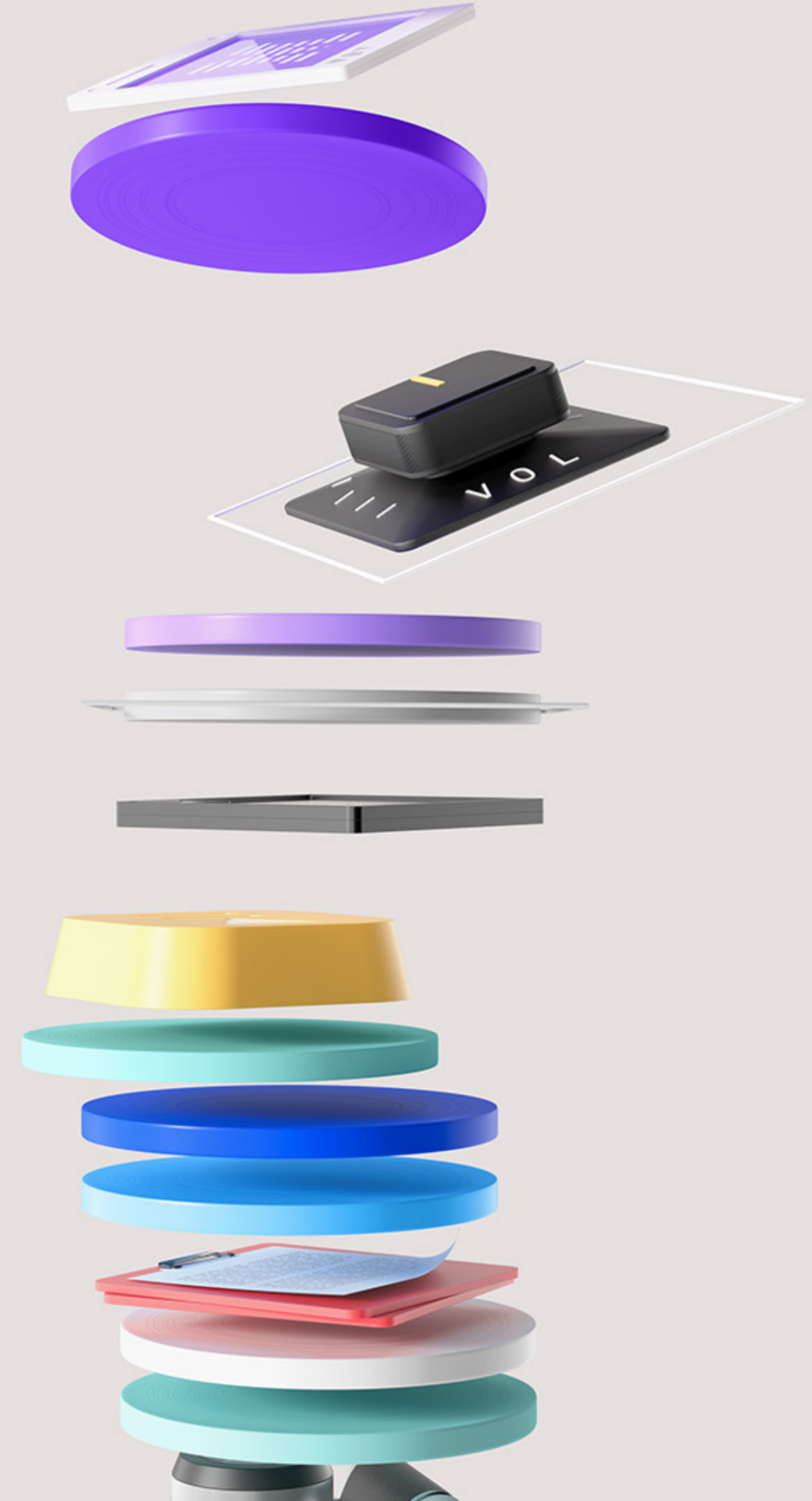


watsonx.data

El único lakehouse  
de datos híbrido y  
abierto para analytics  
e IA empresarial



# Índice

01 →

Por qué su empresa  
necesita un data  
lakehouse preparado  
para la IA generativa

02 →

Retos actuales y nuevos  
en materia de datos

03 →

Definición de lakehouse  
de datos

04 →

Conozca watsonx.data

05 →

¿Por qué elegir  
watsonx.data?

06 →

Próximos pasos



# Por qué su empresa necesita un lakehouse de datos preparado para la IA generativa



Los datos genéricos producen resultados genéricos. Si no tiene datos específicos de su empresa, los resultados tampoco lo serán. Si todo el mundo usa los mismos datos públicos para entrenar la misma IA generativa, ¿cómo puede usted superar a sus competidores?

Su ventaja competitiva radica en sus datos únicos. Los datos no estructurados (el tipo de datos que se encuentran ocultos en correos electrónicos, archivos PDF, imágenes y videos) son especialmente valiosos, pero muy difíciles de usar en la práctica. Las organizaciones necesitan una arquitectura de datos inteligente para acceder, preparar y entregar datos estructurados y no estructurados y así poder desbloquear todo el potencial de sus datos empresariales para la IA.

Los datos son el combustible de la IA, sin embargo, se estima que los modelos de IA están aprovechando menos del 1 % de los datos empresariales.<sup>1</sup> Y el 90 % de los datos empresariales no están estructurados.<sup>2</sup>

Un lakehouse de datos híbrido y abierto puede ayudarle a potenciar su IA e infundir sus datos en una amplia variedad de casos de uso modernos. Este unifica todos los tipos de datos, simplifica la integración y hace que la gestión y la gobernanza de los datos, tanto estructurados como no estructurados, sean más fáciles que nunca. No es de extrañar que se esté convirtiendo en la arquitectura de datos a largo plazo preferida para las cargas de trabajo de analytics y las nuevas cargas de trabajo de IA generativa.



90%

de los datos empresariales  
se estima que no están  
estructurados.<sup>2</sup>





## Retos actuales y nuevos en materia de datos

A medida que ha aumentado la importancia de los datos empresariales, también lo han hecho los retos asociados a estos. El gran volumen de datos se ha vuelto abrumador y, a menudo, se encuentran en silos por toda la organización. Además, las nuevas variedades de datos han complicado la integración y los datos de baja calidad han puesto en riesgo el impacto en la IA. La IA generativa puede ayudar a resolver estos problemas, pero una arquitectura de datos resiliente y flexible es esencial para ello.

Hoy en día, las limitaciones de la generación aumentada por recuperación (RAG) impiden que las empresas se den cuenta del valor de los datos no estructurados para la IA generativa. Entonces, ¿cómo puede llevar los datos no estructurados a las cargas de trabajo de IA generativa y tradicional de analytics? Mediante la adopción de un lakehouse de datos abierto e híbrido.





## Definición de lakehouse de datos



Los datos empresariales son fundamentales para cualquier arquitectura de IA. Las organizaciones que busquen herramientas de IA generativa en 2025 deberán asegurarse de que sus datos estén listos para la IA. Una solución de lakehouse proporciona una base moderna para escalar los datos listos para IA.

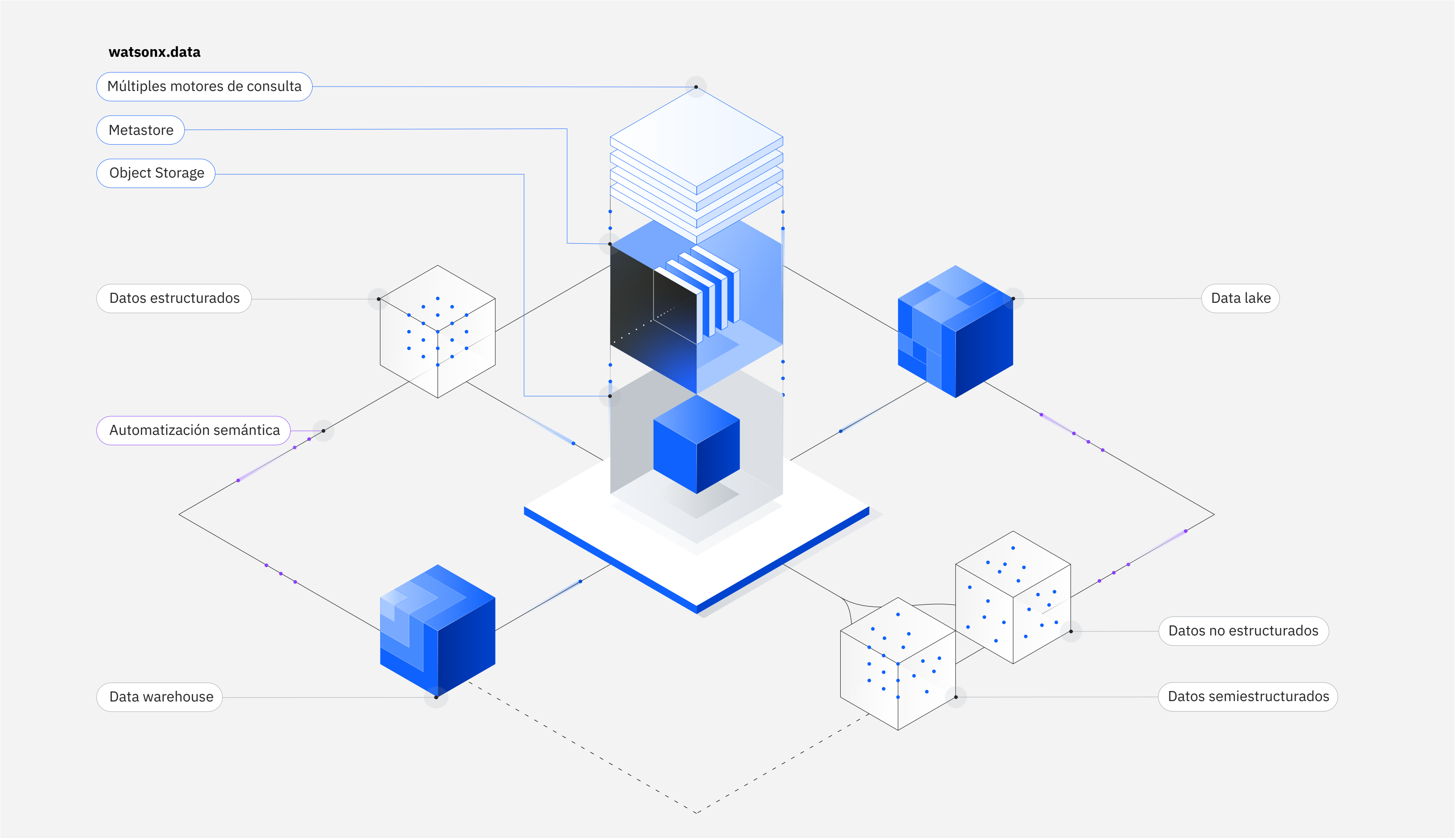
El lakehouse de datos es una arquitectura emergente que ofrece la flexibilidad de un lake de datos con el rendimiento y la estructura de un almacén de datos. La mayoría de las soluciones de lakehouse ofrecen un motor de consulta de alto rendimiento sobre almacenamiento de bajo costo junto con una capa de gobernanza de metadatos. Las capas de metadatos inteligentes facilitan a los usuarios la categorización y clasificación de datos no estructurados, como video y voz, y semiestructurados, como XML, JSON y correos electrónicos.

El mejor lakehouse de datos ofrecerá tecnologías de código abierto que reducen la duplicación de datos y simplifican los pipelines complejos. Sin embargo, algunos lakehouses de primera generación tienen restricciones de extracción, transformación y carga (ETL) que limitan su capacidad para resolver los desafíos de costo y complejidad. Por ejemplo, un motor de consulta único diseñado para cargas de trabajo de business intelligence (BI) o machine learning podría ser ineficaz cuando se usa para otro tipo de carga de trabajo.

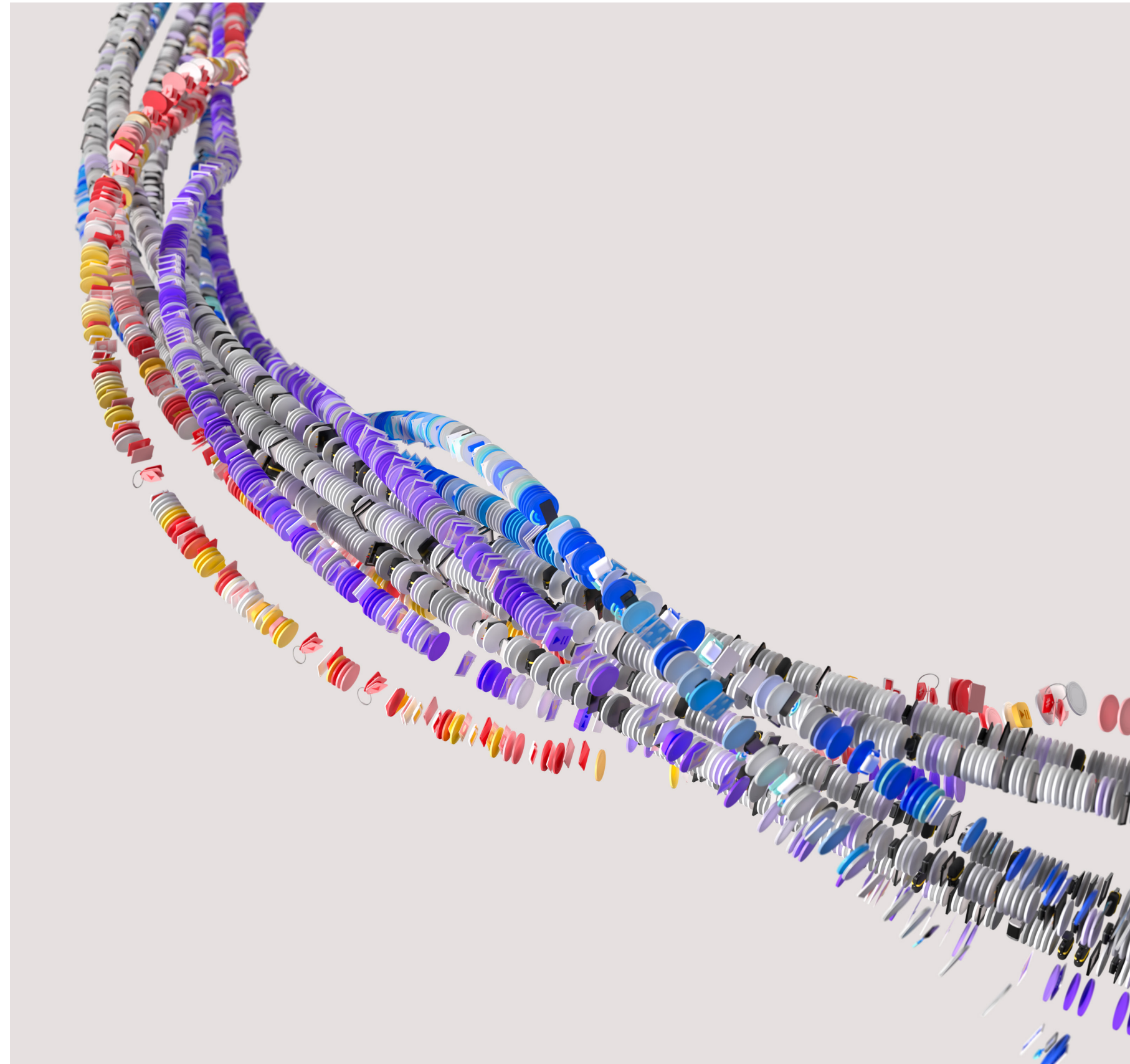
Los equipos de datos e IA de IBM creen que cada carga de trabajo es única y debe optimizarse con el entorno más adecuado que mantenga el costo al mínimo y el rendimiento al máximo. En última instancia, convertir los datos sin procesar en insights aplicables en la práctica a escala requiere un lakehouse que ofrezca un rendimiento rentable y flexibilidad de formato.



Gestione todo el ciclo de vida de los datos para la IA



## Conozca watsonx.data



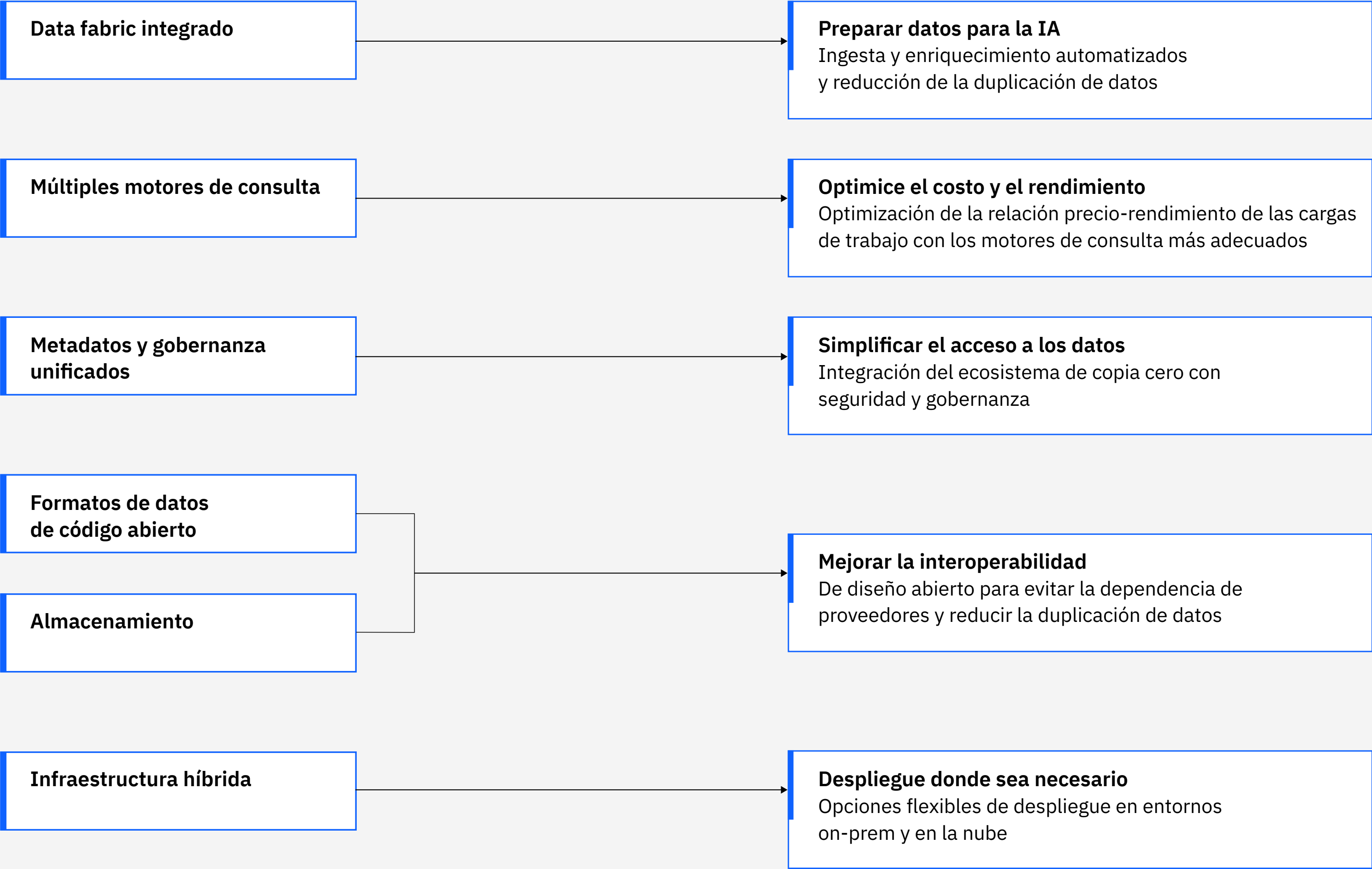
IBM® watsonx.data® es el único lakehouse de datos híbrido y abierto para analytics e IA empresarial.

Ayuda a gestionar todo el ciclo de vida de los datos empresariales para IA dentro de su lakehouse de datos para potenciar las aplicaciones de IA generativa de siguiente generación y BI infundida con IA.

Con watsonx.data, puede simplificar y escalar la integración, la gestión y la gobernanza de datos estructurados y no estructurados en entornos on-premises, en la nube y multinube. Es parte de una pila de datos abierta y moderna que emplea tecnología innovadora de código abierto que se integra con su patrimonio de datos existente, sin vendor lock-in (dependencia de proveedores).



IBM watsonx.data integra perfectamente las capacidades de un lakehouse y tejido de datos.



¿Cuáles son los componentes de watsonx.data?

Una arquitectura de lakehouse de datos abierta e híbrida proporciona una plataforma unificada para resolver los desafíos clave de datos que enfrentan las empresas en la actualidad: grandes volúmenes de datos, múltiples ubicaciones y formatos, calidad de datos y gobernanza.

A continuación, descubra cómo IBM watsonx.data aborda estos desafíos.

**Infraestructura**

La infraestructura es donde se desplegará su lakehouse totalmente gestionado a través de cualquier entorno en la nube u on-premises.

**Almacenamiento**

La capa de almacenamiento es donde los datos se almacenan físicamente como archivos o en formatos de datos abiertos, como Apache Parquet y Avro. Los formatos de datos abiertos son especificaciones de archivos y protocolos que se ponen a disposición de la comunidad de código abierto para que cualquiera pueda ingerirlos y mejorarlos.

**Motores de consulta adecuados al propósito**

Los motores de consulta múltiples y adecuados para su propósito permiten ejecutar consultas de forma eficaz para grandes conjuntos de datos. Con la optimización de la carga de trabajo en varios motores de consulta y niveles de almacenamiento, puede minimizar el costo de su almacén de datos, al tiempo que proporciona un procesamiento de datos a gran escala rápido, confiable y eficaz.

**Formatos de tabla abiertas**

Los formatos de tablas abiertas, como Apache Iceberg, ayudan a proporcionar estructura y ofrecen la confiabilidad y simplicidad de SQL con big data. Estos formatos permiten que distintos motores accedan a los mismos datos al mismo tiempo, lo que ayuda a evitar el vendor lock-in (dependencia de proveedores). Los datos también se pueden compartir entre múltiples herramientas y repositorios de datos, como su almacén de datos. Una copia única de los datos le permite reducir la duplicación de datos y eliminar los silos.

**Almacén de metadatos**

El almacenamiento de metadatos en watsonx.data se basa en Hive Metastore de código abierto, que ayuda a garantizar la interoperabilidad y permite que herramientas externas accedan fácilmente a sus metadatos.



**Gobernanza**

Los metadatos también se almacenan en formatos de tabla abierta. Sirve para definir los formatos de archivos para cualquier herramienta que pueda leer o escribir formatos de datos abiertos.

**Servicio técnico de metadatos**

El componente de servicio de metadatos técnicos es necesario para comprender qué datos están disponibles en la capa de almacenamiento. El motor de consulta usa metadatos para proporcionar información completa sobre el linaje para que usted sepa dónde se encuentran sus datos, cómo se ven y cómo leerlos.

**Catálogos de datos**

Los catálogos de datos ayudan a los usuarios a encontrar los datos correctos para el trabajo y brindan información semántica para políticas y reglas. Los metadatos empresariales, como las terminologías y las etiquetas empresariales, se almacenan para respaldar la búsqueda y la protección de datos.

**Motor de políticas**

El motor de políticas permite a los usuarios definir políticas de protección de datos y al motor, aplicarlas. Para crear una infraestructura de gobernanza que sea escalable, a menudo se despliega un motor de políticas con el servicio de metadatos técnicos y el catálogo de datos.

**Motor de consultas**

El motor de consultas, a menudo conocido como componente de cálculo, es el corazón del lakehouse de datos abiertos. Puede ser de código abierto o propio, y accede a los datos en formatos de tabla abierta.

En una arquitectura de lakehouse abierto, el motor de consulta es totalmente modular, lo que significa que el motor puede escalar dinámicamente para satisfacer las demandas de carga de trabajo y concurrencia. Los motores de consulta también pueden anexarse a cualquier catálogo y sistema de almacenamiento.

¿Qué distingue a IBM watsonx.data del resto?

**Híbrido y abierto**

Watsonx.data está desarrollado con tecnología fundacional de código abierto para la innovación y el apoyo a la comunidad, sin vendor lock-in (dependencia de proveedores). Puede acceder a los datos y desplegarlos en cualquier lugar en entornos de nube híbrida, in-client VPC, entornos multinube y on-premises, así como SaaS.

**Optimizado para la carga de trabajo**

Con el apoyo de capacidades integradas de data fabric, que incluyen gobernanza de datos, calidad y linaje, watsonx.data le ayuda a gestionar volúmenes de datos crecientes a través de una arquitectura única de múltiples motores optimizada para el precio-rendimiento.

**Listo para la IA generativa**

Watsonx.data permite la autorización coherente del usuario de extremo a extremo desde los sistemas de origen de documentos hasta la recuperación.

**¿Cómo funciona todo en conjunto?**

Watsonx.data permite a las organizaciones desbloquear la información crítica para el negocio en sus datos empresariales no estructurados para obtener IA generativa y analytics más precisas y potentes.

40%

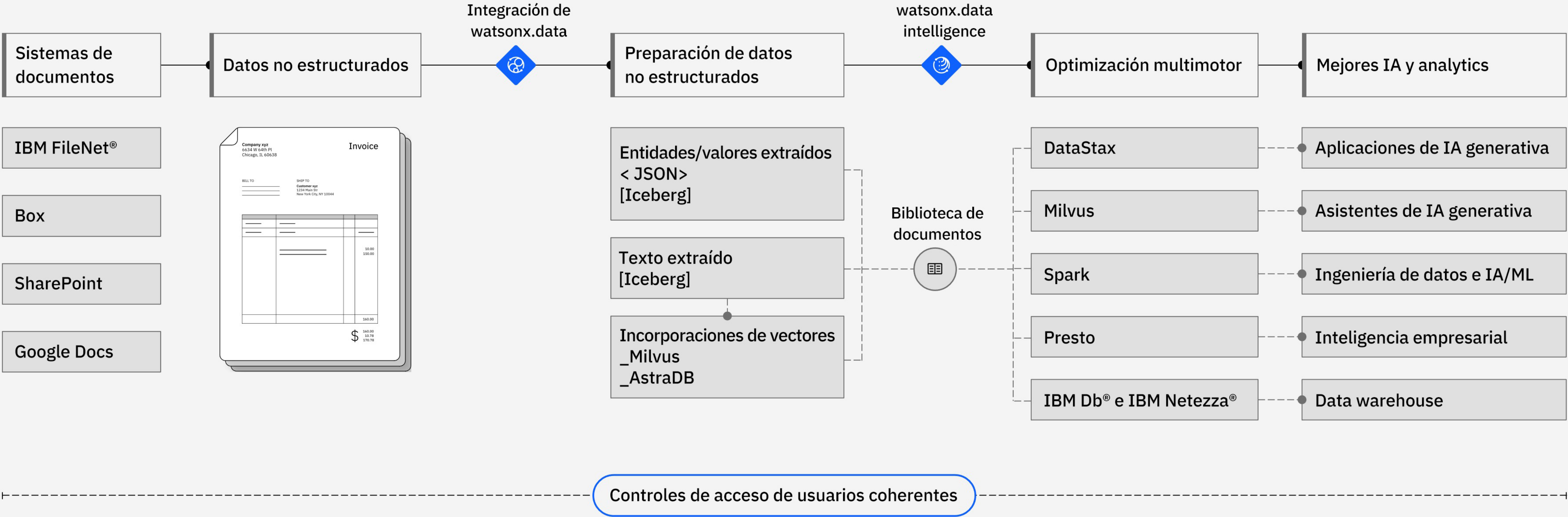
IA más precisa que la RAG convencional con watsonx.data<sup>3</sup>








Ingesta a escala

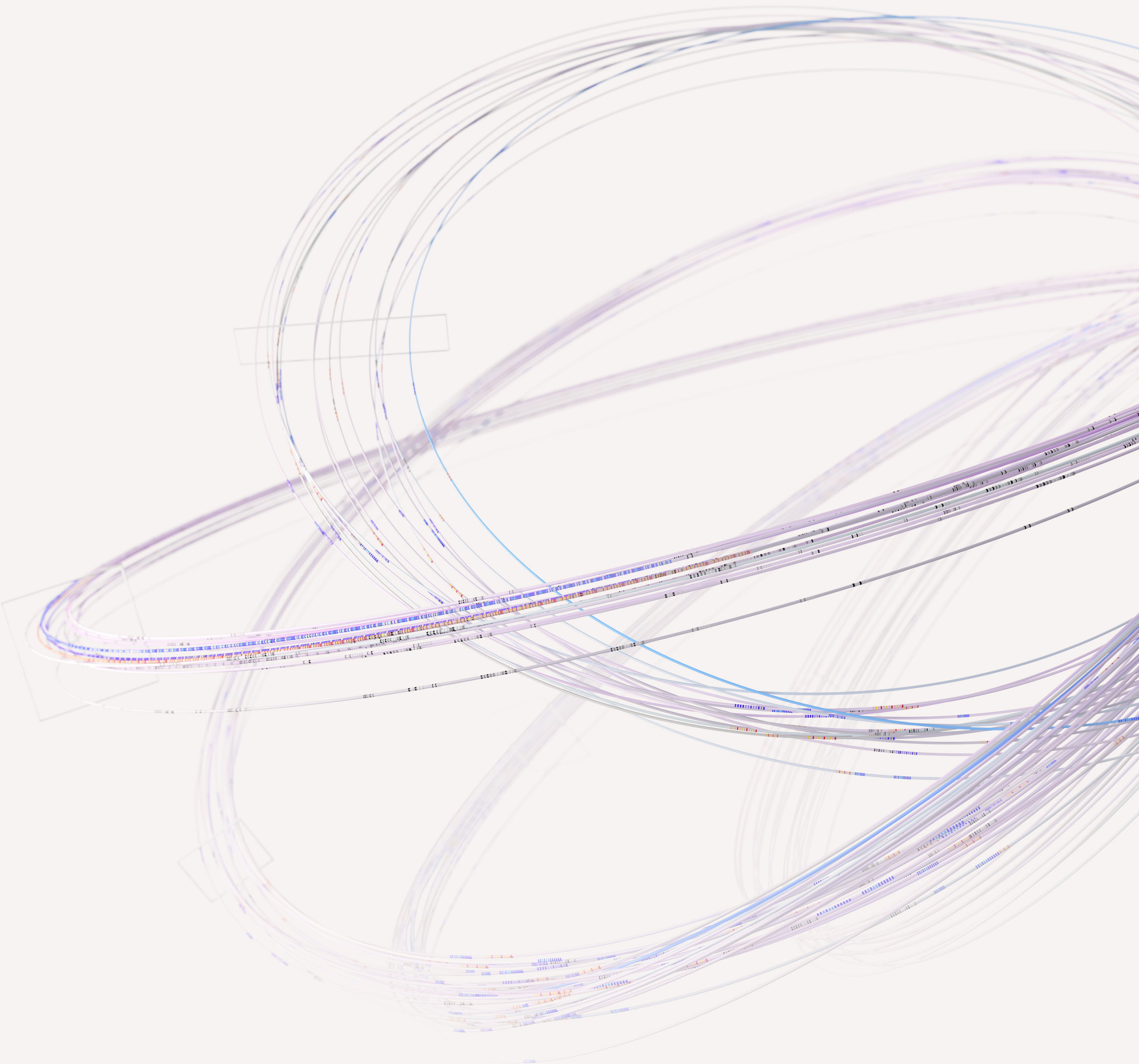
Curar y gobernar

Entrega a la IA



# ¿Por qué elegir watsonx.data?

-  Para unificar los datos estructurados y no estructurados a fin de mejorar la precisión de la IA y el retorno de la inversión (ROI).
-  Para desbloquear datos sin explotar para analytics conversacionales.
-  Para simplificar y escalar la integración, la gestión y la gobernanza de datos estructurados y no estructurados.
-  Para proporcionar datos confiables y de alta calidad a todos los usuarios de los datos.
-  Para proteger sus inversiones en datos existentes con una pila de datos abierta y moderna.





# Próximos pasos

Benefíciase de los conocimientos de IBM sobre gestión y optimización de datos, perfeccionados durante décadas de manejo de las cargas de trabajo de datos más exigentes del mundo. Vea lo rápido que puede maximizar el retorno de la inversión (ROI) de sus datos con watsonx.data.

Empiece su prueba gratuita →

Solicite una demo →







- 1. [The Future of AI is Open, IBM, 2024](#)
- 2. [IDC, UNTAPPED VALUE: What Every Executive Needs to Know About Unstructured Data, 2023](#)
- 3. Basado en pruebas internas que comparan la corrección de las respuestas de los resultados del modelo de IA utilizando la capa de recuperación de watsonx.data Premium Edition con RAG solo vectorial en tres casos de uso comunes con conjuntos de datos de IBM utilizando el mismo conjunto de modelos de inferencia, evaluación e incorporación de código abierto seleccionados y variables adicionales. Los resultados pueden variar.

© Copyright IBM Corporation 2025

IBM, el logotipo de IBM, Db2, FileNet, Netezza, y watsonx.data son marcas comerciales o marcas registradas de International Business Machines Corporation, en los Estados Unidos o en otros países. Otros nombres de productos y servicios pueden ser marcas registradas de IBM o de otras empresas. Encontrará una lista actualizada de las marcas registradas de IBM 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.

Es responsabilidad del usuario verificar el funcionamiento de cualesquier productos o programas ajenos a IBM con productos y programas de IBM. IBM no se hace responsable por productos y programas ajenos a IBM

LA INFORMACIÓN INCLUIDA EN ESTE DOCUMENTO SE PROPORCIONA “TAL CUAL” SIN NINGUNA GARANTÍA, EXPRESA O IMPLÍCITA, INCLUSO SIN NINGUNA GARANTÍA DE COMERCIABILIDAD, 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 un sistema, producto o servicio sea inmune o hará que la empresa de usted 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 que sus servicios o productos garantizarán que el cliente cumpla con cualquier ley o reglamento en particular.