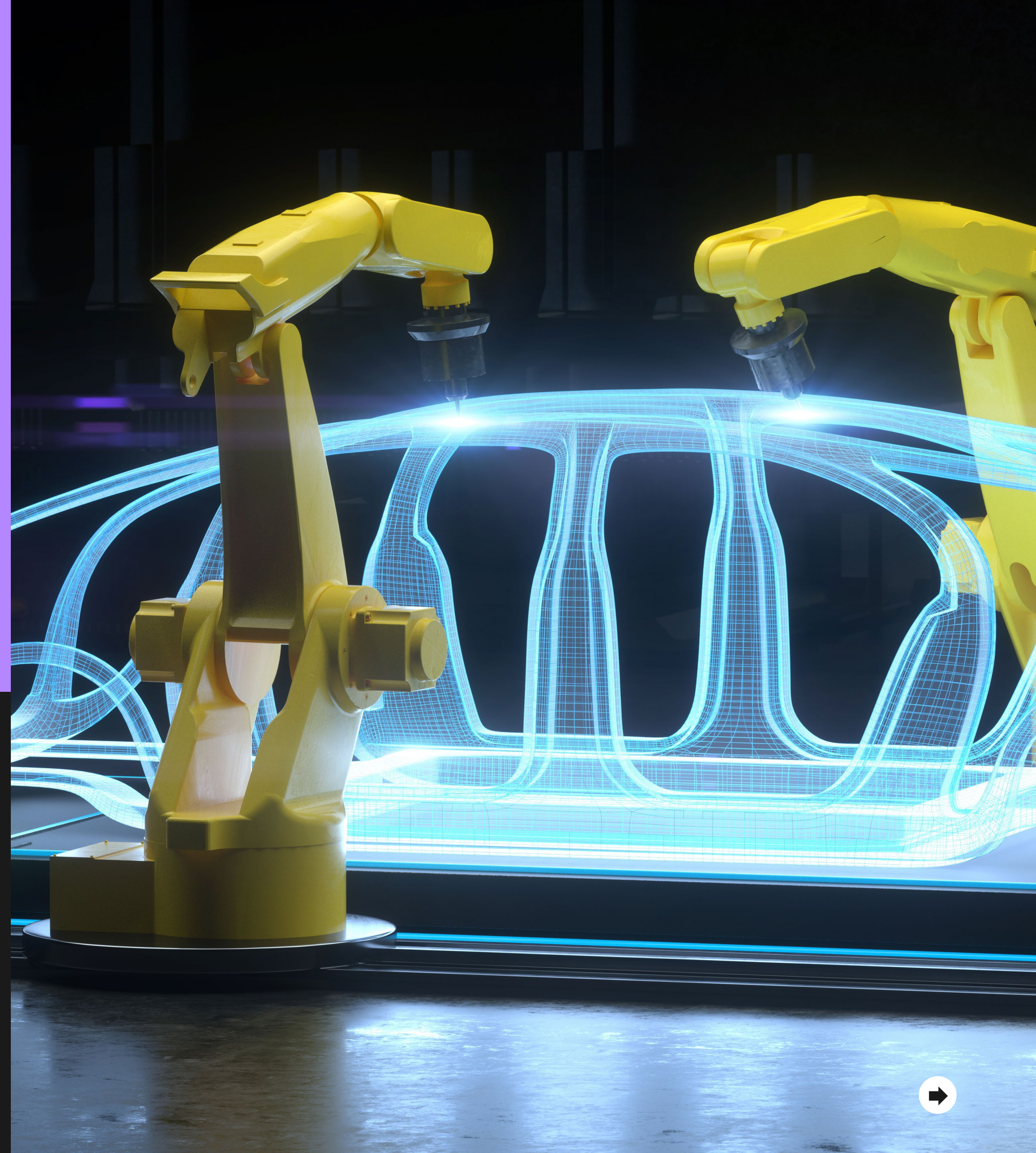


E-BOOK

# Transformamos el sector con la IA condicionada por los datos

➔ [netapp.com/ai](https://netapp.com/ai)

 **NetApp**





# La IA condicionada por los datos líder

## Resumen ejecutivo

La inteligencia artificial (IA) está transformando miles de empresas en diferentes ámbitos de todo el mundo. Los rápidos avances requieren unas cantidades enormes de datos. Sin embargo, la IA vale lo que valen los datos que usa. Por tanto, la calidad y el movimiento de los datos resultan vitales. Esto significa que usted necesita una arquitectura de datos inteligente, potente y fiable que le permita desatar todo el potencial de la IA.

Da igual en qué etapa del camino hacia la IA se encuentre, este e-Book puede ayudarle a trazar un recorrido más directo hacia el éxito.

Siga leyendo para obtener más información acerca de:

- La importancia de maximizar su canalización de datos
- Cómo la división de silos de datos mejora el flujo de datos
- Los enormes beneficios empresariales del rendimiento y la escalabilidad
- Recomendaciones para cumplir con los requisitos de la arquitectura de IA
- Transformar su empresa con soluciones de IA fiables

➔ [netapp.com/ai](https://netapp.com/ai)





# Contenido

- 1. IA condicionada por los datos 4 ➔
- 2. Domine la canalización de sus datos 6 ➔
- 3. Integre su solución de Data Fabric en el perímetro, el núcleo y cloud 9 ➔
- 4. Ofrezca la potencia del rendimiento con escalabilidad 10 ➔
- 5. Supere los retos de la arquitectura de IA 14 ➔
- 6. Soluciones de IA inteligentes, potentes y fiables de NetApp 16 ➔
- 7. Acelere su paso a la IA 18 ➔



# 1. IA condicionada por los datos

Desde el compromiso del cliente hasta los procesos de optimización y la creación de una ventaja competitiva sostenible, la IA ha abierto un amplio abanico de oportunidades de negocio en muchos sectores.

Miles de empresas de todo el mundo están trabajando duro para acceder a este valor. Se espera que la IA se encuentre entre las prioridades en cuanto a inversión en tecnología durante al menos cinco años, lo que afectará a la transformación digital en todos los sectores.

Con la ayuda de la IA, la información que antes era demasiado compleja de concebir se encuentra ahora al alcance de la mano. Ya no se trata solo de datos cuantitativos, pues las potentes tecnologías actuales de aprendizaje profundo y aprendizaje automático utilizan imágenes y voz para seguir avanzando de cara a los análisis de datos.

Este e-Book le ayudará a acelerar su viaje hacia la IA sin importar en qué etapa del camino hacia la transformación digital se encuentre su organización. Descubra cómo una arquitectura de IA que es inteligente, potente y fiable puede ofrecerle una plataforma en la que pueda desatar el potencial de sus datos.

**«Hubo un reconocimiento casi universal (del 96 %) de que el Big Data y la IA estaban mostrando resultados»<sup>1</sup>**

NewVantage Partners, encuesta ejecutiva de 2021 sobre Big Data e IA

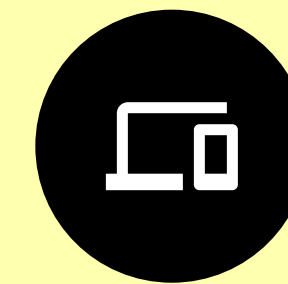
# IA condicionada por los datos

## Los datos: la fuerza que impulsa la IA

Unas cantidades enormes de datos son fundamentales para el éxito de la IA. Así, todos sus datos importantes deben estar disponibles, accesibles y protegidos. Por supuesto, es muy fácil decirlo. Hacerlo ya es otra cosa. En las empresas de hoy en día, las fuentes de datos que alimentan los sistemas de IA son muy diferentes a las del pasado. Proporcionan datos de mayor calidad, pero hay muchas más en muchos más sitios. Estamos hablando de marañas de complejidad distribuidas, diversas y dinámicas.

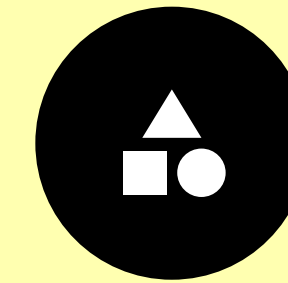
### PPDAI está haciendo lo imaginable

PPDAI está transformando los servicios financieros en China al ofrecer servicios en línea a los desatendidos, es decir, aquellos residentes que no tienen datos de información crediticia. Como la primera empresa financiera de Internet del país, PPDAI depende de unas enormes cantidades de datos que provienen de diferentes fuentes para cubrir las necesidades de sus 60 millones de clientes. Cuando no hay calificaciones crediticias, PPDAI usa el aprendizaje automático y la IA para analizar los datos de móviles, redes sociales y comportamientos y así evaluar el riesgo. El proceso es rápido. La gente que solicita préstamos a través de la aplicación móvil de PPDAI y que cumple con los requisitos recibe el dinero en minutos. El alcance y el éxito de la plataforma de PPDAI es posible solo gracias a un análisis y un movimiento de datos rápidos y seguros.



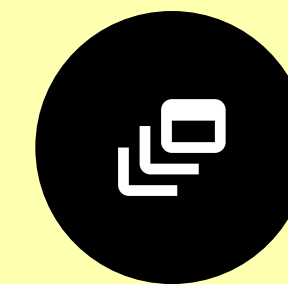
### Distribuidos.

Para superar un gran reto mediante la IA, los datos que necesita ya no están solamente en su centro de datos. Podrían venir de dispositivos del Internet de las cosas (IoT) en el perímetro de su red, de cloud e incluso de fuentes de terceros.



### Diversos.

Vídeo. Audio. Imágenes. Datos transaccionales generados automáticamente. Y mucho más... Los proyectos de IA suelen requerir una variedad de fuentes de datos tanto estructurados como no estructurados. Justo cuando creía que tenía controlada la complejidad, llegan fuentes y tipos de datos nuevos que podrían complicar aún más la gobernanza y la gestión de los datos.



### Dinámicos.

Los conjuntos de datos de IA están en constante crecimiento y cambio. Por tanto, no resulta fácil saber dónde están los datos, de dónde vienen y dónde se almacenan los archivos.

Piense en una organización que esté usando datos agregados desde fuentes de todo el mundo, incluidos dispositivos de punto de venta, aplicaciones móviles y redes sociales. Para obtener información útil, la organización debe mantener dichos datos actualizados prácticamente en tiempo real.

## 2. Maximice la canalización de sus datos

¿Cómo de importante es la gestión de datos para los procesos de aprendizaje automático y aprendizaje profundo? Viene a ser como un entrenamiento de alta calidad para un equipo deportivo profesional. Ignorar lo más básico puede conllevar errores y chapuzas a pesar del calibre de los jugadores. Y los equipos tecnológicos que son nuevos en el ámbito de la IA suelen subestimar esta relación. Una encuesta de la Sloan del MIT identificó los problemas en cuanto a los datos más comunes que surgen al adoptar la IA. Los encontrará a continuación.

Falta de datos.

Los algoritmos de IA se basan en los datos para predecir eventos del mundo real. Para crear un algoritmo de IA que se anticipe a los fallos, necesita más que los datos de las operaciones normales. Necesita también una gran cantidad de datos registrados de fallos que ya han ocurrido.

### **Preparativos inesperados.**

Identificar las fuentes de datos adecuadas, recopilar los datos y prepararlos puede requerir mucho tiempo.

### **Propiedad de datos confusa.**

Es posible que le resulte difícil obtener acceso a determinados datos valiosos porque su propiedad no está clara.

### **Unir los puntos de cara a los datos..**

Especialmente en grandes organizaciones, los datos pueden estar fragmentados en múltiples sistemas corporativos, lo cual complica el proceso de formación.

Todos estos retos surgen del flujo de datos. Pasar por alto cualquiera de ellos podría hacer que sus esfuerzos en cuanto a la IA se fueran a pique. Por eso es tan crucial la canalización de sus datos.

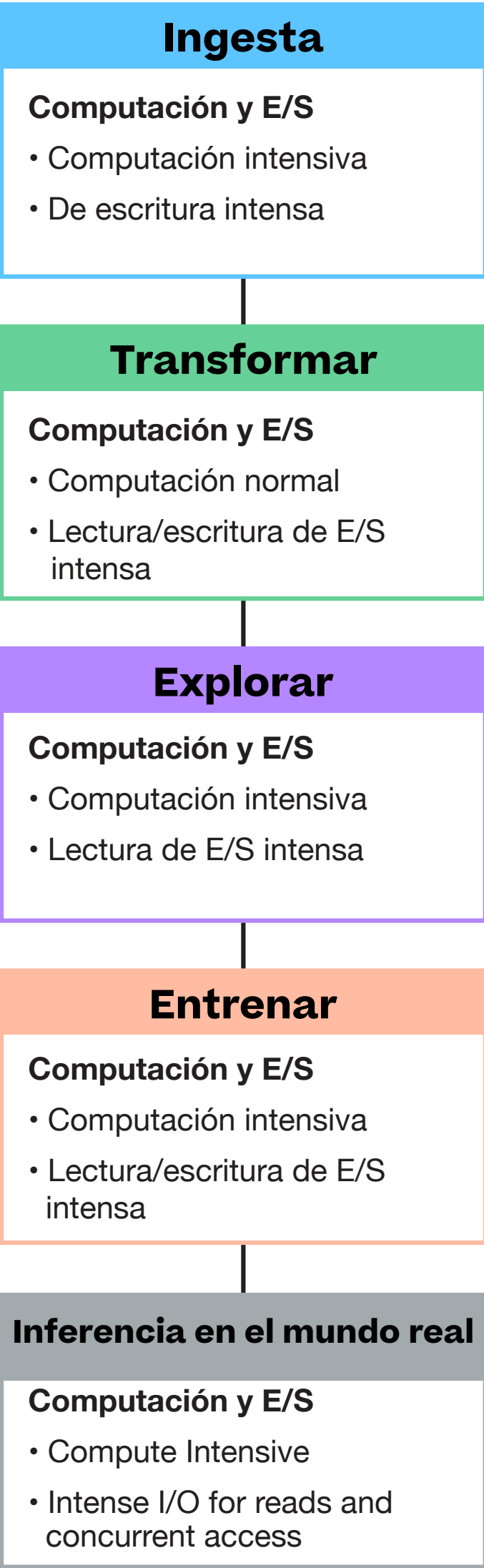


# Maximice la canalización de sus datos

## ¿Qué es la canalización de datos?

Una canalización de datos permite que los datos fluyan sin problemas desde el punto de creación pasando por todo el proceso de IA. Y los puntos de creación son prácticamente infinitos... ¿Un dispositivo IoT en una fábrica? Sí. ¿Equipos del cliente? Desde luego. ¿Una plataforma de redes sociales o un sistema de registros corporativo? Exacto.

Las etapas de la canalización incluyen las que se muestran en la figura 1.



La creación de datos a menudo ocurre en el perímetro de su red en cosas como los dispositivos inteligentes de punto de venta. Los dispositivos y sensores conectados por el IoT se están extendiendo cada vez más por todos los sectores.

Antes de empezar cualquier formación, necesita normalizar los datos. Las opciones de preprocesamiento incluyen un lago de datos, un almacén de objetos en cloud de Amazon Simple Storage Service (Amazon S3) o un almacén de archivos en las instalaciones.

Durante la fase de exploración, se determinan los modelos de aprendizaje profundo más prometedores en función de sus conjuntos de datos y el resultado que desea obtener.

La formación de los modelos de inferencia incluye el movimiento regular de los datos al clúster de formación. Es un proceso repetitivo que tiene lugar a intervalos regulares a lo largo de la vida de un modelo.

Los modelos de inferencia formados y contrastados se ponen en marcha en el mundo real. A menudo, los modelos se almacenan en un repositorio de tipo DevOps, donde se benefician de un acceso de lectura de latencia ultrabaja.

Figura 1) Una canalización de datos bien diseñada permite que fluyan libremente por las diferentes fases de la IA, cumpliendo los requisitos de I/O particulares de cada una de ellas y evitando los cuellos de botella.

# Maximice la canalización de sus datos

## Los inhaladores inteligentes usan un enfoque del tipo perímetro-núcleo-cloud para mejorar la salud de los pacientes

Tanto si realiza los procesos de IA en cloud o en las instalaciones, una canalización de datos debe ser capaz de satisfacer los requisitos informáticos y de E/S particulares de cada fase. Un enfoque de almacenamiento de datos que funcione bien en toda la canalización resulta esencial. Piense en cómo cambian los requisitos de computación y E/S a medida que los datos avanzan a lo largo de la canalización con el caso de los inhaladores inteligentes.

En Estados Unidos, 25 millones de personas sufren de asma, que es lo mismo que decir 1 de cada 13 personas. Imagínese el sufrimiento que se podría evitar si se relaciona el uso de los inhaladores con la información en tiempo real acerca de la ubicación, el tiempo, la calidad del aire y la cantidad de polen. Así, se podría ayudar a los pacientes a evitar los detonantes. Se puede hacer añadiendo sensores en los inhaladores para el asma y confiando en la IA.

Cambridge Consultants, un cliente de NetApp y un partner de innovación en IA, demostró el potencial de un inhalador inteligente que dependía de las GPU de NVIDIA y de la gestión de datos de NetApp. Es un gran ejemplo para determinar los requisitos de la IA a escala:

- Los datos fluyen desde miles de dispositivos en el perímetro.
- Los datos se combinan con conjuntos de datos del exterior durante la formación en un centro de datos acelerado por GPU.
- El modelo de inferencia resultante se pone en marcha en cloud para analizar nuevos puntos de datos e identificar y actuar en eventos desencadenantes.



### 3. Integre su solución de Data Fabric en el perímetro, el núcleo y cloud

¿Desea aprovechar las tecnologías de IA para introducir nuevos servicios y mejorar la información obtenida a partir de los datos de la empresa? ¡Estupendo!

Pero no olvide que los flujos de trabajo de IA son complejos. Algunos pasan del perímetro de su red al núcleo de los centros de datos y, finalmente, a cloud. Otros puede que empiecen en el perímetro de su red y vayan directamente a cloud y, desde allí, al núcleo o de nuevo al perímetro.

Dejando a un lado los enredos, puede que los datos también necesiten un preprocesamiento extenso antes de realizar la formación. Este trabajo se puede llevar a cabo en el perímetro, en los centros de datos del núcleo, en cloud o en estos tres sitios. Los modelos de IA pueden formarse en el núcleo o en cloud y es posible que necesite archivar los datos en cloud para usarlos en el futuro o por motivos de cumplimiento de normativas.

NetApp reúne los datos a lo largo de la canalización de IA para que estén disponibles allá donde sean necesarios.

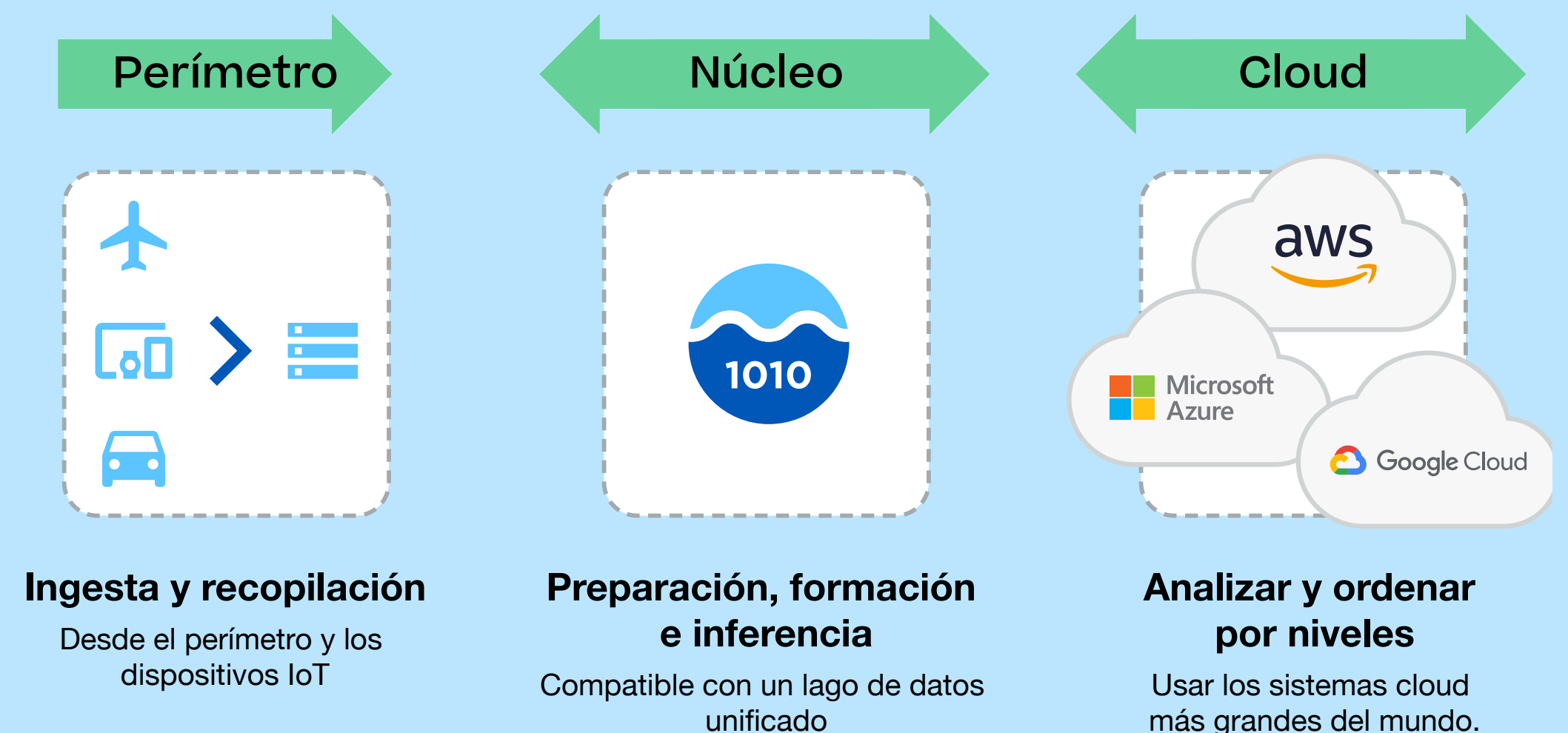


Figura 2) Solo NetApp le permite integrar su Data Fabric en el perímetro, el núcleo y cloud.





## 4. Ofrezca la potencia del rendimiento con escalabilidad

Aquí tiene una frase que merece la pena leer dos veces: para una IA satisfactoria, la canalización de datos requiere un rendimiento y una escalabilidad extremos en cada fase: en el perímetro, en el núcleo y en cloud. Lo que pasa es que demasiadas organizaciones subestiman los retos que surgen a raíz de reunir y gestionar cantidades enormes de datos. Y suele ser una lección dolorosa de aprender.

Los cuellos de botella en cualquier punto de la canalización provocan una infraestructura costosa, aumentan los costes y desperdician el tiempo de los científicos de datos. Estas consecuencias ya son negativas de por sí, pero, en muchos casos de uso, los cuellos de botella también ponen en peligro los resultados. Por ejemplo, si un inhalador inteligente proporciona una alerta acerca de un evento detonante, pero llega tarde y el paciente ya está luchando por recobrar el aliento tras un ataque de asma, no es que haya añadido mucho valor.



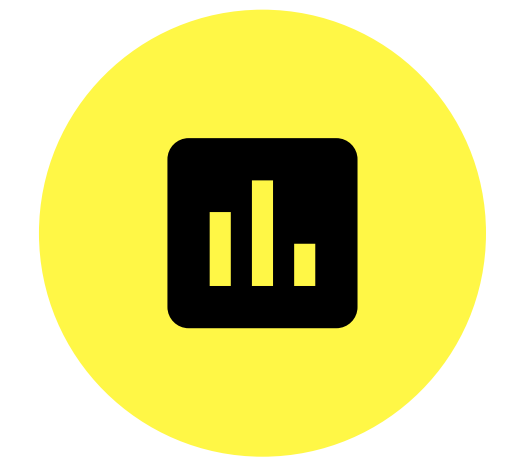
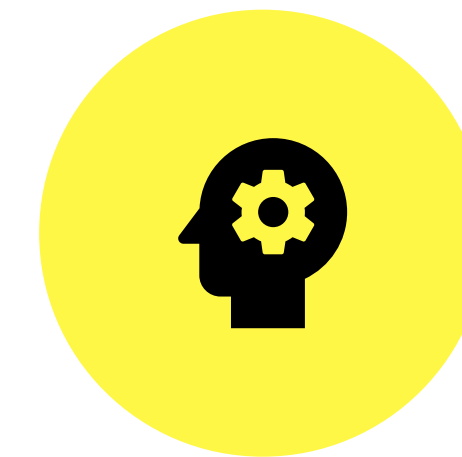
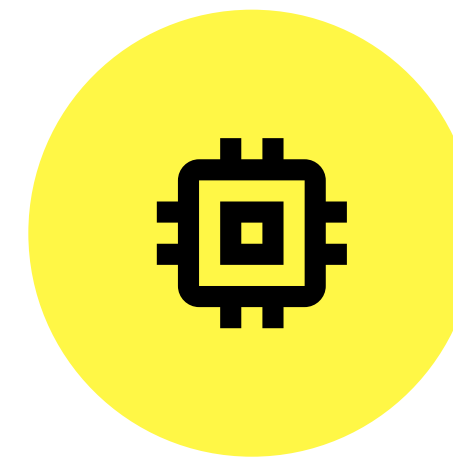
# Ofrezca la potencia del rendimiento con escalabilidad

## Potencia/Rendimiento de GPU: un delicado juego de equilibrio

Para generar conocimientos útiles basados en los datos más recientes, algunos modelos deben estar formados casi en tiempo real. Sin embargo, si el rendimiento de acceso a los datos no coincide con la potencia de las GPU, puede ser casi imposible conseguir inmediatez. ¿Qué hace falta para beneficiarse del rendimiento de las GPU caras?

Un almacenamiento de datos de alto rendimiento.

Tanto si lleva a cabo la formación de IA en las instalaciones como si lo hace en cloud, necesita GPU que funcionen lo más eficientemente posible. Dejarlas inactivas mientras se espera a los datos alarga los ciclos de formación, reduce la productividad de la formación y hace mella en los resultados.



### ¿Por qué son las GPU tan importantes para el aprendizaje profundo?

A principios de la década de los 2000, los científicos de computación se dieron cuenta de que las funcionalidades de manipulación de vídeo de las GPU se podían aplicar a una mayor variedad de problemas de computación en paralelo.

Los algoritmos de aprendizaje profundo esenciales para la IA condicionada por los datos de hoy en día requieren un gran volumen de multiplicaciones de matrices ejecutadas en paralelo. Dicha necesidad hace que las GPU, u otro hardware de procesamiento paralelo similar, sean esenciales para la exploración de datos y la formación de modelos de inferencia. Las GPU actuales, de empresas como NVIDIA, ofrecen cientos de miles de núcleos sencillos.

El rendimiento de las GPU está aumentando rápidamente, así que su canalización de datos tiene que seguirle el ritmo. Es como la diferencia entre conducir un coche deportivo de gama alta en un circuito vacío frente a conducirlo en una autopista en hora punta... A pesar de que quede elegante y dé sensación de velocidad, está a merced de los demás conductores.



# Ofrezca la potencia del rendimiento con escalabilidad

## Ofrezca la potencia del rendimiento con escalabilidad

El tipo de datos de su canalización de IA afecta al patrón de I/O, el cual puede afectar a la velocidad con la que finalizan las formaciones. Los datos no estructurados pueden fusionarse en un conjunto de flujos, lo cual crea un I/O secuencial. Fuentes como bases de datos, registros de sensores, registros de archivos y correos electrónicos deben ser accesibles mediante lecturas aleatorias.

Estos dos tipos de I/O, secuencial y aleatorio, son bastante diferentes. Normalmente, se usan tanto fuentes de datos estructuradas como no estructuradas para formar un modelo de inferencia, por lo que su arquitectura debe ser compatible con ambos tipos de I/O para evitar los cuellos de botella que atascan la canalización.

## Transición sin problemas desde el prototipo hasta la producción

Mientras pasa del prototipo a la producción, es esencial que piense en las necesidades de sus usuarios de datos: los ingenieros de datos y los científicos de datos. Los científicos de datos, por ejemplo, no pueden esperar meses para que el entorno de canalización de datos vaya a producción. Buscan una puesta en marcha rápida, ya sea en las instalaciones o en cloud. Además, una vez en la producción, esperan un escalado del entorno rápido.

Resulta difícil predecir el éxito de una aplicación desde el inicio. Dicho éxito puede conllevar muchos más datos y formaciones mucho más frecuentes. Y eso crea mayores demandas tanto para la GPU como para el E/S, así como para los recursos necesarios para soportar la carga adicional. Por eso, merece la pena considerar arquitecturas potentes que sean sencillas de implementar en línea.



# Ofrezca la potencia del rendimiento con escalabilidad

## Lleve la IA al siguiente nivel

Una vez que su equipo tenga varios proyectos de IA en producción, puede utilizar lo que ya haya creado para expandir nuevos casos de uso de IA. Busque oportunidades para usar un enfoque de IA similar para agilizar procesos empresariales, crear una ventaja competitiva o expandirse a nuevos mercados.

¿Cómo puede lograr el éxito? El éxito reside en escalar horizontalmente la infraestructura y los procesos que ya ha creado e incorporar las mejores y más recientes funcionalidades con soluciones fiables. Recuerde: las inversiones en IA por adelantado pueden ser significativas, ya sea en cloud o en las instalaciones.

Puede empezar mediante un aprovisionamiento de GPU en cloud y luego descubrir que sus datos crecen tan rápidamente que poner en marcha un clúster de formación *in situ* es más rentable. O es posible que empiece *in situ* y pase a cloud según sus datos vayan creciendo. La flexibilidad es la clave del éxito.

«La IA se verá como una arquitectura y ayudará a descubrir una transformación radical de los procesos empresariales y el rendimiento laboral»<sup>2</sup>

IDC, Market Analysis Perspective, Worldwide Artificial Intelligence Software, 2020



## 5. Supere los retos de la arquitectura de IA

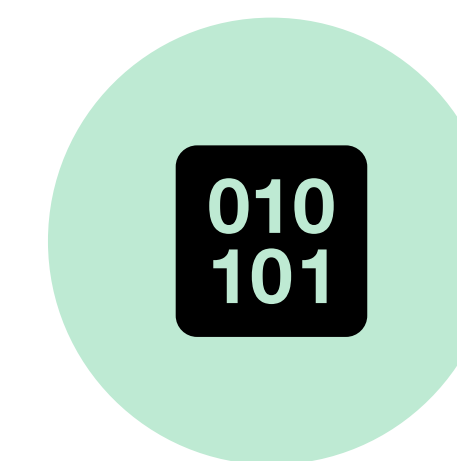
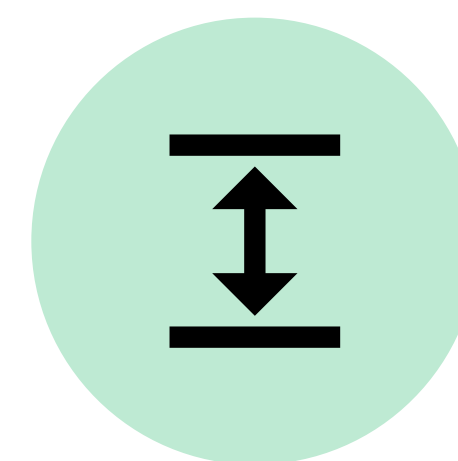
¿Qué opina? ¿Suenan complejo construir una infraestructura de IA efectiva desde el perímetro hasta el núcleo y cloud? Puede ser, pero también existen maneras de construir la infraestructura de modo que sea compatible con un crecimiento significativo.

### Evite las soluciones de punto

Una canalización de datos de IA puede consistir en cinco o más fases diferentes. Confiar en soluciones de punto en cada fase es, sin duda, la mejor receta para la complejidad. Una arquitectura de IA óptima le permite usar la mejor solución de computación y almacenamiento de datos para los requisitos de cada fase al mismo tiempo que ofrece una eficiencia y una gestión de datos uniformes en todas las fases.

### Elija tecnologías que ofrecen una escala superior

Un proyecto de IA puede empezar como algo pequeño, pero puede llegar a crecer hasta alcanzar un tamaño difícil de predecir incluso para los expertos. Esto significa que puede que la canalización de sus datos deba escalar para acomodar petabytes o incluso exabytes de datos. En definitiva, un único vehículo autónomo genera un terabyte o más de datos cada hora en que está operativo, así que es fácil ver cómo algunos conjuntos de formaciones pueden crecer por encima de los 500 PB. Procesar petabytes de datos requiere el equilibrio de un ancho de banda de E/S enorme y el rendimiento de computación. Si su solución no puede escalar para satisfacer los requisitos de rendimiento y capacidad a medida que vayan creciendo... Bueno, prepárese para poner en pausa su iniciativa mientras reestructura sus operaciones.



# Supere los retos de la arquitectura de IA

## Considere el multicloud híbrido para las nuevas tecnologías

El ámbito de la IA está evolucionando a gran velocidad, así que es esencial prepararse para los cambios. Obtener la capacidad de poner en marcha proyectos en las instalaciones o de elegir entre los mejores servicios de IA en cloud le ayudará a maximizar la flexibilidad. La base para el éxito es una canalización de IA integral. Así es como mantendrá las cargas de trabajo en funcionamiento sin importar dónde estén los datos o a dónde se muevan, ya sea al perímetro, al núcleo o a cloud.

## Elija soluciones validadas e integradas para una puesta en marcha rápida

¿Está buscando maneras de reducir el tiempo necesario para poner en marcha la infraestructura de IA? La clave es elegir soluciones validadas e integradas siempre que sea posible. Una solución contrastada e integrada le puede ayudar a evitar errores costosos y a reducir el tiempo en que obtiene información.

También puede resultarle útil para:

- Eliminar conjeturas de diseño al ofrecer un rendimiento predecible que escala.
- Reducir la complejidad de la puesta en marcha al simplificar la adquisición, la instalación y la solución de problemas.
- Agilizar el soporte al eliminar múltiples puntos de contacto.

**Recuerde:** asegúrese de que la solución que elija escale a lo largo de un amplio abanico de rendimientos y capacidades. No se olvide de buscar una que no le obligue a usar diferentes soluciones puntuales en cada fase de la canalización de datos.



## 6. Soluciones de IA inteligentes, potentes y fiables de NetApp

En definitiva, puede ser muy difícil cumplir todos sus objetivos de cara a la IA, a no ser que reciba ayuda con la infraestructura de IA y el software. Por eso, merece la pena trabajar con expertos. No podemos agilizar el flujo del tráfico en su ciudad para que su viaje al trabajo en su supercoche sea más divertido. Sin embargo, somos expertos en flujos de datos y sí podemos ayudarle a desatar todo el poder de sus datos mediante el diseño y la implementación de una estrategia para su infraestructura de IA que sea compatible con la innovación a escala en todo el entorno de multicloud híbrido.

Solo NetApp le permite integrar su Data Fabric en el perímetro, el núcleo y cloud, y le ofrece soluciones inteligentes, potentes y fiables para ayudarle a desatar el potencial de los datos mediante el uso del aprendizaje profundo y el aprendizaje automático. Elija las soluciones de NetApp® para permitir una plataforma apta para el futuro que dé cabida a su paso a la IA condicionada por los datos, desde el análisis predictivo hasta las decisiones autónomas.





# Soluciones de IA inteligentes, potentes y fiables de NetApp

Las empresas están adoptando la IA más rápido que nunca para mantener la competitividad, impulsar el crecimiento y reducir los gastos. Da igual que esté implementando bots de chat, mantenimiento predictivo o medicina genómica, las puestas en marcha de IA efectivas dependen por completo de los datos. Necesita una canalización de datos fluida en las instalaciones, en cloud y en las puestas en marcha de clouds híbridos. Las soluciones de IA de NetApp pueden ayudar.

NetApp le ayuda a crear un Data Fabric a medida para acelerar su camino hacia la IA. Solo NetApp le permite integrar su Data Fabric sin importar dónde esté.

Con las soluciones de NetApp, puede:

- Simplificar la gestión de datos con una tecnología líder del sector que optimiza las puestas en marcha de IA para que, así, sus científicos de datos puedan centrarse en la ciencia y no en la tecnología.
- Integrar sin problemas sus datos en las instalaciones, en cloud y en entornos de cloud híbrido.
- Configurar su infraestructura de IA con integración en Ansible en aproximadamente 20 minutos.
- Transferir **cinco veces más datos por su canalización de datos**, lo que reduce el tiempo que transcurre entre los datos y la información.
- Reducir el tiempo necesario para copiar un conjunto de datos **de días a segundos**.
- Mantener los datos seguros independientemente de dónde se encuentren gracias al conjunto nativo más completo del sector de la IA en cuanto a funciones de seguridad y protección de datos.

Descubra qué puede hacer con las soluciones de IA de NetApp.

Tabla 2) El Data Fabric de NetApp cumple con los requisitos de gestión de datos tanto en el perímetro como en el núcleo y cloud.



## 7. Acelere su paso a la IA

Para obtener más información acerca de la cartera de soluciones de IA de NetApp, que incluye ONTAP® AI de NetApp, visite [netapp.com/ai](https://netapp.com/ai).

Descubra diez buenas razones para elegir las soluciones de IA de NetApp.

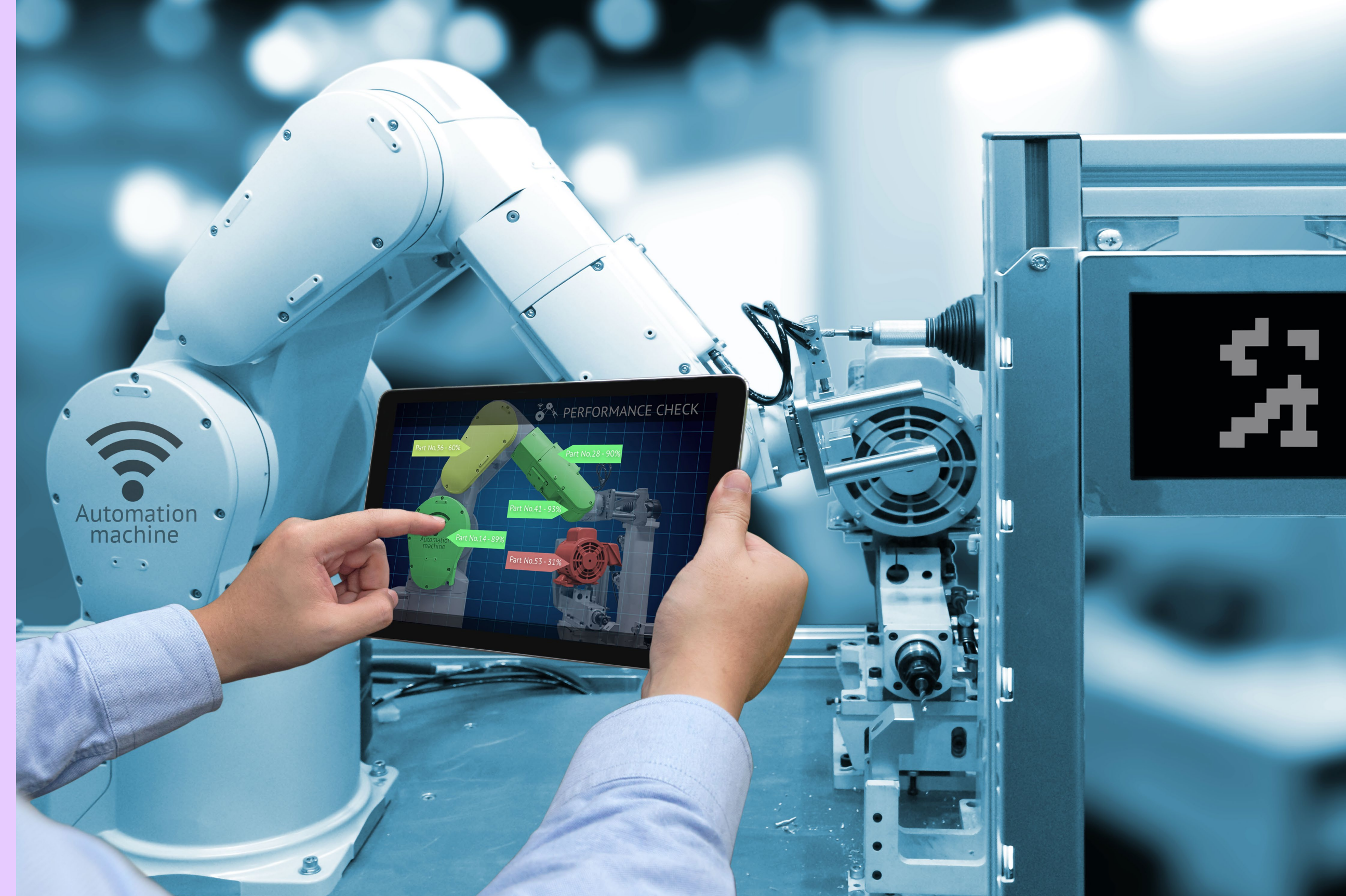
**¿Preguntas?** Póngase en contacto directo con un experto en soluciones de IA de NetApp hoy mismo.>



**Conozca a nuestros expertos.**

<sup>1</sup> NewVantage Partners, encuesta ejecutiva de 2021 sobre Big Data e IA.

<sup>2</sup> IDC, Market Analysis Perspective, Worldwide Artificial Intelligence Software, 2020.



### Acerca de NetApp

En un sector lleno de generalistas, NetApp es un especialista. Nos centramos en una cosa: ayudar a que su empresa aproveche al máximo sus datos. NetApp incorpora a cloud los servicios de datos de clase empresarial en los que confía, y lleva la sencilla flexibilidad de cloud al centro de datos. Nuestras soluciones líderes del sector funcionan en diversos entornos del cliente y en los clouds públicos más grandes del mundo.

Como empresa de software centrado en datos y orientado a cloud, solo NetApp puede ayudar a crear su Data Fabric exclusivo, a simplificar y conectar su cloud, y a proporcionar con seguridad los datos, los servicios y las aplicaciones correctos a las personas adecuadas en cualquier momento y lugar.