física del centro de datos, incluida la administración de alimentación, enfriamiento, racks y software. A continuación, proporcionamos orientación sobre cómo abordar estos desafíos.<sup>5</sup> Por último, proporcionamos una visión prospectiva de lo que está por venir en el diseño de centros de datos. Este documento no trata sobre la aplicación de la IA a los sistemas de infraestructura física. **Si bien los sistemas de infraestructura física de última generación eventualmente aprovecharán más IA, este informe se centra en respaldar las cargas de trabajo de IA con los sistemas existentes que están disponibles hoy mismo.**

<sup>1</sup> Federico Cabitza, et al., [Rams, hounds and white boxes: Investigating human-AI collaboration protocols in medical diagnosis](#), Artificial Intelligence in Medicine, 2023, vol. 138

<sup>2</sup> Jongsuk Lee, et al., [Key Enabling Technologies for Smart Factory in Automotive Industry: Status and Applications](#), International Journal of Precision Engineering and Manufacturing, 2023, vol. 1

<sup>3</sup> Christian Birchler, et al., [Cost-effective simulation-based test selection in self-driving cars software](#), Science of Computer Programming, 2023, vol. 226

<sup>4</sup> Consulte la definición en la sección "Atributos y tendencias de la IA".

<sup>5</sup> Esta guía también se aplica a otras cargas de trabajo de alta densidad, como la computación de alto rendimiento (HPC). Una diferencia importante con las aplicaciones HPC es que tienden a ser instalaciones únicas que pueden emplear soluciones personalizadas de TI, alimentación, enfriamiento o rack. Por el contrario, la demanda masiva de aplicaciones de IA requiere un equipo estándar (TI e infraestructura de apoyo) a escala.

## Atributos y tendencias de la IA

Cuatro atributos y tendencias de la IA subyacen a los desafíos de la infraestructura física:

- Cargas de trabajo de IA
- **Potencia de diseño térmico** (TDP, Thermal design Power) de las GPU (Graphics processing units)
- Latencia de red
- Tamaño del clúster de IA

### Cargas de trabajo de IA

Las cargas de trabajo de IA se dividen en dos categorías generales: entrenamiento e inferencia.

Las cargas de trabajo de **entrenamiento** se utilizan para entrenar modelos de IA como modelos de lenguaje grande (LLM, large language models). El tipo de carga de trabajo de entrenamiento al que nos referimos en este documento es el **entrenamiento distribuido** a gran escala (gran número de máquinas funcionando en paralelo<sup>6</sup>), debido a los desafíos que plantea a los centros de datos en la actualidad. Estas cargas de trabajo requieren grandes cantidades de datos suministrados a servidores especializados con procesadores conocidos como aceleradores. Una unidad de procesamiento de gráficos (GPU) es un ejemplo de acelerador<sup>7</sup>. Los aceleradores son muy eficientes para realizar tareas de procesamiento paralelo como los que se utilizan en el entrenamiento de LLM. Además de los servidores, el entrenamiento también requiere almacenamiento de datos y una red para conectarlos todos. Estos elementos se ensamblan en una matriz de racks conocida como clúster de IA, que básicamente entrena un modelo como una sola computadora. Los aceleradores de un clúster de IA *bien diseñado* funcionan con una utilización cercana al 100 % durante la mayor parte de su duración de entrenamiento, que oscila entre horas y meses. Esto significa que el consumo de electricidad promedio de un clúster de entrenamiento es casi igual a su consumo de electricidad pico (relación pico/promedio  $\approx 1$ ).

Cuanto más grande sea el modelo, más aceleradores serán necesarios. Las densidades de rack en grandes clústeres de IA pueden oscilar entre 30 kW y 100 kW, dependiendo del modelo y la cantidad de GPU. Los clústeres pueden variar desde unos pocos racks hasta cientos de racks y se describen comúnmente por la cantidad de aceleradores utilizados. Por ejemplo, un clúster de **22,000 unidades GPU H100** utiliza aproximadamente 700 racks y requiere aproximadamente 31 MW de electricidad, con una densidad de rack promedio de 44 kW. Hay que tener en cuenta que esta alimentación excluye los requisitos de infraestructura física como el enfriamiento. Por último, las cargas de trabajo de entrenamiento guardan el modelo como "**puestos de control**". Si el clúster falla o pierde alimentación, puede continuar desde donde se quedó.

**Inferencia** significa que el modelo previamente entrenado se pone en producción para predecir la salida de nuevas consultas (entradas). Desde la perspectiva del usuario, existe una compensación entre la precisión de una salida y el tiempo de inferencia (es decir, la latencia). Si soy científico, tal vez esté dispuesto a pagar una prima y esperar más tiempo entre consultas para obtener resultados altamente precisos. Por otro lado, si soy un redactor publicitario en busca de ideas para escribir, quiero un chatbot gratuito con respuestas instantáneas. En resumen, la necesidad empresarial determina el tamaño del modelo de inferencia, pero muy rara vez se utiliza todo el modelo original entrenado. En su lugar, se implementa una versión ligera del modelo para reducir el tiempo de inferencia con una pérdida aceptable de precisión.

Las cargas de trabajo de inferencia tienden a utilizar aceleradores para modelos grandes y también pueden depender en gran medida de las CPU, dependiendo de la aplicación. Es probable que las aplicaciones, como los vehículos autónomos, los motores de recomendaciones y ChatGPT tengan un cúmulo de equipo de TI diferentes, "ajustadas" a sus requisitos.

<sup>6</sup> La gran cantidad de **parámetros** y **tokens** en un modelo requiere que la carga de trabajo de procesamiento se **divida en varias GPU** para reducir el tiempo que tarda en entrenar el modelo.

<sup>7</sup> Otros ejemplos de aceleradores son las unidades de procesamiento de tensores (TPU), los conjuntos de computadoras programables en campo (FPGA) y los circuitos integrados específicos de la aplicación (ASIC).

Dependiendo del tamaño del modelo, los requisitos de hardware por instancia pueden variar desde un dispositivo Edge (por ejemplo, un smartphone) hasta varios racks de servidores. Esto significa que las densidades del rack pueden variar desde unos pocos cientos de vatios a más de 10 kW. A diferencia del entrenamiento, el número de servidores de inferencia aumenta con el número de usuarios/consultas. De hecho, es probable que un modelo popular (por ejemplo, ChatGPT) requiera muchas más veces la cantidad de racks para la inferencia que para el entrenamiento, ya que sus consultas se cuentan ahora por **millones al día**. Por último, las cargas de trabajo de inferencia a menudo son críticas para el negocio, lo que requiere resiliencia (por ejemplo, una redundancia de UPS o geográfica).

## Potencia de diseño térmico (TDP) de las GPU

Si bien el entrenamiento o la inferencia son imposibles sin almacenamiento y red, nos centramos en la GPU porque representa aproximadamente la mitad del consumo de electricidad de un clúster de IA.<sup>8</sup> La potencia de la GPU tiende a ser mayor con cada nueva generación. El consumo de corriente de un chip, medido en vatios, se especifica comúnmente con la **TDP**. Si bien analizamos la GPU específicamente, esta tendencia general de aumento de la TDP también se aplica a otros aceleradores. El aumento de la TDP por generación de GPU es una consecuencia del diseño de la GPU para un mayor número de operaciones, con el fin de entrenar modelos e inferir en menos tiempo y al menor costo. En la **tabla 2** se comparan tres generaciones de GPU Nvidia en términos de TDP y rendimiento<sup>9</sup>.

GPU	TDP (W) <sup>10</sup>	TFLOPS <sup>11</sup> (Entrenamiento)	Rendimiento sobre V100	TOPS <sup>12</sup> (Inferencia)	Rendimiento sobre V100
V100 SXM2 32 GB	300	15.7	1X	62	1X
A100 SXM 80 GB	400	156	9.9X	624	10.1X
H100 SXM 80 GB	700	500	31.8X	2,000	32.3X

**Tabla 2**

*TDP y rendimiento en diferentes generaciones de GPU NVIDIA*

## Latencia de red

Con entrenamiento distribuido, **cada GPU debe tener un puerto de red** para establecer el tejido de la red informática. Por ejemplo, si un servidor de IA tiene ocho GPU, ese servidor requerirá ocho puertos de red de computación. Esta estructura informática permite que todas las GPU de un clúster de IA grande se comuniquen en concierto a altas velocidades (por ejemplo, 800 gigabit/segundo). A medida que aumentan las velocidades de procesamiento de la GPU, también deben hacerlo las velocidades de la red, en un esfuerzo por reducir el tiempo y el costo de los modelos de entrenamiento. Por ejemplo, el uso de GPU que procesan datos de la memoria a 900 GB/s con una estructura computacional de 100 GB/s reduciría la utilización promedio de la GPU porque está esperando en la red para orquestar lo que las GPU harán a continuación. Esto es como comprar un vehículo autónomo de 500 caballos de fuerza con una gama de sensores rápidos que se comunican a través de una red lenta; la velocidad del automóvil estará limitada por la velocidad de la red y, por lo tanto, no utilizará completamente la potencia del motor.

Los cables de red de alta velocidad son costosos. Por ejemplo, las conexiones ópticas InfiniBand son del orden de 10 veces el costo del cobre. Por lo tanto, los científicos de datos, en colaboración con los equipos de TI, intentan especificar clústeres de entrenamiento de IA con cobre, de modo que las distancias de cableado de la red se mantengan dentro de

<sup>8</sup> Con 400 W, la GPU NVIDIA V100 representa el 55 % en este clúster y con 700W, la H100 representa el 49 %

<sup>9</sup> Si bien la GPU es clave para estas mejoras de rendimiento, se han realizado otras mejoras del sistema para aprovechar las GPU mejoradas, como el aumento de la memoria y la comunicación entre GPU.

<sup>10</sup> [V110](#), [A100](#), [H100](#)

<sup>11</sup> TFLOPS, teraFLOPS (billón) de operaciones de punto flotante por segundo: medida del rendimiento de la multiplicación de matrices con precisión de tensor float 32 ([TF32](#)), generalmente utilizada con cargas de trabajo de entrenamiento. [V100](#), [A100](#), [H100](#)

<sup>12</sup> TOPS, teraTOPS (billón) de operaciones por segundo: medida del rendimiento matemático de enteros con una precisión de enteros de 8 bits ([INT8](#)), generalmente utilizada con cargas de trabajo de inferencia. [V100](#), [A100](#), [H100](#)

latencias aceptables. Al aumentar los puertos por rack se reducen las distancias de cableado, pero aumenta la cantidad de GPU por rack y, por lo tanto, la densidad del rack. Con el tiempo, el clúster de racks crece tanto que la latencia obliga a los diseñadores a cambiar a la fibra, lo que aumenta el costo. Hay que tener en cuenta que es más difícil paralelizar las GPU para las cargas de trabajo de inferencia y, por lo tanto, esta relación de densidad de rack generalmente no se aplica a la inferencia.<sup>13</sup>

## Tamaño del clúster de IA

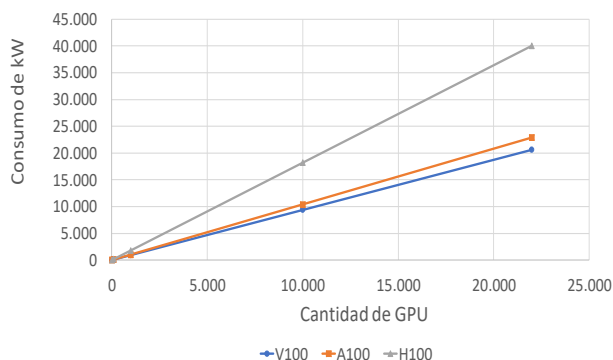
Como se ha explicado anteriormente, el entrenamiento de modelos de gran tamaño puede requerir que miles de GPU actúen en armonía. Dado que la GPU representa aproximadamente la mitad del consumo de electricidad de un clúster, la cantidad de unidades GPU se convierte en un indicador útil para estimar el consumo de electricidad del centro de datos. En la **figura 1** se estima el consumo de electricidad del centro de datos en función de la cantidad de GPU en un clúster de entrenamiento de IA en tres generaciones de GPU (de la **tabla 2**). Para poner estos valores en perspectiva, una central eléctrica de 40,000 kW es capaz de alimentar aproximadamente 31,000 [hogares estadounidenses promedio](#). Hay que tener en cuenta que las tres líneas de tendencia no equivalen a la misma productividad. En otras palabras, mientras que el consumo de electricidad de un centro de datos con unidades GPU H100 supera a uno con unidades GPU V100, las ganancias de productividad del centro de datos H100 superan con creces su prima de consumo de electricidad.

**Figura 1**

*Consumo energético estimado del centro de datos en función de la cantidad de GPU*

*PUE del centro de datos = 1.3*

*Hay que tener en cuenta que las ganancias de productividad no se presentan en este gráfico.*



Los cuatro atributos y tendencias descritos tienen un impacto directo en la densidad de energía del rack. La mayoría de los centros de datos de hoy en día pueden admitir densidades de energía de rack pico de aproximadamente 10 a 20 kW.<sup>14</sup> Sin embargo, la implementación de decenas o cientos de racks de más de 20 kW en un clúster de IA presentará desafíos de infraestructura física a los operadores de centros de datos. Pueden ser específicos de la energía o pueden afectar a dos o más categorías de infraestructura física. Estos desafíos no son insuperables, pero los operadores deben proceder con un entendimiento completo de los requisitos, no solo con respecto a la TI, sino también a la infraestructura física, especialmente para las instalaciones de centros de datos existentes. Cuanto más antigua sea la instalación, más difícil será admitir las cargas de trabajo de entrenamiento de IA. En las secciones principales a continuación se explican estos desafíos con más detalle para cada categoría de infraestructura física y proporcionan orientación para superar estos desafíos. Hay que tener en cuenta que algunos de los enfoques de diseño recomendados solo se aplican a las nuevas construcciones de centros de datos, mientras que otros son relevantes tanto para los edificios nuevos como para los de campo de navegación (modernización).

## Energía

Las cargas de trabajo de IA presentan seis retos clave que afectan al tren de potencia, incluidos los tableros de distribución, la distribución y las unidades de distribución de potencia en rack (rPDU).

- La distribución de 120/208 V no es práctica de implementar
- Los pequeños tamaños de bloques de distribución de energía desperdician espacio de TI

<sup>13</sup> [Aceleración de la inferencia de aprendizaje profundo con el paralelismo de hardware y software](#) Abril de 2020

<sup>14</sup> Uptime Institute, [Rack Density is Rising](#), 12/2022

- Las PDU para rack estándar de 60/63 A no son prácticas de implementar
- El mayor riesgo de peligro de arco eléctrico (arc-flash) complica las prácticas de trabajo
- La falta de diversidad de carga aumenta el riesgo de disparo del interruptor automático aguas arriba
- Las altas temperaturas del rack aumentan el riesgo de fallas y peligros

## La distribución de 120/208 V no es práctica de implementar

120/208 V, una tensión utilizada históricamente en centros de datos de América del Norte, cumplió su propósito cuando las densidades eran relativamente bajas (del orden de dos a tres kilovatios por rack) y los servidores recibieron cables de alimentación de 120 V. En la actualidad, con cargas de alta densidad como los clústeres de IA, esta tensión es demasiado baja. Aunque todavía es posible alimentar estas cargas a 120/208 V, esto plantea problemas, que se derivan de la siguiente relación: la potencia es igual a voltios por amperios ( $P = V \times A$ ). Como muestra esta ecuación, cuanto menor sea la tensión, más corriente necesitará para la misma potencia. Por consiguiente, el conductor debe ser más grande para proporcionar de manera segura una mayor corriente.

Ahora, considera un rack de entrenamiento de IA de (8) servidores acelerados por GPU HPE Cray XD670, con una densidad total de rack de 80 kW. Con 120/208 V, se necesitarían cinco circuitos de 60 amperios para alimentar el rack (cada circuito es igual a  $120 \text{ V} \times 3 \text{ fases} \times 60 \text{ A} \times 80 \% \text{ de reducción de potencia} = 17.280 \text{ W} = 17.3 \text{ kW}$ ) con redundancia 1N. Si se requiriera 2N (aunque es poco común para las cargas de entrenamiento de IA), este número se duplicaría a diez. Con 5 a 10 circuitos por rack, hay que imaginar el caos de los cables de alimentación distribuidos en un clúster de IA de 100 racks. El resultado es probablemente una instalación improvisada de cables de alimentación colgando por encima o cerca del rack, que podría conducir a problemas, incluidos errores humanos y restricciones de flujo de aire. Esto no resulta práctico. Además, la instalación y administración de una cantidad excesiva de circuitos tiene implicaciones económicas.

**GUÍA:** Debido a que duplicar la tensión significa duplicar la energía, los centros de datos existentes con distribución de 120/208 V deben modernizar su distribución a 240/415 V. Los nuevos centros de datos ya deben diseñarse teniendo en cuenta 240/415 V. Consulte el informe técnico n.º 128, [Distribución de energía de CA de alta eficiencia para centros de datos](#), para obtener más información sobre este tema. Esto lleva al siguiente desafío, que está relacionado con las limitaciones sobre cómo distribuir la alimentación de 240/415 V.

Hay que tener en cuenta que gran parte del mundo no tiene este mismo desafío, ya que muchos países distribuyen energía a una tensión más alta de 230/400 V, lo que es adecuado para lograr las demandas de energía en racks de IA.

## Los pequeños tamaños de bloques de distribución de energía desperdician espacio de TI

Existen tres tipos principales de distribución de energía en el centro de datos: unidades de distribución de energía (PDU) basadas en transformadores, tableros de energía remotos (RPP) y electroducto (también llamado ducto barra o busway). El tamaño del bloque de distribución representa la capacidad (kW) de cada solución de distribución. Incluso con una mayor tensión de distribución de 240/415 V (países con IEC de 230 V), los tamaños de los bloques de distribución tradicionales son demasiado pequeños para admitir las capacidades de clúster de IA de hoy en día. Hace diez años, un bloque de distribución de 300 kW (833 A a 120/208 V) podría admitir 100 racks (cinco filas de 20 racks a una densidad de rack promedio de 3 kW/rack). Hoy en día, ese mismo bloque ni siquiera podía soportar la configuración mínima de un [NVIDIA DGX SuperPOD](#) (una sola fila de 10 racks de 358 kW a 36 kW/rack). El uso de múltiples bloques de distribución para una sola fila de racks no es práctico por varias razones. Por ejemplo, como mínimo se duplica el espacio de las PDU y los RPP. Varios

bloques también aumentan el costo en comparación con un solo bloque de mayor capacidad.

**GUÍA:** Los tamaños de los bloques de distribución deben aumentar para satisfacer las demandas de los clústeres de alta densidad. Se recomienda elegir un tamaño de bloque de distribución lo suficientemente alto como para alojar, como mínimo, una fila completa del clúster. Un tamaño de bloque de 800 amperios es un tamaño de capacidad estándar disponible actualmente para los tres tipos de distribución en un arreglo de 240/415 V. Esto proporciona 576 kW (461 kW disminuidos).

### Las PDU para rack estándar de 60/63 A no son prácticas de implementar

Incluso a una tensión más alta, sigue siendo un desafío proporcionar capacidad suficiente con una rPDU estándar. La mayoría de los responsables de la toma de decisiones prefieren las rPDU estándar porque tienen plazos de entrega más cortos, son fáciles de conseguir, rentables y varios proveedores las venden con configuraciones similares.

En la actualidad, la rPDU estándar de mayor capacidad está clasificada en 60 A (NEMA) / 63 A (IEC). En la **tabla 3** se ilustra la capacidad utilizable de las rPDU con diversas clasificaciones de corriente y tensiones. En base a esto, vemos que el valor nominal de 60 y 63 A limita la capacidad de una única rPDU a 34.6 kW y 43.5 kW respectivamente. Esto lleva al dilema de cómo manejar mejor las densidades de rack que esto.

**Tabla 3**

*Densidad de energía trifásica útil por unidad rPDU basada en el valor nominal de potencia y la tensión del interruptor automático (línea a neutro)*

*Parte superior: NEMA (por ejemplo, Norteamérica)  
Parte inferior: IEC (por ejemplo, Europa)*

	Estándar		Personalizado			
NEMA	40 A	60 A	100 A	125 A	150 A	175 A
120/208 V	11.5 kW	17.3 kW	28.8 kW	36.0 kW	43.2 kW	50.4 kW
240/415 V	23.0 kW	34.6 kW	57.6 kW	72.0 kW	86.4 kW	100.8 kW

	Estándar		Personalizado			
IEC	32 A	63 A	100 A	125 A	150 A	160 A
230/400 V	22.1 kW	43.5 kW	69.0 kW	86.3 kW	103.5 kW	110.4 kW

Hay que tener en cuenta que estos valores se reducen al 80 % en función de los requisitos de códigos normales.

**GUÍA:** Para densidades de rack mayores a 34.6 kW (NEMA) y 43.5 kW (IEC), existen dos enfoques.

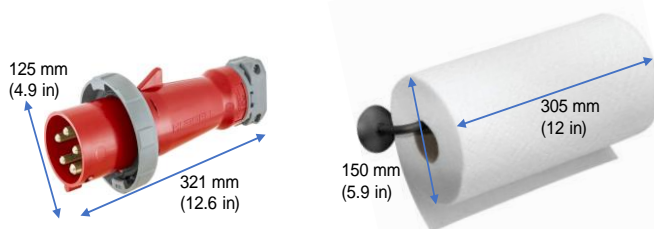
1. Múltiples rPDU estándar listas para usar
2. rPDU personalizada superior a 60 A y 63 A

En la actualidad, la mayoría de las rPDU en U de cero tienen aproximadamente 2 metros (80 in) de altura. Con estas ofertas estándar, es probable que se instalen, como máximo, 4 unidades rPDU en un solo rack enfriado por aire (por ejemplo, las unidades rPDU de 4 x 60/63 A son de 138 kW/174 kW). O si se requiere un colector de enfriamiento líquido, entonces 2 rPDU en un solo rack (por ejemplo, 2 rPDU de 60/63 A son de 69 kW/87 kW). Estas rPDU se pueden combinar para aumentar la capacidad o aplicarse para fines de redundancia (por ejemplo, 2N).

Si existe una restricción de espacio debido a la cantidad de rPDU, se recomienda las rPDU personalizadas. Por ejemplo, como se muestra en la **tabla 3**, es posible alimentar un rack de 100 kW con una unidad rPDU de 175 A en Norteamérica o de 150 A en Europa. Las rPDU personalizadas pueden venir con un conector de pines y mangas o estar cableadas y brindan la flexibilidad de la cantidad y el tipo de receptáculos. Con valores nominales de corriente más elevados, los conectores de clavija y casquillo requieren más trabajo para instalarlos y pasarlos por un rack debido a su tamaño físico (ver la **figura 2**). Hay que tener en cuenta que, con valores nominales de corriente mayores de 60 A, la instalación y el funcionamiento pueden requerir de un electricista.

## Figura 2

Conector de clavija y casquillo de 240/415 V 125 A comparado con el tamaño de un rollo de toalla de papel. Acoplar un par de conectores tan grandes es un desafío para una sola persona.



## El mayor riesgo de peligro de arco eléctrico (arc-flash) complica las prácticas de trabajo

De acuerdo con el informe técnico n.º 194, [Consideraciones sobre el arco eléctrico para el espacio de TI del centro de datos](#), el término “arco eléctrico” describe lo que ocurre cuando la corriente de cortocircuito eléctrico fluye a través del aire. En un destello por arqueo, la corriente literalmente viaja por el aire de un punto a otro, liberando una gran cantidad de energía, conocida como energía incidente<sup>15</sup>, en menos de un segundo. Esta energía se libera en forma de calor, sonido, luz y presión explosiva, todos los cuales pueden causar lesiones. Algunas lesiones específicas pueden incluir quemaduras, ceguera, descarga eléctrica, pérdida de la audición y fracturas.

Una consecuencia del aumento de las corrientes nominales de la rPDU es que tienen diámetros de conductor más grandes que permiten más corriente de falla a través de la rPDU. Si la corriente de falla disponible en la rPDU da como resultado una energía incidente de 1.2 calorías/cm<sup>2</sup> o más, no se permite la presencia de trabajadores en esa área sin la capacitación adecuada y el equipo de protección personal (EPP).<sup>16</sup> El riesgo aumenta a medida que aumenta el valor nominal de amperaje de la rPDU. La seguridad del personal del centro de datos es un desafío que se debe atender.

**GUÍA:** Con tantas variables involucradas, se recomienda comenzar con una evaluación de riesgo de destello por arqueo para analizar la corriente de falla disponible, ya que esto permite determinar las mejores soluciones para un sitio específico. Es importante que este estudio se realice desde el equipo de media tensión hasta el nivel del rack. Algunos ejemplos de soluciones son:

- Especificación de transformadores con mayor impedancia aguas arriba
- Uso de reactores de línea (es decir, inductores) para impedir el flujo de corriente de cortocircuito
- Uso de [bloques limitadores de corriente](#)
- Uso de [interruptores automáticos limitadores de corriente](#)

Consulte el informe técnico [Mitigación de arco eléctrico](#), y el informe técnico n.º 253, [Beneficios de limitar la corriente de cortocircuito de las MV en centros de datos grandes](#), para obtener más detalles sobre cómo enfrentar los peligros de destello por arqueo.

## La falta de variación de carga aumenta el riesgo de disparo del interruptor automático aguas arriba

<sup>15</sup> De acuerdo con la norma NFPA 70E (2015), la energía incidente es “La cantidad de energía térmica impresionada en una superficie, a una cierta distancia de la fuente, generada durante un evento de destello por arqueo”.

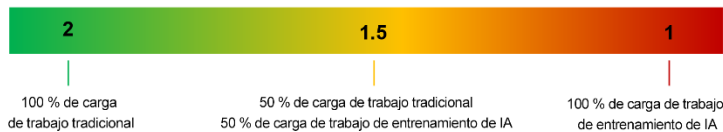
<sup>16</sup> Para obtener más información, consulte los informes técnicos n.º 13, [Mitigación de los riesgos eléctricos al intercambiar equipos energizados](#) y 194, [Consideraciones sobre el destello por arqueo para el espacio de TI del centro de datos](#)”.



El consumo de energía de las diversas cargas de trabajo del centro de datos suele alcanzar el pico en momentos aleatorios. Estadísticamente hablando, hay una probabilidad muy baja de que todos estos picos ocurran al mismo tiempo. Por lo tanto, si sumara el pico de todas las cargas de trabajo individuales y dividiera este valor por el consumo de energía promedio total, encontraría una relación pico/promedio de 1.5 a 2 o más para un centro de datos grande típico. Esto es lo que permite a los diseñadores “sobre - suscribir” los sistemas de energía y enfriamiento. Sin embargo, como se analizó en la sección “Atributos y tendencias de la IA”, las cargas de entrenamiento de IA carecen de diversidad. Estas cargas de trabajo pueden ejecutarse durante horas, días o incluso semanas a una potencia máxima. El resultado es una mayor probabilidad de disparar un interruptor automático grande aguas arriba. Esto es como lo que sucede cuando muchas cargas grandes se ejecutan simultáneamente en un hogar y el interruptor automático del tablero principal se abre. En la **figura 3** se ilustra el espectro típico de la relación pico/promedio (también conocido como factor de diversidad) a medida que las cargas en un centro de datos pasan a ser el 100 % de las cargas de IA.

### Figura 3

Espectro de relaciones típicas de pico/promedio desde 100 % de cargas mixtas tradicionales hasta 100 % de cargas de trabajo de entrenamiento de inteligencia artificial



**GUÍA:** En el caso de una nueva sala de un centro de datos con más de 60-70 % de cargas de trabajo de entrenamiento de IA, se recomienda dimensionar el interruptor automático principal en función de la suma de los interruptores de alimentación individuales aguas abajo. En otras palabras, supongamos una relación pico/ promedio de 1, donde el consumo de energía promedio es igual al consumo de energía pico. No se aconseja la práctica de la sobre - suscripción y ni depender de la diversidad.

Para los centros de datos existentes, calculemos la carga de IA total que puede admitir el interruptor automático aguas arriba. Por ejemplo, si hay un interruptor automático principal de 1,000 A aguas arriba de los clústeres de carga de trabajo de IA, hay que asegurarse de que las cargas de IA no sumen más de 1,000 A.

### Las altas temperaturas del rack aumentan el riesgo de fallas y peligros

Entre el aumento de las densidades y el enfoque en la eficiencia operativa, los entornos de TI se están calentando. Las altas temperaturas de funcionamiento mejoran la eficiencia del sistema de enfriamiento, pero también causan mayor tensión en los componentes. Cuando los componentes están expuestos a temperaturas para las que no están clasificados, el resultado puede ser el siguiente:

- **Fallas de componentes prematuras:** aunque el primer día los sistemas operan como está previsto, la vida útil de los componentes puede reducirse significativamente cuando se exponen a condiciones fuera de la gama especificada.
- **Peligros de seguridad:** el uso de cables no clasificados para el rango de funcionamiento podría conducir a peligros de seguridad, como la fusión de los cables.

IEC 60320 es el estándar internacional reconocido utilizado por la mayoría del mundo para la conexión de cables de suministro de energía. Existen conectores IEC específicos para temperaturas nominales más elevadas. En la **tabla 4** se comparan los conectores C19/C20 estándar con los conectores C21/C22 para alta temperatura.

Tabla 4

Comparación de conectores estándar [IEC 60320](#) y de alta temperatura para 250 V y 16/20 A

	Hembra	Macho	Límite	Notas
Estándar	 C19	 C20	65 °C	C20 se utiliza comúnmente como un cable puente, que proporciona energía desde una PDU de rack a dispositivos de TI de alta potencia.

	Hembra	Macho	Límite	Notas
Alto temperatura	 C21	 C22	155 °C	C21 se acopla con conectores C22 o C20 y se utiliza cuando las temperaturas superan el valor nominal C19.

**GUÍA:** Se recomienda analizar todas las cargas dentro del clúster de IA para garantizar que se utilicen los conectores y receptáculos adecuados. Los conectores C21/C22 se están volviendo más comunes, con cargas de equipo de cómputo de mayor densidad, como los servidores de IA. Los servidores de IA a menudo se configuran con estos cables/receptáculos de alta temperatura, pero es posible que otros dispositivos del rack no lo hagan, como el interruptor de la parte superior del rack. Es importante comprender el entorno en el que operará su equipo y asegurarse de que todos los dispositivos tengan la clasificación correspondiente, incluida la PDU de rack y todos sus subcomponentes.

Al especificar las PDU de rack, es importante no solo observar la tensión, el amperaje y la cantidad de tomacorrientes, sino también su valor nominal de temperatura. Para este tipo de aplicación, existen en el mercado unidades rPDU con alta temperatura nominal. Aunque por lo general tienen un costo adicional, ese costo agregado generalmente supera el costo de las fallas latentes que están a la espera de ocurrir. Debe tenerse en cuenta que también se recomienda colocar sensores de temperatura en la parte posterior del rack (monitoreados por DCIM) para validar que las condiciones de funcionamiento sean las esperadas.

## Enfriamiento

La densificación de los clústeres de servidores de entrenamiento de IA está forzando una evolución del enfriamiento por aire al enfriamiento por líquido para enfrentar sus crecientes TDP. Si bien los clústeres menos densos y los servidores de inferencia seguirán usando enfriamiento de centros de datos más convencional, vemos los siguientes seis desafíos clave del enfriamiento que los operadores de centros de datos necesitan atender:

- El enfriamiento por aire no es adecuado para clústeres de IA superiores a 20 kW/rack
- La falta de diseños estandarizados y las limitaciones del sitio complican las modernizaciones de enfriamiento líquido
- Las TDP futuras desconocidas aumentan el riesgo de obsolescencia del diseño de enfriamiento
- La inexperiencia complica la instalación, operación y mantenimiento
- El enfriamiento líquido aumenta el riesgo de fugas dentro de los racks de TI
- Existen opciones limitadas de líquidos para operar el enfriamiento líquido de manera sostenible

### El enfriamiento por aire no es adecuado para clústeres de IA superiores a 20 kW/rack

El enfriamiento líquido para TI ha existido durante más de medio siglo para la computación especializada de alto rendimiento. El enfriamiento por aire ha sido la opción principal y puede soportar densidades de energía de rack promedio de aproximadamente 20 kW cuando se diseña correctamente con contención de pasillos calientes. Con un único servidor de IA 8-10U que consume **12 kW**, es fácil exceder este umbral de 20 kW. A este desafío se añade que los servidores de grandes clústeres de IA no se pueden distribuir (para reducir la densidad del rack) debido a las limitaciones de latencia. Cada vez hay más versiones enfriadas por líquido de los servidores de entrenamiento de IA, y algunas son exclusivamente de enfriamiento líquido, impulsadas por el aumento de la TDP.

**GUÍA:** Los clústeres de IA más pequeños y los racks de servidores de inferencia que están configurados a 20 kW por rack o menos pueden enfriarse con aire. Para estos racks, las buenas prácticas de administración de flujo de aire (por ejemplo, paneles ciegos (blanking panels), [contención de pasillos](#)) para garantizar un enfriamiento más eficaz y eficiente. Si un sistema de enfriamiento por aire sigue teniendo restricciones, distribuir los servidores de IA en múltiples racks es una estrategia para reducir la densidad del rack. Por ejemplo, si un clúster tiene 20 racks a 20 kW/rack, la distribución de los servidores a 40 racks reduciría la densidad del rack a 10 kW/rack. Hay que tener en cuenta que es posible que no se puedan distribuir los racks si las mayores distancias de cableado de la red degradan el rendimiento del clúster de IA.

Cuando las densidades del rack de IA superan los 20 kW, se debe prestar especial atención a los servidores enfriados con líquido. Existen varias tecnologías y arquitecturas de enfriamiento líquido. Directo al chip (DTC), a veces llamada placa conductiva o fría, y la inmersión, son las dos categorías principales. En comparación con la inmersión, la opción directo al chip es actualmente la preferida, ya que es más compatible con el enfriamiento por aire existente y también es más fácil para aplicaciones de modernización. Si se les da la opción, los operadores de centros de datos deben seleccionar servidores enfriados por líquido para mejorar el rendimiento y reducir el costo de energía, lo que puede compensar la prima de inversión. Por ejemplo, el servidor acelerado por GPU HPE Cray XD670 consume 10 kW cuando se enfría con aire, frente a 7.5 kW cuando se enfría con líquido, debido a la reducción de los requisitos de alimentación del ventilador y a las menores corrientes de fuga en el silicio. Para obtener más información sobre enfriamiento líquido, consulte el informe técnico n.º 279, [Cinco razones para adoptar el enfriamiento líquido](#), y el informe técnico n.º 265, [Tecnologías de enfriamiento líquido para centros de datos y aplicaciones perimetrales](#).

Hay que tener en cuenta que los líquidos tienen una capacidad mucho mayor para capturar calor por volumen de unidad, lo que permite que las tecnologías de enfriamiento líquido eliminen calor de manera más eficiente que el enfriamiento por aire. Sin embargo, si se detiene el flujo de líquido, la temperatura del chip aumentará mucho más rápido que con el aire, lo que provocará un apagado más rápido. Colocar las bombas en una UPS ayuda a resolver este problema.

## La falta de diseños estandarizados y las limitaciones del sitio complican las modernizaciones de enfriamiento líquido

En comparación con los sistemas de agua fría tradicionales, los servidores enfriados con líquido directo al chip tienen requisitos más estrictos en cuanto a temperatura, flujo y química del agua. Esto significa que los operadores no pueden pasar agua directamente desde un sistema de enfriamiento a través de la placa fría de un chip.<sup>17</sup> Si bien la calidad del agua es sin duda parte del desafío de modernizar un centro de datos a enfriamiento líquido, el mayor problema es la falta de diseños estandarizados para las cargas de IA a esta escala (es decir, cientos de racks). Hay que considerar que existen varias opciones de montaje y ubicaciones para las unidades de distribución de refrigerante (CDU)<sup>18</sup>. Puede instalarse en el piso en el perímetro de la sala, al final de la fila o en el rack en cada rack de servidores. Existen varios métodos para distribuir la tubería a los racks, muchas ubicaciones para el equipo del sistema de enfriamiento, varios métodos para controlar las temperaturas, etc. Para ayudar a visualizar los componentes de un sistema enfriado por líquido, en la **figura 3** se ilustran diferentes circuitos de agua y CDU.

<sup>17</sup> El flujo de agua sin tratar a través de la placa fría de un servidor puede causar corrosión, crecimiento biológico y acumulación de suciedad. Todos estos factores comprometen la transferencia de calor de las GPU, lo que finalmente lleva a que estas se aceleren o se apaguen para evitar daños.

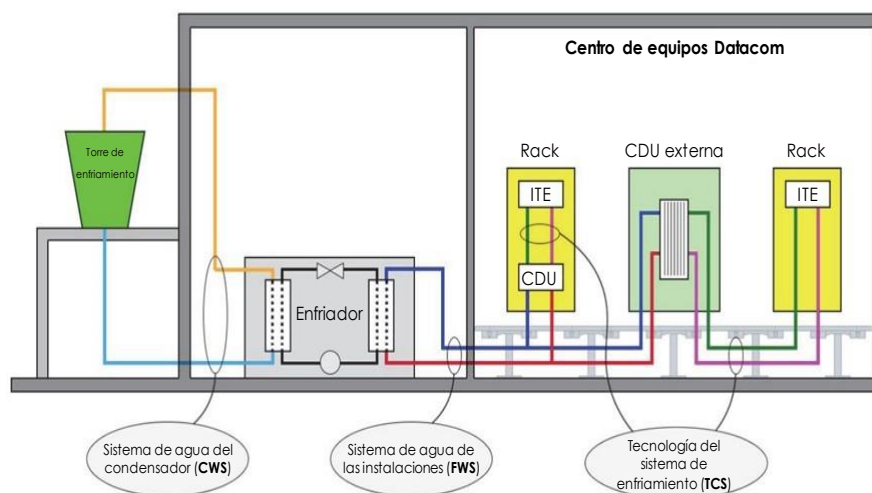
<sup>18</sup> Una CDU aísla físicamente el circuito de agua fría del circuito de agua "limpia" que suministra los servidores.

La modernización para el enfriamiento líquido también es disruptiva para un centro de datos de producción y puede toparse con restricciones físicas como espacio limitado en el piso y altura insuficiente del piso elevado para hacer funcionar la tubería de agua. Incluso si el 100 % de los servidores se enfrían con líquido directo al chip, todavía hay una necesidad de enfriamiento por aire adicional para enfriar otros equipos como interruptores de red y conducción de calor de los servidores enfriados por líquido. En resumen, la modernización es un desafío debido al alto número de permutaciones de diseño combinadas con análisis limitados y pocas implementaciones de enfriamiento líquido a gran escala de las cuales aprender. Hay que tener en cuenta que algunos centros de datos no tienen un sistema de agua fría, lo que hace que la modernización sea aún más desafiante.

### Figura 3

Enfriamiento líquido mediante CDU en un centro de datos

Fuente: ASHRAE, [Water-Cooled Servers: Common Designs, Components, and Processes](#), página 10



**GUÍA:** Se recomienda a los operadores de centros de datos que realicen una evaluación de diseño de las cargas enfriadas por líquido propuestas y de las condiciones existentes de la instalación antes de implementar el enfriamiento líquido. La revisión por expertos es esencial para evaluar posibles diseños y evitar las repercusiones en los costos de limitaciones imprevistas en la construcción. Por ejemplo, las tuberías pueden obstruir el flujo de aire debajo del piso elevado o interferir con la alimentación o las bandejas de cables. Para obtener más información, consulte el informe técnico n.º 133, [Prácticas de diseño de centros de datos para integrar cargas de trabajo de IA enfriadas por líquido](#).

## Las TDP futuras desconocidas aumentan el riesgo de obsolescencia del diseño de enfriamiento

Las tecnologías de IA están evolucionando a un ritmo tan rápido que es probable que las GPU de próxima generación tengan mayores TDP y mayores requerimientos de enfriamiento. Por ejemplo, un servidor actual con ocho GPU puede tener dieciséis en su próxima generación. Como resultado, la distribución de enfriamiento de un centro de datos diseñada para las cargas actuales puede volverse insuficiente para soportar las cargas del futuro.

**GUÍA:** Se recomienda diseñar el sistema de enfriamiento para adaptarse al enfriamiento por aire y líquido, escalar según sea necesario y admitir diferentes generaciones de aceleradores. Por ejemplo, el uso actual de enfriadores de alta temperatura para enfriamiento por aire puede cambiarse fácilmente a enfriamiento líquido a mayor temperatura. Otra práctica recomendada es diseñar el sistema de tuberías de agua fría con derivaciones para futuras CDU. Esto permitirá una transición a cargas 100 % enfriadas por líquido directo al chip, combinadas con intercambiadores de calor de puerta trasera para un enfriamiento por aire complementario.

## La inexperiencia complica la instalación, operación y mantenimiento

Los operadores de centros de datos están bastante familiarizados con los sistemas enfriados por aire, ya que se han utilizado durante décadas el enfriamiento líquido es nuevo para la mayoría de los operadores. El enfriamiento líquido utiliza componentes como placas frías, colectores, válvulas de acoplamiento ciego, etc. Estos componentes vienen acompañados de procedimientos adicionales de instalación, funcionamiento y mantenimiento desconocidos para estos operadores. Por ejemplo, los diminutos canales de agua de los servidores de placa fría directa al chip son más susceptibles a las incrustaciones, lo que significa que los operadores pueden tener que aprender nuevos procedimientos de operación y mantenimiento para controlar la química del agua. Otro ejemplo es pasar el agua por los servidores, lo que presenta el riesgo de fugas.

**GUÍA:** El diseño del enfriamiento líquido desempeña un papel fundamental en la minimización del trabajo de instalación, operación y mantenimiento. Se recomienda que los operadores de centros de datos, que no están familiarizados con la compatibilidad con servidores enfriados por líquido, busquen expertos que proporcionen una revisión exhaustiva de su diseño y emitan procedimientos operativos estándar (SOP) y métodos de procedimientos (MOP) detallados para las operaciones diarias. Esto minimizará las fallas y los errores humanos, especialmente los relacionados con fugas.

## El enfriamiento líquido aumenta el riesgo de fugas dentro de los racks de TI

Las tecnologías directo al chip utilizan agua (por ejemplo, agua desionizada, soluciones basadas en alcohol) en placas frías dentro de un servidor. Las fugas de agua son un problema de seguridad y confiabilidad y deben considerarse en la fase de diseño y adquisición.

**GUÍA:** Se recomienda trabajar con proveedores acreditados para garantizar que sus sistemas pasen por pruebas de presión rigurosas para minimizar el riesgo de fugas. Además, la detección de fugas a nivel de servidor y rack puede ayudar a detectar fugas antes de que se vuelvan más graves. En lugar de los sistemas de bombeo tradicionales de la CDU, hay que considerar las CDU con sistemas innovadores de prevención de fugas (LPS). Los LPS mantienen el circuito de agua a un ligero vacío (presión negativa) para eliminar el riesgo de fugas dentro de los equipos de TI. En el enfriamiento líquido por inmersión se utilizan fluidos dieléctricos que también eliminan el riesgo de fugas de agua dentro de los servidores. Estas pueden ser una opción de tu proveedor de servidores de IA o de integración. Por último, se deben desarrollar procedimientos de operación de emergencia (EOP) para resolver las fugas si se presentan.

## Existen opciones limitadas de líquidos para operar el enfriamiento líquido de manera sostenible

En comparación con la TI tradicional enfriada por aire, el enfriamiento líquido ofrece algunos beneficios de sostenibilidad ambiental en el sentido de que reduce tanto el consumo de energía como el uso de agua. Esto se debe a la mayor eficiencia energética tanto de los servidores de TI como de los sistemas de enfriamiento, ya que se eliminan la mayoría o la totalidad de los ventiladores del servidor, y las mayores temperaturas del agua permiten aumentar las horas del economizador.<sup>19</sup> Sin embargo, algunos sistemas enfriados por líquido utilizan productos químicos diseñados que son perjudiciales para el medio ambiente. Por ejemplo, los fluidos de fluorocarbono se utilizan ampliamente como fluidos dieléctricos en tecnologías de

<sup>19</sup> La economización se produce cuando la temperatura exterior es más baja que la temperatura del agua. Las temperaturas del agua de retorno de los servidores con enfriamiento directo al chip son mucho más altas que las temperaturas tradicionales de retorno de agua fría. Con estas temperaturas más altas, hay más horas al año para enfriar el agua.

enfriamiento líquido de inmersión<sup>20</sup> debido a su rendimiento de transferencia de calor. Desafortunadamente, algunos fluorocarbonos tienen potenciales de calentamiento global (**GWP**) del orden de 8,000. A modo de comparación, el refrigerante HFC-143a, común en los refrigeradores, tiene un GWP de 1,430. Además, las presiones de la sociedad han llevado a los fabricantes a eliminar las PFAS (sustancias per y polifluoroalquilo) de productos como los refrigerantes (para mitigar el impacto ambiental) y a pasar a refrigerantes con un GWP más bajo. La sostenibilidad se ha convertido en una prioridad para la mayoría de los operadores de centros de datos, ya que les quedan menos opciones.

**GUÍA:** Se recomienda evitar los fluidos de fluorocarbono. En el pasado, estos se utilizaban en sistemas de enfriamiento por inmersión y directo al chip. Hoy en día, la tecnología directo al chip utiliza agua y, por lo tanto, no es un problema. Si se implementa enfriamiento líquido de inmersión, se recomienda utilizar fluidos dieléctricos basados en aceite que tengan cero GWP (a diferencia de los fluidos de ingeniería bifásicos). Sin embargo, debido a que los fluidos dieléctricos basados en aceite no son tan efectivos para transferir calor como los de agua directo al chip, la tecnología directo al chip se ha convertido en la arquitectura de enfriamiento líquido preferida en la actualidad. Hay que tener en cuenta que es probable que los proveedores desarrollen dieléctricos sostenibles como alternativa a los fluidos de fluorocarbono. Esto mejoraría significativamente la eficiencia de la eliminación de calor del enfriamiento líquido de inmersión y tal vez precipitaría un cambio en la arquitectura de enfriamiento. Consulte el informe técnico n.º 291, [Comparación de fluidos dieléctricos para enfriamiento líquido envolvente de equipos de cómputo](#) para obtener más información.

## Racks

Algunos de los problemas de alimentación y enfriamiento descritos en las secciones anteriores también afectan al rack de TI (es decir, al armario o gabinete de TI). Veamos los siguientes cuatro desafíos del sistema de racks impulsados por las cargas de trabajo de IA:

- Los racks de ancho estándar carecen de espacio para los equipos de energía y enfriamiento necesarios
- Los racks de profundidad estándar carecen de espacio para el cableado y los servidores de IA profundos
- Los racks de altura estándar carecen de espacio para la cantidad requerida de servidores
- Los racks estándar carecen de la capacidad suficiente para soportar el peso de los equipos de IA

### Los racks de ancho estándar carecen de espacio para los equipos de energía y enfriamiento necesarios

Debido a que los servidores de IA son cada vez más profundos, hay menos espacio en la parte posterior del rack para montar unidades PDU de rack y colectores de enfriamiento líquido. A medida que las densidades de potencia de los servidores siguen aumentando, se volverá muy difícil, si no imposible, acomodar la distribución de energía y enfriamiento necesaria en la parte posterior de un rack de ancho estándar (es decir, 600 mm / 24 in). Además, es probable que los racks estrechos congestionen el flujo de aire de escape detrás del rack debido a los cables de red y alimentación.

**GUÍA:** Se recomienda racks de al menos 750 mm (29.5 in) de ancho para alojar las PDU del rack y, en el caso de enfriamiento líquido, colectores para servidores enfriados con líquido. Aunque estos racks no se alinearán a través de las placas del piso elevado de 600 mm de ancho como lo hacen los racks estándar de 600 mm, esto ya no es una restricción relevante. Esto se debe a que los servidores de IA enfriados por aire requieren altos índices de flujo de

<sup>20</sup> En el enfriamiento por inmersión se sumerge todo el chip o incluso el servidor en el fluido dieléctrico.

aire y los pisos elevados no se utilizan normalmente para la distribución de aire, sino más bien para las tuberías y el cableado.

## Los racks de profundidad estándar carecen de espacio para el cableado y los servidores de IA profundos

Los servidores optimizados para cargas de trabajo de IA pueden alcanzar profundidades que superan la profundidad máxima de montaje de algunos racks estándar. Incluso si un servidor profundo puede montarse en un rack poco profundo, se necesita suficiente espacio libre en la parte posterior para acomodar el cableado de la red y, al mismo tiempo, permitir un flujo de aire suficiente.

**GUÍA:** Los racks de TI tienen rieles de montaje ajustables para acomodar diferentes profundidades del equipo de TI, sin embargo, las profundidades máximas de montaje varían. Se recomienda racks de al menos 1,200 mm (47.2 in) de profundidad con profundidades de montaje máximas mayores que 1,000 mm (40 in).

## Los racks de altura estándar carecen de espacio para la cantidad requerida de servidores

Dependiendo de la altura de los servidores de IA, los racks comunes de 42U de altura probablemente sean demasiado cortos para alojar todos los servidores, interruptores y otros equipos. Por ejemplo, un interruptor de red de 64 puertos implica que el rack tendría 8 servidores, cada uno con 8 GPU. A esta densidad, y suponiendo una altura de servidor de 5U, los servidores por sí solos consumirían 40U, dejando solo 2U de espacio restante para alojar otros dispositivos.

**GUÍA:** Se recomienda implementar clústeres de entrenamiento de IA en racks de 48U o superior, con el supuesto de que las puertas de los centros de datos son lo suficientemente altas como para alojarlos. [1U](#) es igual a 44.45 mm (1.75 in)<sup>21</sup>.

## Los racks estándar carecen de la capacidad suficiente para soportar el peso de los equipos de IA

Con servidores de IA pesados, un rack de alta densidad puede pesar más de 900 kg (2000 lb). Esto coloca una carga significativa en los racks de TI y los pisos elevados, tanto en términos de capacidad de carga estática como dinámica (de balanceo). Los racks que no están clasificados para estos pesos pueden experimentar deformación en marcos, patas niveladoras o ruedas. Además, es posible que los pisos elevados no soporten estos racks pesados.

**GUÍA:** Las capacidades de soporte de peso del rack de TI se especifican como estáticas y dinámicas. Estático se refiere al peso que un rack puede soportar mientras está estacionario. Dinámico se refiere al peso que un rack puede soportar mientras se mueve. Se recomienda especificar racks con una capacidad de peso estático superior a 1,800 kg (3,968 lb) y una capacidad de peso dinámico superior a 1,200 kg (2,646 lb). Estas capacidades de rack deben ser validadas por un tercero independiente.<sup>22</sup> Incluso si tu implementación actual de IA es pequeña y aún no requieres estas capacidades, los racks tienden a tener una vida útil más larga que los equipos de TI. Es probable que la próxima generación de tu implementación de IA requiera algunas o todas estas recomendaciones de rack. Finalmente, en algunos casos, los racks de TI se preconfiguran fuera del sitio y luego se transportan al centro de

<sup>21</sup> Por ejemplo, 48U significa que hay 2.13 m (84 in) de espacio vertical interior disponible para el equipo.

<sup>22</sup> Se recomienda Underwriters Laboratory (UL) y la International Safe Transit Association (ISTA). Para obtener más información, consulte el Informe técnico n.º 201, "[Cómo elegir un rack de TI](#)".

## Herramientas de software

datos. Estos racks deben ser capaces de soportar las fuerzas dinámicas generadas durante el transporte, y el embalaje asociado también debe proteger los racks y el valioso equipo de TI que soportan.

Los pisos del centro de datos, y los pisos elevados en particular, deben evaluarse para garantizar que puedan soportar el peso de un clúster de IA. Esto es especialmente importante para la capacidad dinámica del piso elevado cuando se mueven racks pesados alrededor del centro de datos.

Las herramientas de software de infraestructura física apoyan el diseño y el funcionamiento del centro de datos e incluyen [DCIM \(Data Center Infrastructure Management\)](#), [EPMS \(Energy Power Management System\)](#), [BMS \(Building Management System\)](#) y [herramientas digitales de diseño eléctrico](#). Contar con clústeres de TI de alta densidad de potencia y enfriados por líquido junto con equipos de TI tradicional enfriados por aire significa que ciertas funciones de software se vuelven más críticas. Aunque es posible que algunas cargas de trabajo de entrenamiento de IA no requieran alta disponibilidad, un diseño y una supervisión deficientes pueden provocar riesgos de inactividad para los racks e inquilinos adyacentes que probablemente sean críticos para el negocio. Los siguientes dos desafíos destacan importantes funciones del software de administración que se vuelven más relevantes en el contexto de las cargas de trabajo de entrenamiento de IA de alta densidad:

- La densidad de potencia extrema y la demanda de los clústeres de IA lleva a la incertidumbre del diseño
- Un menor margen de error aumenta el riesgo operativo en un entorno dinámico

### La densidad extrema de energía y la demanda de energía del clúster de IA conducen a la incertidumbre del diseño

Antes de actualizar un sitio existente para acomodar nuevos clústeres de IA, se necesita un estudio de viabilidad para confirmar que hay suficiente capacidad de alimentación y enfriamiento, así como la infraestructura necesaria para distribuir esa capacidad a las nuevas cargas. En casos típicos con densidades de energía del rack muy por debajo de los 10 kW y con un exceso de energía a granel y capacidad de enfriamiento, agregar TI estándar podría ser relativamente fácil y no requerir tanto escrutinio y verificación. Las mediciones de energía y enfriamiento en un momento dado se pueden utilizar junto con componentes comunes de distribución de energía y unidades de enfriamiento existentes con las que estés familiarizado. Este enfoque de diseño de modernización más manual y "a ojo" no será suficiente para grandes grupos de entrenamiento de IA de alta densidad. Un clúster de IA que consume cientos de kilovatios tiene consecuencias mucho mayores si se comete un error de diseño (por ejemplo, no conocer el consumo de corriente real entre el pico y el promedio, no estar seguro de qué cargas hay en qué circuitos, etc.). No puedes permitirte tener incógnitas e incertidumbres con el diseño. Además, debido a que los diseños de clústeres de IA son tan particulares (por ejemplo, rPDU/electroductos de alto amperaje no estándar, uso de enfriamiento líquido, etc.), existe una mayor incertidumbre sobre cómo funcionará el clúster en el arranque.

**GUÍA:** Se recomienda utilizar **EPMS** y **DCIM** para proporcionar una visión precisa de la capacidad de energía actual y sus tendencias, tanto a nivel de energía masiva como de distribución dentro del espacio de TI. Estas herramientas mostrarán cuál es el consumo máximo real de energía durante un largo período de tiempo. Es importante comprender esto para asegurarte de no disparar un interruptor automático sin darte cuenta. Esta evaluación de la capacidad ayudará a determinar la capacidad de alojar cargas de IA. Hay que tener en cuenta que esto supone que los medidores de energía necesarios están en su lugar. A continuación, antes de cualquier



cambio, se recomienda realizar estudios técnicos y de seguridad, que incluyen análisis de capacidad, coordinación de protecciones, estudio de arco eléctrico (arc-flash) y evaluación de cortocircuitos<sup>23</sup>. El uso de **herramientas de software de diseño eléctrico (también conocido como ingeniería de sistemas de energía)** simplifica la cantidad de cálculos y recolección de datos.

Después de la evaluación, es probable que se requieran cambios en la red eléctrica para agregar los clústeres de IA. En este caso, las herramientas de software de diseño eléctrico garantizan que se cuente con los datos correctos para seleccionar el equipo eléctrico óptimo, prevenir fallas eléctricas, desarrollar métodos eficaces de procedimiento e implementar protocolos de seguridad adecuados al trabajar y mantener la red eléctrica en el espacio de TI.

Es importante tener en cuenta que los centros de datos existentes con **diagramas unifilares digitalizados (iSLD)**<sup>24</sup> podrían simplificar el proceso de evaluación descrito anteriormente. Cuando se utilizan iSLD precisos e inteligentes, el tiempo y la experiencia necesarios para recopilar datos y realizar los cálculos se reducen en gran medida. Un iSLD es un diagrama unifilar más avanzado almacenado y administrado en software especializado que incluye funcionalidad y conocimiento avanzados de las características y comportamiento operativo de los dispositivos. Crea un gemelo digital de la red eléctrica física. Básicamente, esta plataforma de software se puede usar para diseñar la red eléctrica, crear y mantener el SLD, y realizar todos los estudios técnicos y evaluaciones de seguridad.

## Un menor margen de error aumenta el riesgo operativo en un entorno dinámico

Suponiendo que se haya implementado un diseño óptimo del centro de datos siguiendo las orientaciones del primer reto, el funcionamiento del "primer día" debería ir sin problemas. Sin embargo, en comparación con otros tipos de instalaciones, los centros de datos son entornos dinámicos en los que se producen movimientos, adiciones y cambios frecuentes de equipos de TI. A medida que los márgenes de seguridad de la capacidad se reducen, como es probable con la adición de un gran clúster de IA, el riesgo de disparar un interruptor automático y crear una zona activa o infrutilizar recursos aumenta a medida que las cargas cambian con el tiempo dentro del espacio de TI. Las razones subyacentes del mayor riesgo son las altas densidades de rack y las bajas relaciones pico/promedio (cerca de 1) de los clústeres de IA, que se analizaron anteriormente. Un menor margen de error significa que los operadores deben conocer cada vez más la situación para evitar el tiempo de inactividad y garantizar un uso eficiente de los recursos disponibles durante toda la vida útil del centro de datos.

**GUÍA:** Se recomienda crear un gemelo digital de todo el espacio de TI (incluyendo el equipo y las máquinas virtuales en los racks) que minimice o evite los desafíos mencionados anteriormente. Este diseño debe mantenerse con el tiempo. Las funciones de modelado y planificación de la DCIM permiten operar planos de piso de espacio de TI efectivos usando una herramienta basada en reglas. Al agregar o mover digitalmente cargas de TI, se puede validar que haya suficientes capacidades de energía, enfriamiento y peso del piso para soportarlas. La DCIM crea un gemelo digital del espacio de TI y documenta todas las dependencias de equipos en los recursos. Esto fundamenta las decisiones para evitar infrutilizar recursos y minimizar el error humano que podría ocasionar el tiempo de inactividad. EPMS y DCIM en conjunto le permiten monitorear las capacidades de energía en todas las PDU, UPS, rPDU, etc. para recibir advertencias tempranas de que se exceden los umbrales de energía para

<sup>23</sup> Es decir, evaluar la capacidad, las clasificaciones de kA y otras especificaciones para la idoneidad para el diseño dado.

<sup>24</sup> Algunos proveedores ofrecen la creación y el mantenimiento de iSLD como un servicio.

evitar el tiempo de inactividad. El software de DCIM aconsejará el mejor lugar para ubicar nuevos equipos en función de los requisitos de energía, enfriamiento, nivel de redundancia, así como el espacio disponible en unidades de rack (U), el puerto de red y la capacidad de peso. Esto se aplica más a equipos que no son de IA y a servidores de inferencia de IA. A diferencia de las cargas de inferencia, las cargas de entrenamiento de IA requieren una configuración prediseñada que rara vez cambia, si acaso cambia.

Muchas herramientas de software de planificación y modelado de DCIM incluyen una herramienta de dinámica de fluidos computacional (CFD) para garantizar un flujo de aire adecuado dada la disposición física del equipo y la carga térmica. El DCIM se puede utilizar para ayudar a optimizar la capacidad de enfriamiento al liberar capacidad de enfriamiento infrautilizada mediante la colocación y configuración óptimas de la infraestructura y las cargas. En términos de movimientos, adiciones y cambios de cargas de IA, la CFD se aplica más a las cargas de inferencia de IA, ya que se agregan más servidores para satisfacer la demanda del usuario (es decir, consultas). Hay que tener en cuenta que, en algunos casos, el clúster de entrenamiento o inferencia de IA está aislado en su propio segmento de alimentación y arquitectura de enfriamiento. En estos casos, las cargas que no son de IA son menos susceptibles a los efectos del clúster de IA. Sin embargo, en ambos casos es beneficioso establecer un gemelo digital de estos espacios.

Hasta ahora, la orientación se centra en las tecnologías y los enfoques de diseño disponibles en la actualidad. En esta sección se describen brevemente algunas tecnologías y enfoques de diseño *futuros* que creemos que ayudarán a superar los retos planteados.

## Perspectivas futuras de la infraestructura para soportar la IA

- **rPDU estándar optimizadas para IA:** los factores de forma cambiarán para ajustarse a servidores con mayor densidad de potencia y menos receptáculos trenzados. La eliminación de receptáculos innecesarios permite más rPDU en cada rack o una única rPDU de mayor capacidad (con capacidad nominal para hasta 150 amperios a 240 V, 86 kW menos). Estas rPDU también proporcionarían receptáculos para equipos de baja densidad como interruptores.
- **Transformadores de media tensión a 415/240 V en el espacio técnico/de TI:** la distribución de energía a media tensión (por ejemplo, 13 kV) reduce la cantidad de cobre, requiere menos conductores y reduce el tiempo de instalación. Por ejemplo, la distribución de TI utilizaría un transformador de 2 MW para alimentar un electroducto de 3 000 A a 415/240 V, lo que alimentaría todo un clúster de IA o una parte de uno mayor a 2 MW. Esta arquitectura de distribución también elimina los tradicionales transformadores de 13 kV a 480/277 V y el cuadro de distribución aguas arriba de la distribución de TI. Esto también puede mitigar las restricciones de la cadena de suministro del equipo de distribución de 480 V.
- **Transformadores de estado sólido:** se trata esencialmente de convertidores de componentes electrónicos de potencia. Utilizan componentes semiconductores para cambiar la tensión primaria a una tensión secundaria. Utilizan un transformador de media frecuencia (MFT) [aislado galvánicamente](#) en los lados primario y secundario. Mientras que los transformadores tradicionales son pesados y funcionan solo con corriente alterna (CA), los transformadores de estado sólido son pequeños y ligeros, y se convierten entre tensión de CA y CC.
- **Interruptores automáticos de estado sólido:** estos interruptores automáticos utilizan semiconductores para activar o desactivar el flujo de corriente. Esto es de particular importancia cuando se interrumpe el flujo de corriente hacia una falla. Sin embargo, para que se considere un interruptor automático, también debe utilizar un interruptor mecánico en serie con los semiconductores para proporcionar [aislamiento galvánico](#). Los interruptores de estado sólido permitirían una operación más rápida y la capacidad de controlar más estrechamente las corrientes de falla. Esto sería muy beneficioso para reducir la energía del arco eléctrico en racks de IA de alta densidad.

- **Fluidos dieléctricos sostenibles:** estos pueden sustituir el agua como la opción actual para el enfriamiento directo al chip si aumentan la eficiencia de transferencia de calor y permiten mayores TDP de los chips.
- **Racks de TI ultra profundos:** a medida que se introducen servidores basados en aceleradores más profundos, los racks más profundos se adaptarían no solo al servidor, sino también al cableado de red, las tuberías de agua y las PDU de rack.
- **Mayor interacción/optimización con la red:** la programación de cargas de trabajo basadas en condiciones de servicios y microrredes ayuda a equilibrar la red y ahorrar en electricidad. Migrar cargas a diferentes zonas de redundancia o poner una UPS en funcionamiento con batería son ejemplos de gestión de cargas de trabajo.

## Conclusión

El rápido crecimiento y la aplicación de la IA están cambiando el diseño y el funcionamiento de los centros de datos. Se calcula que las cargas de trabajo de IA representarán del 15 % al 20 % de la energía total del centro de datos para el año 2028. Aunque se espera que las cargas de trabajo de **inferencia** consuman mucha más energía que los clústeres de entrenamiento, funcionan con una amplia gama de densidades de rack. Por otra parte, las cargas de trabajo de **entrenamiento** de la IA operan siempre a densidades muy altas, que van de 20 a 100 kW por rack o más. Las demandas de red y los costos obligan a agrupar estos racks de entrenamiento. Estos clústeres de extrema densidad de energía son fundamentalmente lo que desafía el diseño de la alimentación, el enfriamiento, los racks y la gestión del software en los centros de datos. En este documento, se proporciona orientación sobre cómo se abordan los desafíos. Se resumen a continuación:

**ENERGÍA:** El uso de la distribución de 120/208 V (en NAM) ya no es suficiente, y en cambio, se recomienda la distribución de 240/415 V para limitar el número de circuitos dentro de racks de alta densidad. Incluso para tensiones más altas, sigue siendo un desafío proporcionar capacidad suficiente con PDU de rack estándar de 60/63 amperios. Por ejemplo, los racks enfriados por líquido están limitados a dos unidades rPDU y proporcionan 69/87 kW. Para la seguridad del personal, se recomienda una evaluación del riesgo de destello por arqueo y un análisis de carga para garantizar que se utilicen los conectores, receptáculos y unidades rPDU adecuadas según sus temperaturas expuestas. Los tamaños de los bloques de distribución aguas arriba deben ser lo suficientemente grandes como para admitir una sola fila de un clúster de IA.

**ENFRIAMIENTO:** Aunque el enfriamiento por aire seguirá existiendo en el futuro cercano, se predice una transición del enfriamiento por aire al enfriamiento líquido como una solución preferida o necesaria para centros de datos con clústeres de IA. En comparación con el enfriamiento por aire, el enfriamiento líquido proporciona muchos beneficios, como una mayor confiabilidad y rendimiento del procesador, ahorros de espacio con densidades de rack más altas, más inercia térmica con agua en las tuberías, mayor eficiencia energética, mejor utilización de la energía (más energía se destina a TI) y menor uso de agua. Los operadores de centros de datos pueden usar nuestra guía propuesta para lograr una transición exitosa del enfriamiento por aire al enfriamiento líquido para admitir las cargas de trabajo de IA.

**RACKS:** Con los clústeres de IA, los servidores son más profundos, las demandas de alimentación son mayores y el enfriamiento es más complejo. Como resultado, se recomienda el uso de racks de mayores dimensiones y capacidad de peso, específicamente: al menos 750 mm (29.5 in) de ancho, 1,200 mm (47.2 in) de profundidad, 48U de alto, con 1,000 mm (40 in) de profundidad de montaje, capacidad de peso estático mayor a 1,800 kg (3,968 lb) y una capacidad de peso dinámico mayor a 1,200 kg (2,646 lb).

**GESTIÓN DE SOFTWARE:** Cuando se administran clústeres de IA, las herramientas de software como DCIM, EPMS, BMS y las herramientas de diseño eléctrico digital se vuelven críticas. Disminuyen el riesgo de comportamientos inesperados con redes eléctricas complejas. También proporcionan un gemelo digital del centro de datos para identificar recursos de energía y enfriamiento restringidos para fundamentar las decisiones de diseño.

## Acerca de los autores

**Víctor Avelares** analista de investigación sénior en el Centro de Investigación de Gestión Energética de Schneider Electric. Es responsable del diseño de centro de datos e investigación de operaciones, y consulta con los clientes sobre la evaluación de riesgos y las prácticas de diseño para optimizar la disponibilidad y la eficiencia de los entornos de sus centros de datos. Víctor posee una licenciatura en Ingeniería Mecánica del Instituto Politécnico Rensselaer y una maestría en Administración de Empresas de Babson College. Es miembro de la AFCOM.

**Patrick Donovan** es analista de investigación sénior en el Centro de Investigación de Gestión Energética de Schneider Electric. Tiene más de 27 años de experiencia en el desarrollo y el servicio técnico de los sistemas de alimentación y de enfriamiento de la división de negocios de Energía Segura de Schneider Electric, incluidas varias soluciones galardonadas de protección, eficiencia y disponibilidad de alimentación. Autor de numerosos documentos técnicos, artículos de la industria y evaluaciones de tecnología, la investigación de Patrick sobre tecnologías y mercados de infraestructura física del centro de datos ofrece orientación y asesoramiento sobre las mejores prácticas para la planificación, el diseño y el funcionamiento de las instalaciones del centro de datos.










**Paul Lin** es director de investigación y experto en Edison, en el Centro de investigación de gestión de energía de Schneider Electric. Es responsable del diseño de centros de datos e investigación de operaciones, y asesora a los clientes en materia de evaluación de riesgos y prácticas de diseño para optimizar la disponibilidad y la sostenibilidad del entorno de sus centros de datos. Es un reconocido experto y un frecuente orador y panelista en eventos de la industria de centros de datos. Antes de unirse a Schneider Electric, Paul trabajó como líder de proyectos de I+D en LG Electronics durante varios años. También es ingeniero profesional registrado y tiene más de 10 patentes. Paul posee una licenciatura y una maestría en ciencias en ingeniería mecánica de la Universidad de Jilin. También posee un certificado del Programa de Liderazgo Transformacional de Schneider en INSEAD.

**Wendy Torell** es analista de investigación sénior en el Centro de Investigación del Centro de Datos de Schneider Electric. En esta función, ella investiga las mejores prácticas en el diseño y operación del centro de datos, publica informes técnicos y artículos, y desarrolla herramientas de compensación para ayudar a los clientes a optimizar la disponibilidad, la eficiencia y el costo de sus entornos de centros de datos. También consulta con los clientes sobre los métodos de la ciencia de la disponibilidad y las prácticas de diseño, para ayudarles a cumplir los objetivos de rendimiento del centro de datos. Recibió su licenciatura en Ingeniería Mecánica de Union College en Schenectady, Nueva York y su maestría en Administración de Empresas de la Universidad de Rhode Island. Wendy es una ingeniera de confiabilidad certificada por ASQ.

**María A. Torres Arango** es analista de investigación en el Centro de Investigación de Gestión Energética de Schneider Electric. En esta función, María investiga temas estratégicos técnicos para fundamentar la toma de decisiones, centrándose actualmente en los sistemas de almacenamiento de energía y la sostenibilidad. María es licenciada en Ingeniería Aeronáutica por la Universidad Pontificia Bolivariana, Colombia; y cuenta con una maestría en Ingeniería Aeroespacial y un doctorado en Ciencia e Ingeniería de Materiales por la Universidad de Virginia Occidental.

**CALIFICA ESTE INFORME** ★★★★★



- 
[Distribución de energía de CA de alta eficiencia para centros de datos](#)  
 Informe técnico n.º 128
- 
[Prácticas de diseño de centros de datos para integrar cargas de trabajo de IA enfriadas por líquido](#)  
 Informe técnico n.º 133
- 
[Consideraciones sobre el destello por arqueo para el espacio de TI del centro de datos](#)  
 Informe técnico n.º 194
- 
[Beneficios de limitar la corriente de cortocircuito de las MV en centros de datos grandes](#)  
 Informe técnico n.º 253
- 
[Tecnologías de enfriamiento líquido para centros de datos y aplicaciones perimetrales](#)  
 Informe técnico n.º 265
- 
[Cinco razones para adoptar el enfriamiento líquido](#)  
 Informe técnico n.º 279
- 
[Comparación de fluidos dieléctricos para enfriamiento líquido envolvente de equipos de TI](#)  
 Informe técnico n.º 291
- 
[Mitigación del arco eléctrico \(arc-flash\)](#)  
 Informe técnico
- 
[Explorar todos los informes técnicos](https://whitepapers.apc.com)  
[whitepapers.apc.com](https://whitepapers.apc.com)
- 
[Explorar todas las TradeOff Tools™](https://tools.apc.com)  
[tools.apc.com](https://tools.apc.com)

**Nota:** Los enlaces de Internet pueden quedar obsoletos con el tiempo. Los enlaces a los que se hace referencia estaban disponibles al momento de redactar este informe, pero es posible que ya no estén disponibles.

## Contáctenos

Para incluir comentarios sobre el contenido de este informe técnico:

Centro de investigación de gestión de energía de Schneider Electric  
[dcsc@schneider-electric.com](mailto:dcsc@schneider-electric.com)

Si tu eres un cliente y tiene preguntas específicas sobre su proyecto de centro de datos:

Comunícate con tu representante de Schneider Electric en  
[www.apc.com/support/contact/index.cfm](https://www.apc.com/support/contact/index.cfm)