

Sus datos **no pueden actuar sobre datos a los que no pueden acceder**





## **El nivel de confiabilidad de la IA depende de los datos que la respaldan.**

Con el auge de la IA generativa y IA agéntica, los datos de su organización se volvieron más importantes que nunca. Con el acceso adecuado a los datos, su IA puede ir más allá de proporcionar respuestas internas para automatizar por completo y escalar de forma confiable los procesos públicos.

Sin embargo, eso significa que hay mucha presión sobre su IA y sobre usted como líder de datos para que sea confiable y correcto, siempre.

Para lograr las mejoras de rendimiento y productividad prometidas por la IA, su organización necesita acceso a datos de alta calidad y confiables. A esto lo llamamos datos listos para la IA. Sin ese cimiento que la sustente, incluso los modelos y agentes más avanzados quedarán muy lejos de ofrecer un valor real. Esa es una de las razones por las que tantas organizaciones lanzan pilotos de IA prometedores que finalmente no logran escalar.

Un primer paso esencial para que sus datos estén listos para la IA es poder acceder a sus datos existentes y unificarlos, a menudo entre entornos en la nube y on-prem.

Los datos empresariales siguen estando en gran medida infrautilizados, con **solo un 1 %** contabilizado en LLM tradicionales.<sup>1</sup> Y **hasta el 90 %** de esos datos generados por la empresa no están estructurados, lo que agrega complejidad y la necesidad de hacer que esos datos estén listos para la IA.<sup>2</sup>

Esto puede representar un gran problema para las iniciativas de IA, porque su IA es tan buena como los datos a los que tiene acceso.



01 →

Por qué los líderes de datos deben empezar con el acceso a los datos

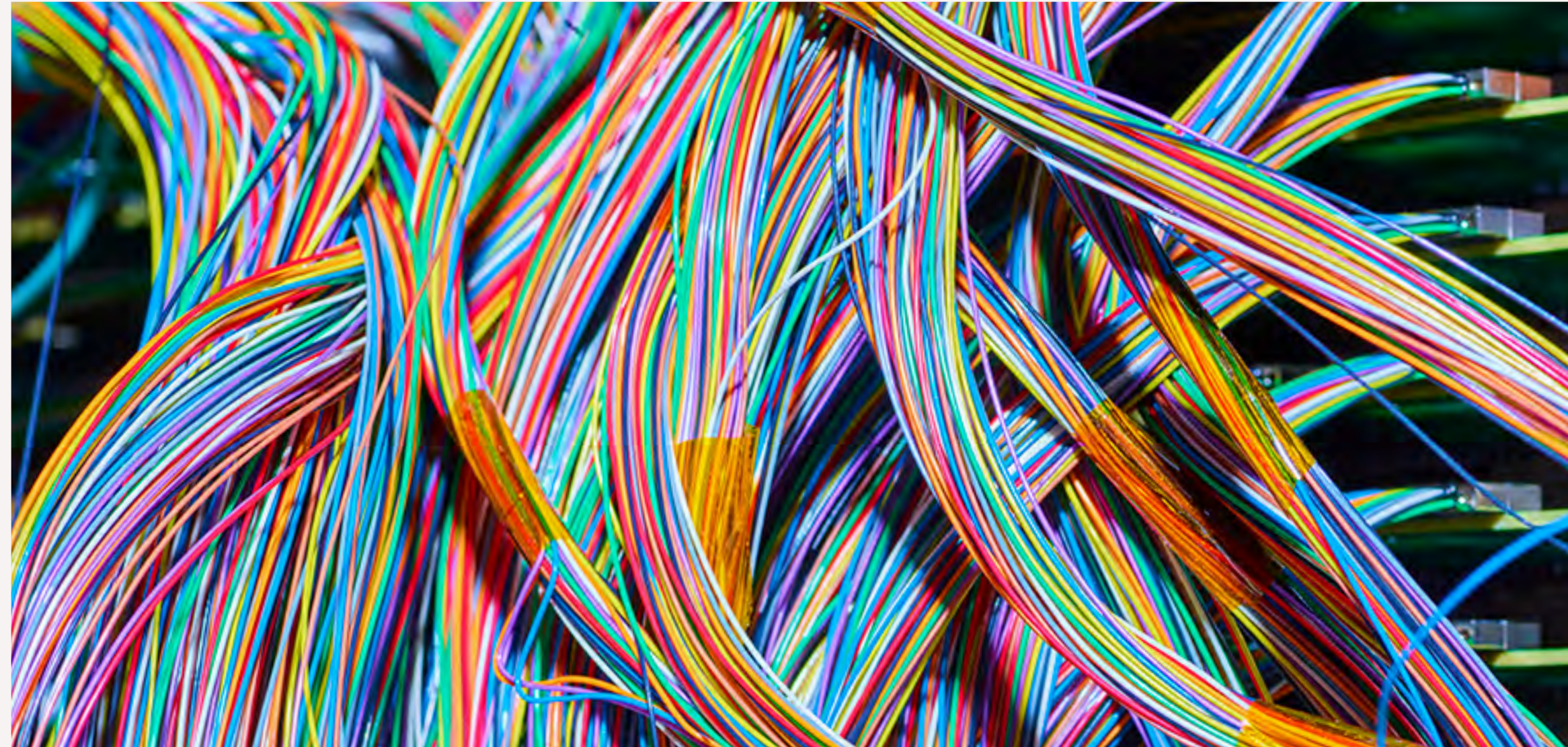
02 →

La complejidad oculta detrás del acceso a los datos

03 →

Dar acceso a la IA a sus datos sin empezar de cero

## Por qué los líderes de datos deben empezar con el acceso a los datos



En la era actual de la IA, muchas organizaciones están intentando crear valor empresarial y obtener retorno de la inversión (ROI) con su IA. Sin embargo, escalar la IA más allá de la fase piloto sigue siendo un reto, a menudo porque los datos no están preparados para la IA.

Para garantizar que los datos empresariales estén preparados para la IA, es esencial un acceso adecuado a los datos. Y eso significa tener acceso y una visión holística de los datos de una organización en cualquier formato, estructurado o no.

Los datos estructurados son el tipo de datos curados que se encuentran en tablas organizadas y etiquetadas, listos para ser extraídos para su uso en IA y otras herramientas y sistemas.



Los datos no estructurados son, esencialmente, todo lo demás. Videos, entradas en el blog, reseñas de clientes, facturas y contratos son solo algunos ejemplos de datos no estructurados. Todos contienen información valiosa que es necesaria para que su IA funcione al máximo.

Los datos no estructurados son mucho más difíciles de acceder que los datos estructurados, lo que significa que las organizaciones a menudo no tienen una visión completa de su escenario y no pueden usar todos sus recursos para alimentar su IA.

La imposibilidad de acceder a sus datos para IA no es un problema en sí mismo, sino que también genera múltiples problemas posteriores. Cuando no tiene una vista completa de sus datos no estructurados, puede duplicarlos de manera involuntaria entre varios repositorios. Esto no solo crea una mala calidad de los datos, sino que también puede impedir que tenga una gobernanza de datos y seguridad unificadas.

Por tanto, crear un acceso unificado, seguro y gobernado a los datos de su organización, y una visión completa de ellos, sin importar el formato, es el primer paso crítico para construir una estrategia sólida en la gestión de datos. Y la gestión de datos es lo que ayuda a permitir resultados de la IA más precisos, eficientes y rentables.

Eso no significa que tenga que migrar todos sus datos o eliminar y reemplazar completamente su infraestructura actual de gestión de datos. La migración de datos es una elección. No necesita mover sus datos para tener una vista unificada, pero puede hacerlo si tiene sentido para su caso de uso.

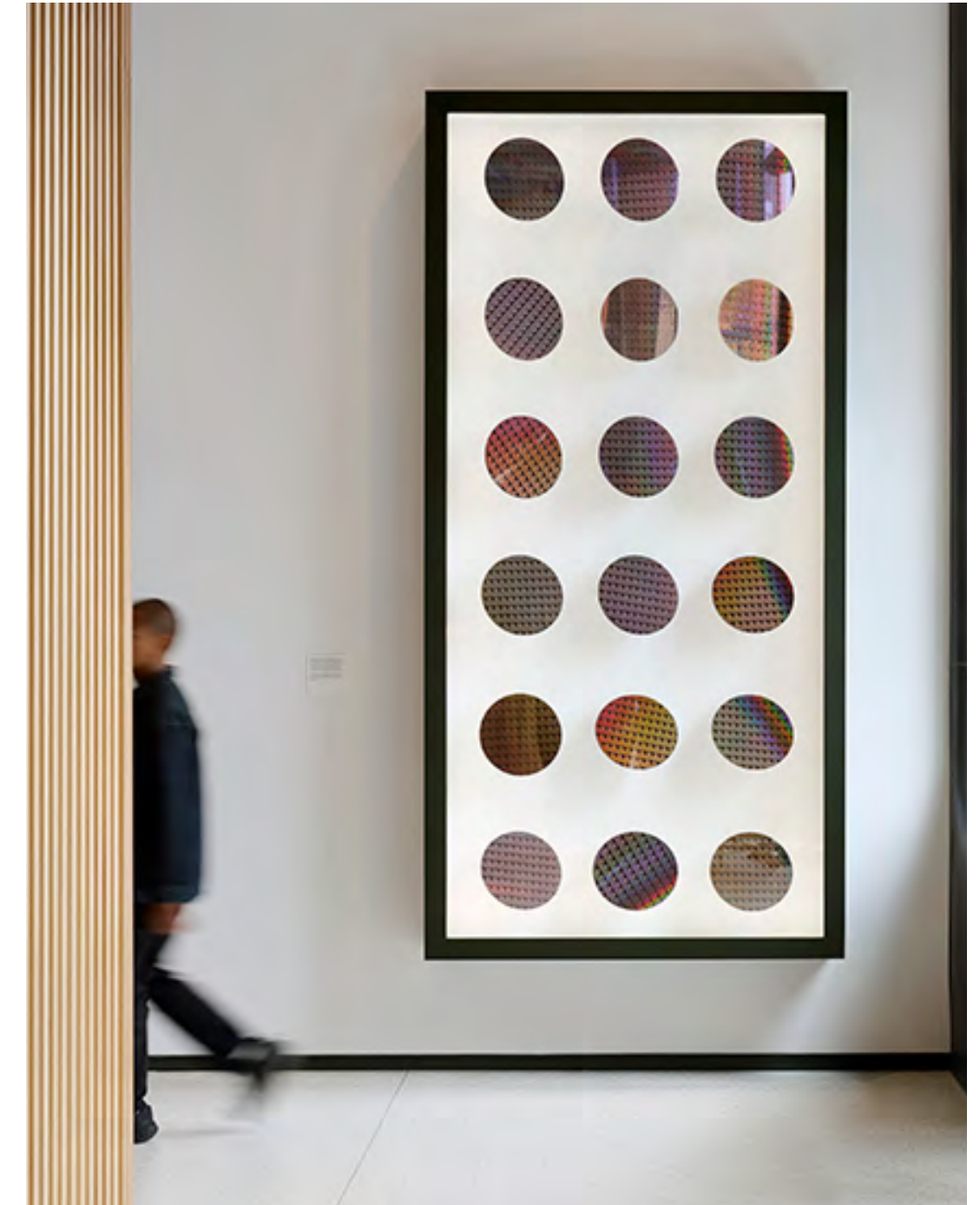
Por qué los líderes de datos deben empezar con el acceso a los datos

Lo importante es encontrar la forma correcta de conectar sus datos en toda su infraestructura. Solo así podrá obtener más valor de sus inversiones de datos existentes usándolos para nuevas cargas de trabajo y resultados impulsados por IA.

No solo está conectando sus datos a la IA; también los está convirtiendo en listos para la IA. Una vez que sus datos estén listos para la IA, puede usar esos mismos datos reiteradamente para muchos casos de uso diferentes.

A cuantos más datos pueda acceder en toda su organización, mayor será la cantidad de insights que obtendrá y sus sistemas de IA más podrán ayudar a impulsar la automatización y la productividad en su empresa.

Por eso es tan importante asegurarse de que su IA pueda acceder a los datos adecuados: datos unificados en todos los formatos, todo su patrimonio de datos y su entorno de nube híbrida. Sin embargo, hay varios obstáculos que impiden alcanzar ese objetivo.



Cuanto más pueda unificar,  
más fácil y rentable será  
gestionar los datos, exponiéndole  
a menos riesgos y dándole más  
agilidad porque tiene menos  
piezas móviles.

**Edward Calvesbert**

Vicepresidente de gestión de productos, watsonx  
IBM



## La complejidad oculta detrás del acceso a los datos



Aunque las empresas enfrentan muchas barreras para acceder a sus datos, el mayor obstáculo suelen ser los silos de datos.

En toda una organización empresarial, los datos se encuentran en múltiples repositorios: desde data lakes hasta bases de datos transaccionales on-premises y repositorios de documentos. Sin embargo, sin una visión centralizada entre los repositorios de datos, las organizaciones a menudo tienen dificultades con registros duplicados, visibilidad limitada y lagunas de comprensión, dejando las iniciativas de IA poco potenciadas desde el principio.

A estos silos se suma el hecho de que los datos están inherentemente bloqueados, de manera predeterminada, en bases de datos seguras u otros repositorios.

Dado que la seguridad es una preocupación comprensible para la mayoría de las organizaciones, el acceso a esos repositorios suele ser limitado. Además, dado que los datos son tan heterogéneos hoy en día, es más probable que acabe con múltiples silos de datos.

En el clima digital-first actual, el volumen, la variedad y la complejidad de los datos son un problema en sí mismo. Los datos se producen en todas partes, todo el tiempo, en cantidades tan grandes que gestionarlos se vuelve difícil.

Sin embargo, volúmenes tan grandes de datos pueden generar insights valiosos, lo que hace que una gestión eficaz sea aún más importante.

Aunque los datos no estructurados comparten muchos de los mismos desafíos que los datos estructurados, incluidos los relacionados con los silos de datos, la seguridad y la calidad también se encuentran literalmente en toda la infraestructura de una organización.

Los desafíos únicos de los datos no estructurados incluyen:

**Correlación:** ¿cómo conecta y obtiene valor de los datos no estructurados y los datos estructurados?

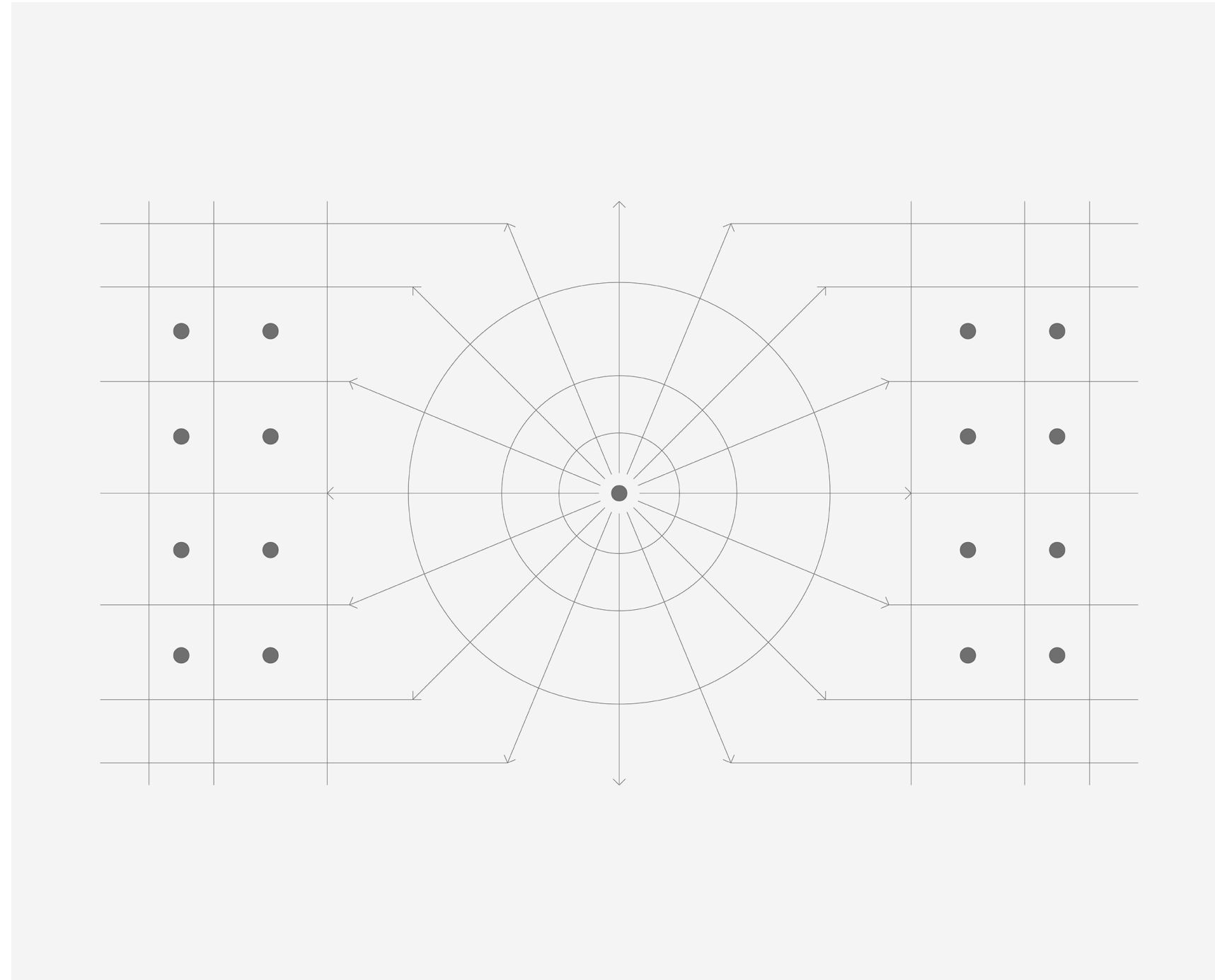
**Precisión:** ¿cómo se asegura que sus resultados de la IA sean correctos y sin sesgo?

**Escalabilidad:** ¿cómo escala la preparación de datos y la gobernanza no estructuradas?

El enfoque convencional para permitir que la IA acceda a datos empresariales es la generación aumentada por recuperación (RAG, por sus siglas en inglés) basada en vectores.

Esta arquitectura optimiza el rendimiento de los modelos de IA conectándolos con bases de conocimiento externas. La RAG combina la recuperación de información y la generación de texto recuperando documentos o fragmentos de información relevantes de grandes conjuntos de datos para generar respuestas contextualmente relevantes.

La RAG ayuda a los LLM a ofrecer respuestas más relevantes sustentadas en bases de conocimiento externas, como datos internos de la organización, revistas académicas y conjuntos de datos especializados. Destaca en la búsqueda semántica y en encontrar rápidamente la información necesaria.



Eso fue lo que dio lugar a que la RAG fuera el enfoque de referencia para acceder a los datos, pero en la era de la IA generativa y la IA agéntica, la RAG convencional empieza a mostrar sus limitaciones.

La RAG convencional es principalmente interna y orientado a los empleados. Admite tipos limitados de cargas de trabajo. Se centra principalmente en tareas informativas, como búsqueda semántica, preguntas frecuentes de documentos y resúmenes.

La RAG convencional se basa únicamente en la incorporación vectorizada para la recuperación de datos no estructurados y no respeta las licencias a nivel de fuente. Por lo tanto, le cuesta normalizar y estructurar esos datos. No puede combinar eficazmente los múltiples tipos de datos necesarios para potenciar la IA actual. También puede ser un proceso arriesgado debido a la gobernanza de datos fragmentada.

Como resultado, las soluciones de RAG pueden ofrecer una precisión limitada y pueden pasar por alto los insights más profundos que se esconden en el contenido empresarial.

Para superar esta limitación y crear agentes de IA y aplicaciones que sean operativos, analíticos y orientados hacia el exterior, las organizaciones deben superar tres desafíos clave:

- 🔗 Los pipelines manuales y frágiles para la ingesta y recuperación de datos
- 🔗 Baja precisión en el manejo de semánticas complejas
- 🔗 Gobernanza incongruente, especialmente para contenido sensible y control de acceso

Afortunadamente, las nuevas soluciones en evolución para acceder a los datos pueden cerrar la brecha entre datos estructurados y no estructurados. Herramientas como la integración inteligente de datos, una arquitectura de lakehouse de datos, enriquecimiento, gobernanza y recuperación híbrida de datos no estructurados y preparados para la IA pueden ayudar a las organizaciones a sacar realmente el máximo partido a sus datos y a generar resultados más confiables de la IA.



## Dar acceso a la IA a los datos correctos sin empezar de cero



Con la reciente proliferación de la IA generativa e IA agéntica, algunas organizaciones se están apresurando a iniciar nuevas iniciativas que crean nuevos silos de datos. Sin embargo, una estrategia de gestión de datos bien pensada puede ayudar a las organizaciones a evitar los problemas que crean estos silos. Una estrategia así puede ayudarle a acceder y unificar los datos de manera gobernada, segura y estratégica, preparándole así para obtener mejores resultados a largo plazo. Además, a corto plazo, acelerará sus iniciativas de IA generativa.

# 01

Varios principios rectores sirven como base para hacer que sus datos sean accesibles y estén listos para la IA:

### Gestión de datos por diseño

Cree una estrategia de gestión de datos con una visión clara de hacia dónde se dirige. Este enfoque ayudará a alinear todas las pequeñas decisiones graduales que tome en el camino.

# 02

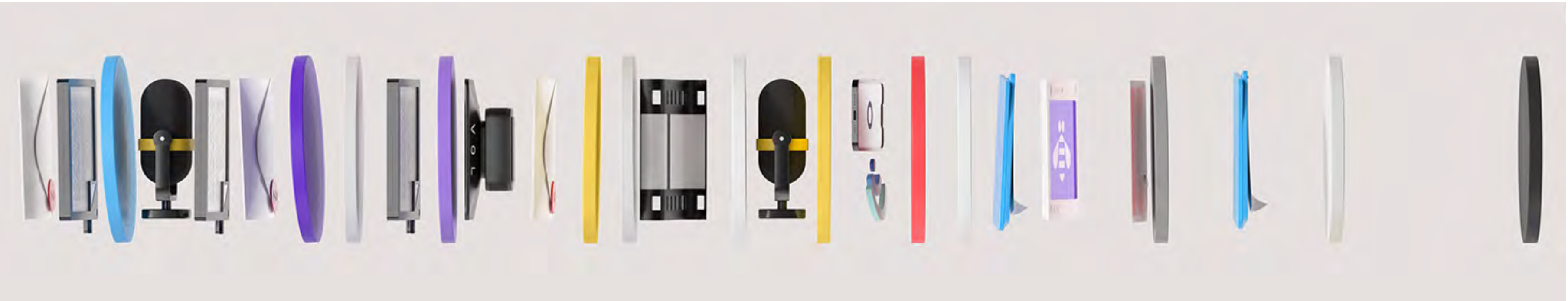
### Maximice el valor de sus inversiones en IA

No reemplace lo que ya construyó; ajústelo. Incorpore una arquitectura de datos que proteja las inversiones e integre nuevas capacidades.

# 03

### Acceso implementado

En lugar de mover grandes cantidades de datos a un repositorio central, cree una mayor interoperabilidad dentro de su patrimonio de datos actual.



# 04

## Flexibilidad

Asegúrese de que, de ahora en adelante, todas sus inversiones puedan adaptarse a situaciones variables con una arquitectura de datos flexible.

# 05

## Preparación operativa

Mantenga la participación de humanos en el proceso como parte de su estrategia de datos, para revisar, refinar y desplegar resultados de la IA de manera responsable.

# 06

## La gobernanza como necesidad

Proporcione a su agente de IA, esencialmente nuevos usuarios que acceden a LLM, un significado semántico real mediante una gobernanza de datos y calidad sólidas.

Con estos principios en mente, puede emprender varias iniciativas para dar a su organización acceso unificado a los datos de su empresa.



## Agregue estructura a los datos no estructurados

Un enfoque para orientar las limitaciones de la RAG convencional comienza con la conectividad y el acceso a fuentes de datos estructuradas y no estructuradas. Necesita poder acceder a sus datos donde residen y crear derivados estructurados a partir de estos.

Al automatizar y escalar la creación de esos derivados estructurados, entrega sus datos a sus aplicación de IA con mayor contexto, permitiendo resultados de la IA más confiables y precisos.

Un componente importante de este enfoque es la gobernanza de datos y la autorización del usuario. Las normas de protección de datos, incluidos los controles de acceso

y las restricciones de información, deben aplicarse a datos sensibles, y los agentes de IA deben estar restringido a la información que sea apropiada y necesaria para su tarea asignada. Sus datos no estructurados deben ser más sofisticados en cuanto a quién tiene acceso a qué datos y por qué motivo. Estas reglas de acceso también deben propagarse de manera descendente hacia los distintos derivados de esos datos.

Con estas medidas de seguridad, debe crear una gestión unificada tanto de datos estructurados como no estructurados. De este modo, puede mejorar la precisión de los resultados y desbloquear sus datos tanto para la IA como para los analytics tradicionales.



## Lakehouse de datos y tejido de datos extendido

Una excelente manera de lograr un acceso unificado a los datos es mediante una arquitectura de lakehouse de datos con capacidades ampliadas de tejido de datos. Este enfoque combina las características de acceso y de optimización de una arquitectura de lakehouse de datos con las fortalezas de gestión y gobernanza de un tejido de datos. También respeta el control de acceso desde el sistema fuente de su documento hasta sus resultados de la IA. Este enfoque puede ayudar a garantizar que los datos adecuados lleguen a los usuarios adecuados, y solo a ellos, incluidos los agentes de IA que son consumidores de datos.

Con este tipo de gestión unificada de datos, puede acabar con los silos de datos y eliminar los riesgos que conlleva la variedad y complejidad de los datos en esos silos. Cuanto más acceso unificado tenga a los datos, más fácil y rentable será gestionar sus datos, y más riesgos podrá mitigar. Con acceso unificado a los datos, puede desplegar diferentes herramientas y admitir múltiples cargas de trabajo con los mismos datos, permitiendo economías de escala.

Además, una arquitectura de lakehouse de datos ofrece la oportunidad de optimizar las cargas de trabajo en función de costos y rendimiento, lo cual ofrece beneficios en varios aspectos:

- Una cobertura más amplia de las cargas de trabajo para hacer más con una sola plataforma
- Elección de herramientas para diferentes prototipos con distintas habilidades
- Mejor retorno de la inversión (ROI) para las cargas de trabajo existentes y un patrimonio de datos optimizado para IA generativa

## Soluciones de nube híbrida

La mayoría de las organizaciones no están 100 % basadas en la nube; hay algunos datos que simplemente no pueden residir en el nube. Su patrimonio de datos es un híbrido de on-prem y múltiples nubes públicas. Por eso, acceder a todos estos datos dondequiera que residan a menudo requiere vincular soluciones de almacenamiento en la nube y on-prem. Un enfoque solo en la nube no será suficiente para que la mayoría de las organizaciones accedan a todos sus datos.

Por eso es importante planear el acceso y el despliegue de datos desde una perspectiva de nube híbrida. Las cargas de trabajo de datos pertenecen donde realmente residen los datos. Sin embargo, una estrategia de centralización de datos puede generar trabajo innecesario y posibles riesgos futuros. En su lugar, aporta analytics, integración e inteligencia a sus datos, no al revés.

## Un enfoque abierto



Adoptar un enfoque abierto en su estrategia de gestión de datos es vital para garantizar el acceso a los datos y el éxito de la IA. ¿Por qué? Porque el código abierto, incluyendo formatos de datos abiertos, formatos de archivo, formatos de tablas, catálogos y motores de consulta, es fundamental para el acceso completo a los datos y la interoperabilidad entre repositorios e inversiones. No solo ayuda a reducir el vendor lock-in (dependencia de proveedores), sino que también proporciona más transparencia sobre la tecnología que se está usando.

Los formatos de archivo abiertos son especialmente importantes porque permiten a la aplicación acceder a los datos sin importar dónde residan. Así, su empresa ya no está ligada a ninguna base de datos específica y tiene la flexibilidad de mover aplicaciones sin problema.

La apertura, en última instancia, es la forma de lograr acceso unificado a los datos sin necesidad de migrar todos sus datos. Sin formatos abiertos, está atrapado en una tecnología propia, lo que limita lo que puede hacer y usar en su ecosistema de datos en general. A través de la apertura, puede aprovechar el potencial de sus inversiones en datos hoy mismo, visualizando y accediendo a sus datos a través de una única capa unificada para la IA.

La promesa de democratizar verdaderamente el acceso a los datos está a la vuelta de la esquina, por lo que todo el mundo va a estar basado en datos.

**Fariya Syed-Ali**

Responsable de marketing de producto, watsonx  
IBM



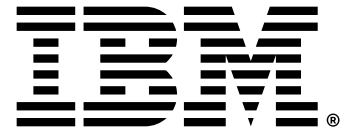
Dar acceso a la IA a los datos correctos sin empezar de cero

## Solutions

La cartera de datos de IBM puede ayudarle a construir y mantener el acceso para que los datos funcionen para usted y sus proyectos de IA.

[Explore ahora estas soluciones para aprender más.](#)





1. [The future of AI is open](#). IBM. 23 de mayo de 2024.
2. [Agentic AI has an unstructured data problem: IBM is unveiling a solution](#). Edward Calvesbert. IBM, 2025.

© Copyright IBM Corporation 2025

IBM, el logotipo de IBM, IBM Consulting y watsonx.data son marcas comerciales o marcas comerciales registradas de International Business Machines Corporation, en Estados Unidos o en otros países. Es posible que otros nombres de productos y servicios sean marcas registradas de IBM o de otras empresas. Una lista actualizada de las marcas de IBM está disponible en [ibm.com/mx-es/legal/copytrade](https://ibm.com/mx-es/legal/copytrade).

Este documento está actualizado a la fecha inicial de publicación e IBM puede modificarlo en cualquier momento.

No todas las soluciones están disponibles en todos los países en los que opera IBM.

LA INFORMACIÓN INCLUIDA EN ESTE DOCUMENTO SE PROPORCIONA “TAL CUAL” SIN NINGUNA GARANTÍA, EXPRESA O IMPLÍCITA, INCLUSO SIN NINGUNA GARANTÍA DE COMERCIALIZACIÓN, IDONEIDAD PARA UN PROPÓSITO PARTICULAR NI GARANTÍA O CONDICIÓN DE NO INFRACCIÓN. Los productos de IBM están amparados de conformidad con los términos y condiciones de los acuerdos en virtud de los que se proveen.

Ningún sistema o producto de TI debe considerarse completamente seguro y ningún producto, servicio o medida de seguridad por sí solo puede ser completamente eficaz para prevenir el uso o acceso no autorizado. IBM no garantiza que ningún sistema, producto o servicio sea inmune o haga que su empresa sea inmune a una conducta maliciosa o ilegal de cualquier parte.

El cliente es responsable de garantizar el cumplimiento de las leyes y regulaciones aplicables. IBM no brinda asesoría legal ni declara o garantiza que sus servicios o productos aseguren que el cliente cumpla con cualquier ley o reglamento.