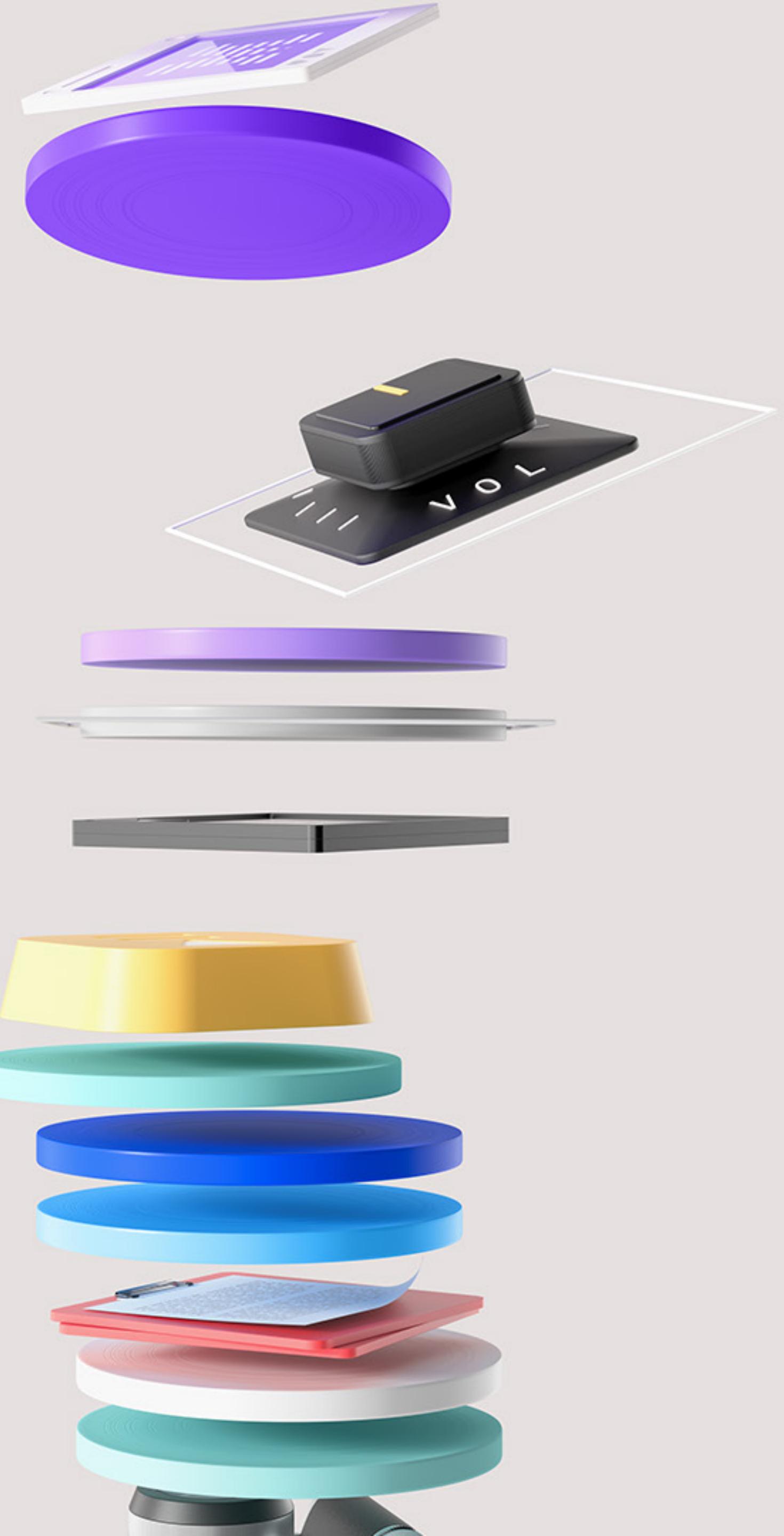


O único data  
lakehouse  
híbrido e aberto  
para IA empresarial  
e análise de dados



## Índice

[01 →](#)

Por que sua empresa  
precisa de um data  
lakehouse preparado  
para a IA generativa

[04 →](#)

Chegou watsonx.data

[05 →](#)

Por que escolher  
o watsonx.data

[02 →](#)

Desafios de dados atuais  
e emergentes

[06 →](#)

Próximas etapas

[03 →](#)

Definição de data  
lakehouse

# Por que sua empresa precisa de um data lakehouse preparado para a IA generativa



Dados genéricos produzem resultados genéricos. Se os dados não forem específicos para sua empresa, os resultados também não serão. Quando todos estão usando os mesmos dados públicos para treinar os mesmos modelos de IA generativa. Como você pode superar seus concorrentes?

Sua vantagem competitiva está em seus dados exclusivos. Dados não estruturados (como aqueles em e-mails, PDFs, imagens e vídeos) são especialmente valiosos, mas muito difíceis de operacionalizar. As organizações precisam de uma arquitetura de dados inteligente para acessar, preparar e fornecer dados estruturados e não estruturados para liberar todo o potencial de seus dados empresariais para a IA.

Os dados são o combustível para a IA, mas estima-se que menos de 1% dos dados empresariais estejam sendo aproveitados por modelos de IA.<sup>1</sup> E 90% dos dados empresariais são não estruturados.<sup>2</sup>

Um data lakehouse híbrido e aberto pode ajudar você a turbinar sua IA e integrar seus dados a uma ampla variedade de casos de uso modernos. Ele unifica todos os tipos de dados, simplifica a integração e torna o gerenciamento e a administração de dados (estruturados e não estruturados) mais fácil do que nunca. Não é de se admirar que esteja emergindo como a arquitetura de dados de longo prazo preferida para análise de dados e novas cargas de trabalho de IA generativa.

# 90%

de dados empresariais  
são estimados como  
sendo não estruturados.<sup>2</sup>



## Desafios de dados atuais e emergentes

À medida que a importância dos dados empresariais aumenta, também aumentam os desafios associados a eles. O enorme volume de dados tornou-se esmagador e, muitas vezes, está isolado em uma organização. Além disso, novas variedades de dados complicaram a integração, e dados de baixa qualidade comprometeram o impacto na IA. A IA generativa pode ajudar a resolver esses problemas, mas uma arquitetura de dados resiliente e flexível é essencial para isso.

Atualmente, as limitações da geração aumentada de recuperação (RAG) impedem que as empresas obtenham o valor dos dados não estruturados para a IA generativa. Então, como trazer dados não estruturados para a IA generativa e as cargas de trabalho analíticas tradicionais? Adotando um data lakehouse híbrido e aberto.



## Definição de data lakehouse



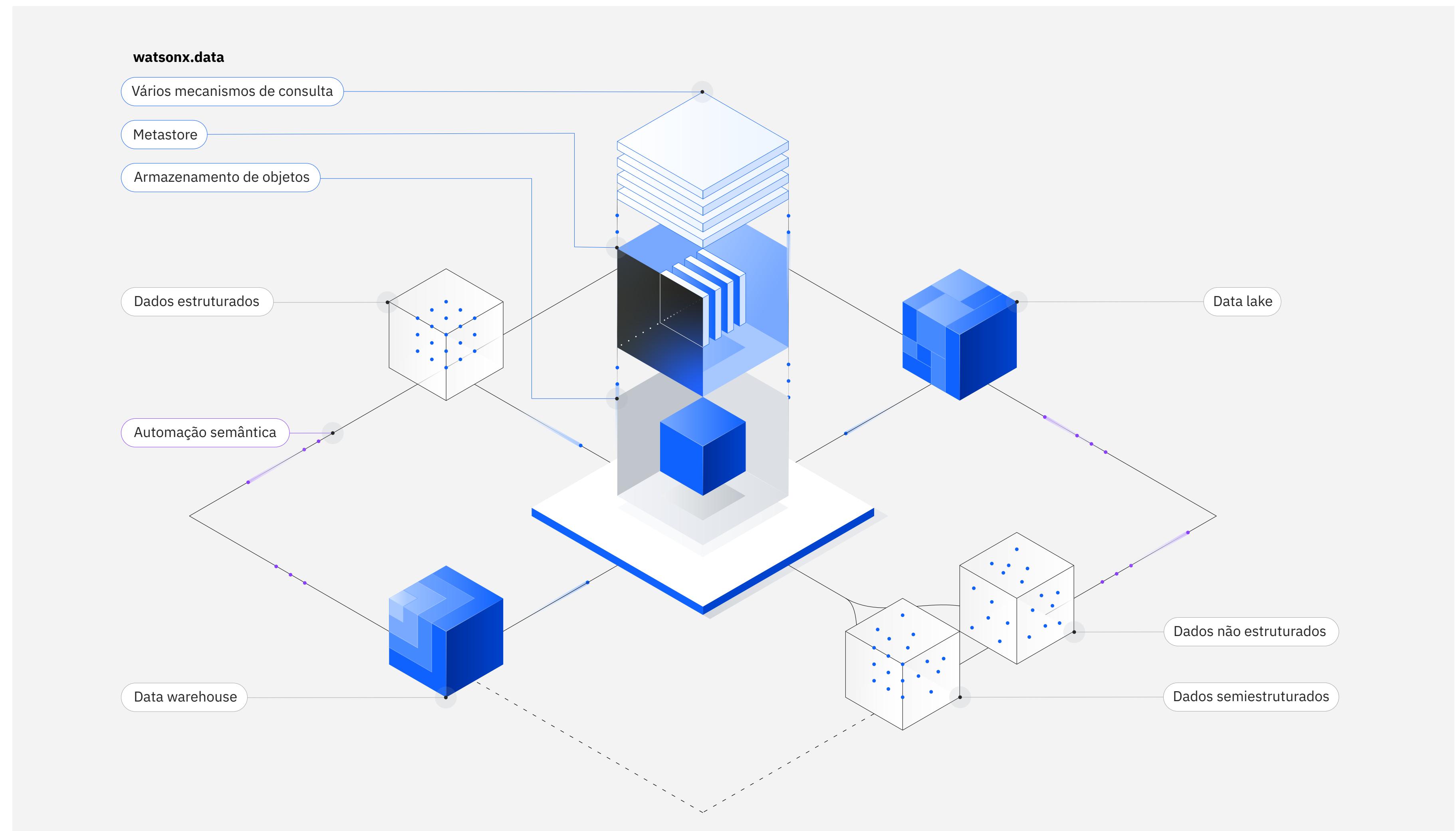
Os dados empresariais estão no centro de qualquer arquitetura de IA. As organizações que buscam ferramentas de IA generativa em 2025 precisarão garantir que seus dados estejam prontos para a IA. Uma solução de lakehouse fornece uma base moderna para escalar dados prontos para a IA.

O data lakehouse é uma arquitetura emergente que oferece a flexibilidade de um data lake com o desempenho e a estrutura de um data warehouse. A maioria das soluções de lakehouse oferece mecanismos de consulta de alto desempenho em um armazenamento de baixo custo, em conjunto com uma camada de governança de metadados. As camadas de metadados inteligentes facilitam o trabalho dos usuários na categorização e classificação de dados não estruturados, como vídeos e voz, e dados semiestruturados, como XML, JSON e e-mails.

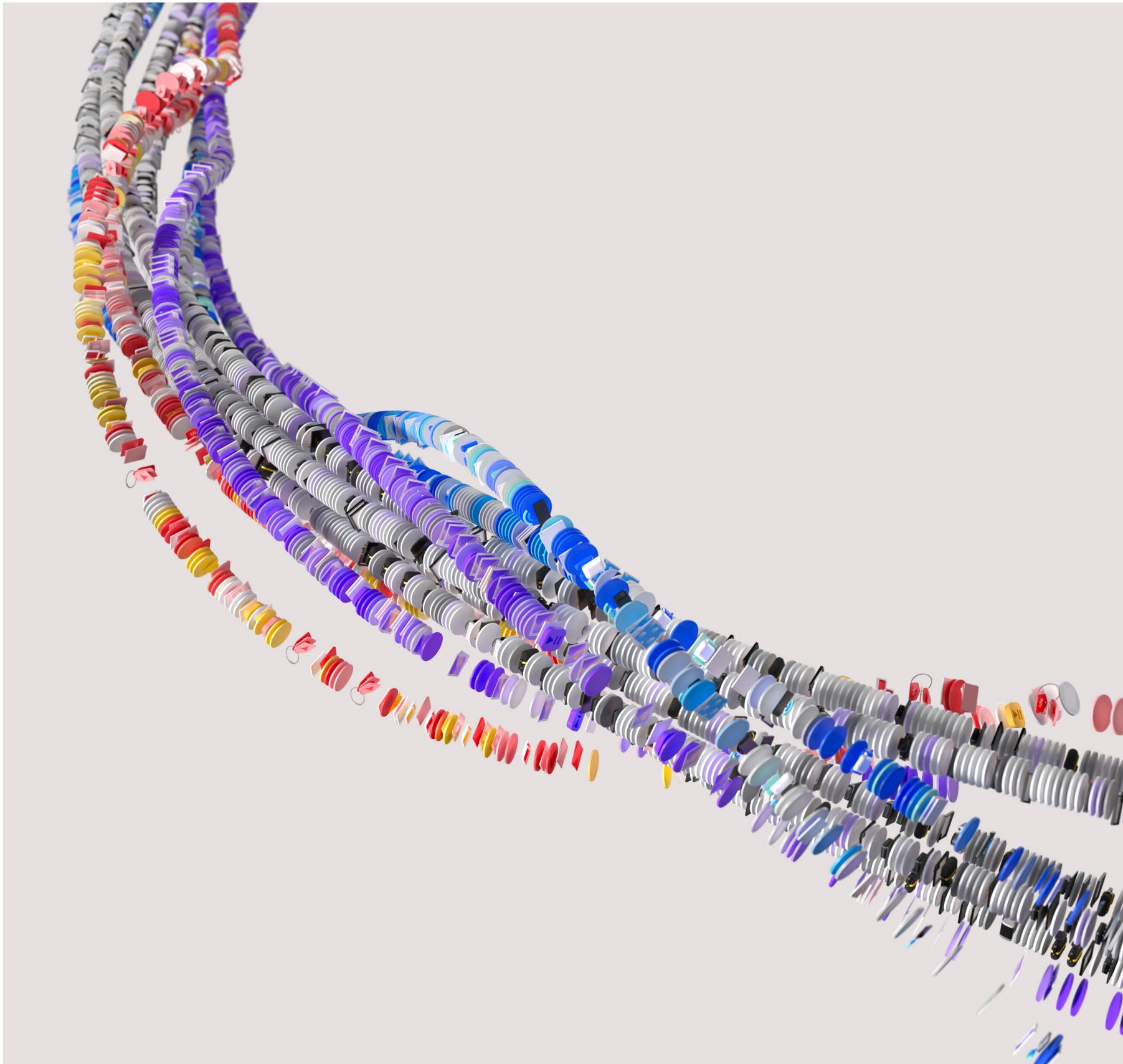
Os melhores data lakehouses oferecem tecnologias de código aberto que reduzem a duplicação de dados e simplificam os complexos pipelines de ETL. No entanto, alguns lakehouses de primeira geração têm restrições para extrair, transformar e carregar (ETL) que limitam sua capacidade de lidar com os desafios de custo e complexidade. Por exemplo, um único mecanismo de consulta projetado para cargas de trabalho de business intelligence (BI) ou aprendizado de máquina (ML) pode ser ineficaz quando usado para outro tipo de carga de trabalho.

As equipes de dados e IA da IBM acreditam que cada carga de trabalho é única e deve ser otimizada com o ambiente mais adequado, que minimize os custos e aumente o desempenho o máximo possível. Em última análise, transformar dados brutos em insights praticáveis em escala requer um lakehouse que ofereça desempenho econômico e flexibilidade de formatos.

Gerencie todo o ciclo de vida dos dados para a IA



## Chegou watsonx.data

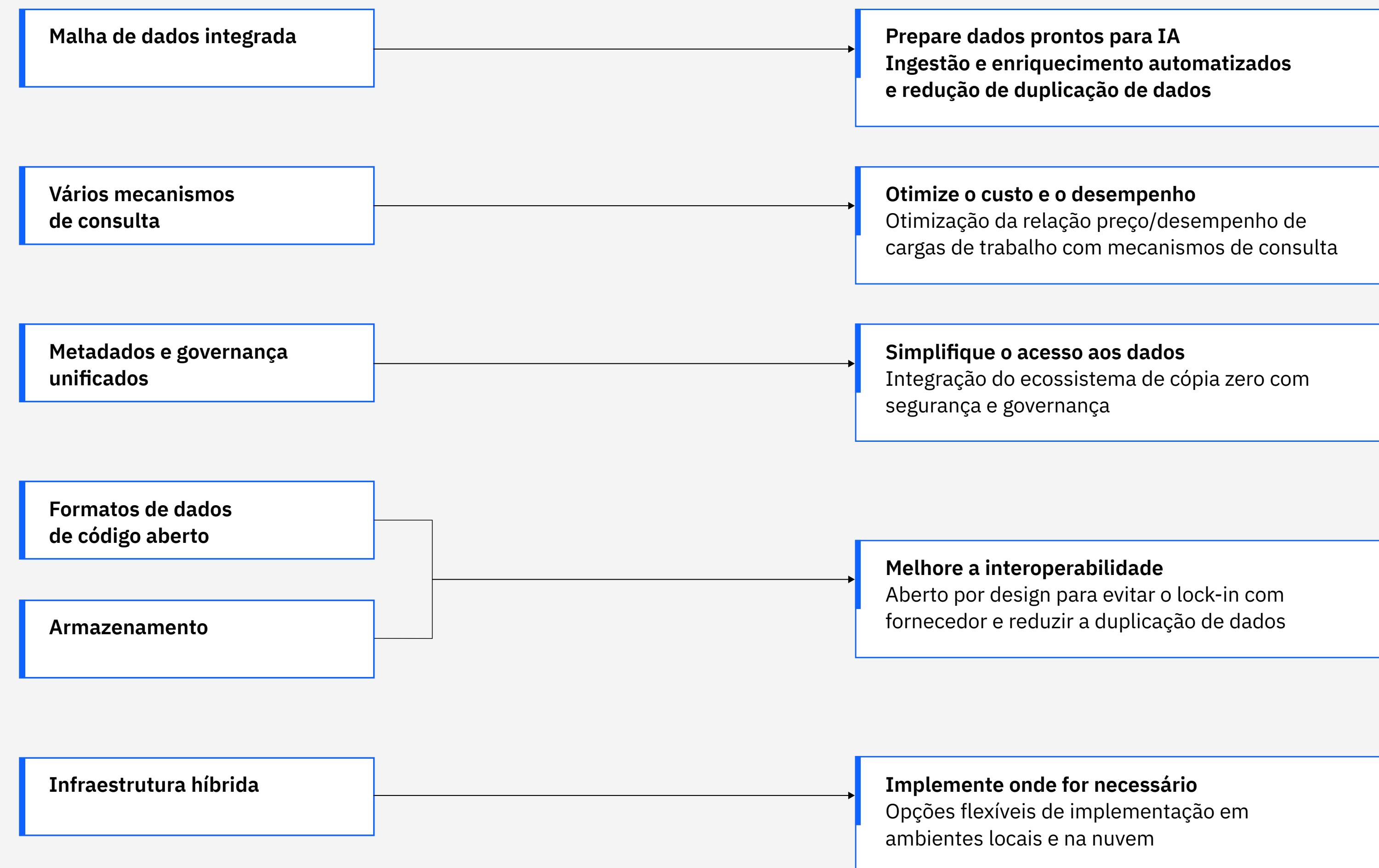


O IBM® watsonx.data® é o único data lakehouse híbrido e aberto para IA empresarial e análise de dados.

Ele ajuda você a gerenciar todo o ciclo de vida dos dados corporativos para IA no seu data lakehouse, impulsionando a próxima geração de aplicações de IA generativa e BI com IA integrada.

Com o watsonx.data, você pode simplificar e escalar a integração, o gerenciamento e a governança de dados estruturados e não estruturados em ambientes locais, na nuvem e multinuvem. Faz parte de um stack de dados moderno e aberto que utiliza tecnologias inovadoras de código aberto e se integra ao seu ambiente de dados existente, sem lock-in com fornecedor.

O IBM watsonx.data integra sem dificuldades os recursos de um data lakehouse e a malha de dados.



Quais são os componentes do watsonx.data?

Uma arquitetura de data lakehouse híbrida e aberta oferece uma plataforma unificada para lidar com os principais desafios de dados que as empresas enfrentam atualmente: grandes volumes de dados, vários locais e formatos, qualidade e governança de dados.

Veja como o IBM watsonx.data lida com esses desafios.

#### Infraestrutura

A infraestrutura é onde seu lakehouse será implementado e completamente gerenciado em todos os ambientes locais e na nuvem.

#### Armazenamento

A camada de armazenamento é onde os dados são fisicamente armazenados como arquivos ou em formatos de dados abertos, como Apache Parquet e Avro. Os formatos de dados abertos são especificações de arquivos e protocolos disponibilizados para a comunidade de código aberto, para que qualquer pessoa possa ingeri-los e aprimorá-los.

#### Mecanismos de consulta adequados para o seu objetivo

Os múltiplos mecanismos de consulta adequados ao propósito permitem que as consultas sejam executadas com eficiência em grandes conjuntos de dados. Com a otimização das cargas de trabalho em vários mecanismos de busca e camadas de armazenamento, é possível minimizar o custo de seu data warehouse e, ao mesmo tempo, proporcionar um processamento de big data rápido, confiável e eficiente em escala.

#### Formatos de tabela abertos

Formatos de abertos tabelas, como o Apache Iceberg, ajudam a fornecer estrutura e proporcionam a confiabilidade e a simplicidade do SQL com big data. Esses formatos permitem que diferentes mecanismos acessem os mesmos dados, ao mesmo tempo, o que ajuda a evitar o lock-in com fornecedor. Os dados também podem ser compartilhados entre várias ferramentas e repositórios de dados, como seu data warehouse. Uma única cópia dos dados permite que você reduza a duplicação de dados e acabe com os silos.

#### Armazenamento de metadados

O armazenamento de metadados no watsonx.data baseia-se no Hive Metastore de código aberto, que ajuda a garantir a interoperabilidade e permite que ferramentas externas acessem facilmente seus metadados.

### Controle

Os metadados também são armazenados em formatos abertos de tabela. Eles servem para definir os formatos dos arquivos para qualquer ferramenta que possa ler ou escrever formatos de dados abertos.

### Serviço de metadados técnicos

Esse componente do serviço técnico de metadados é necessário para entender quais dados estão disponíveis na camada de armazenamento. O mecanismo de consulta usa metadados para fornecer informações completas de linhagem, para que você saiba onde seus dados estão localizados, sua aparência e como lê-los.

### Catálogos de dados

Os catálogos de dados ajudam os usuários a encontrar os dados certos para o trabalho e proporciona informações semânticas para as políticas e regras. Metadados comerciais, como terminologias e tags comerciais, são armazenados para compatibilidade com a pesquisa e a proteção de dados.

### Mecanismo de políticas

O mecanismo de políticas permite que os usuários definam políticas de proteção de dados e garante que essas políticas sejam aplicadas. Para criar um framework de governança escalável, um mecanismo de políticas é frequentemente implementado com o serviço técnico de metadados e o catálogo de dados.

### Mecanismo de consulta

O mecanismo de consulta, frequentemente conhecido como componente de computação, está no centro do data lakehouse aberto. Ele pode ser de código aberto ou proprietário e acessa dados em formatos de tabelas abertos.

Em uma arquitetura aberta de lakehouse, o mecanismo de consultas é totalmente modular, o que significa que ele pode ser escalado dinamicamente para atender às demandas de carga de trabalho e simultaneidade. Os mecanismos de consulta também são capazes de se conectar com qualquer sistema de catálogos e armazenamento.

O que diferencia o IBM watsonx.data dos demais?

#### Híbrido e aberto

O watsonx.data foi construído com tecnologias básicas de código aberto para viabilizar a inovação e o suporte da comunidade — sem lock-in com fornecedor. Você pode acessar os dados e implementar em qualquer lugar: em ambientes de nuvem híbrida, VPC no cliente, multinuvem, ambientes no local e como SaaS.

#### Otimizado para cargas de trabalho

Apoiado por recursos integrados de malha de dados, incluindo gestão de dados, qualidade e linhagem, o watsonx.data ajuda você a gerenciar volumes crescentes de dados com uma arquitetura exclusiva de vários mecanismos com uma relação preço/desempenho otimizada.

#### Preparado para a IA generativa

O watsonx.data viabiliza a autorização de usuário consistente e de ponta a ponta, de sistemas de obtenção de documentos até a recuperação.

#### Como tudo se conecta?

O watsonx.data permite que as organizações liberem as informações críticas para os negócios contidas em seus dados não estruturados, proporcionando que a IA generativa e as análises sejam mais precisas e poderosas.

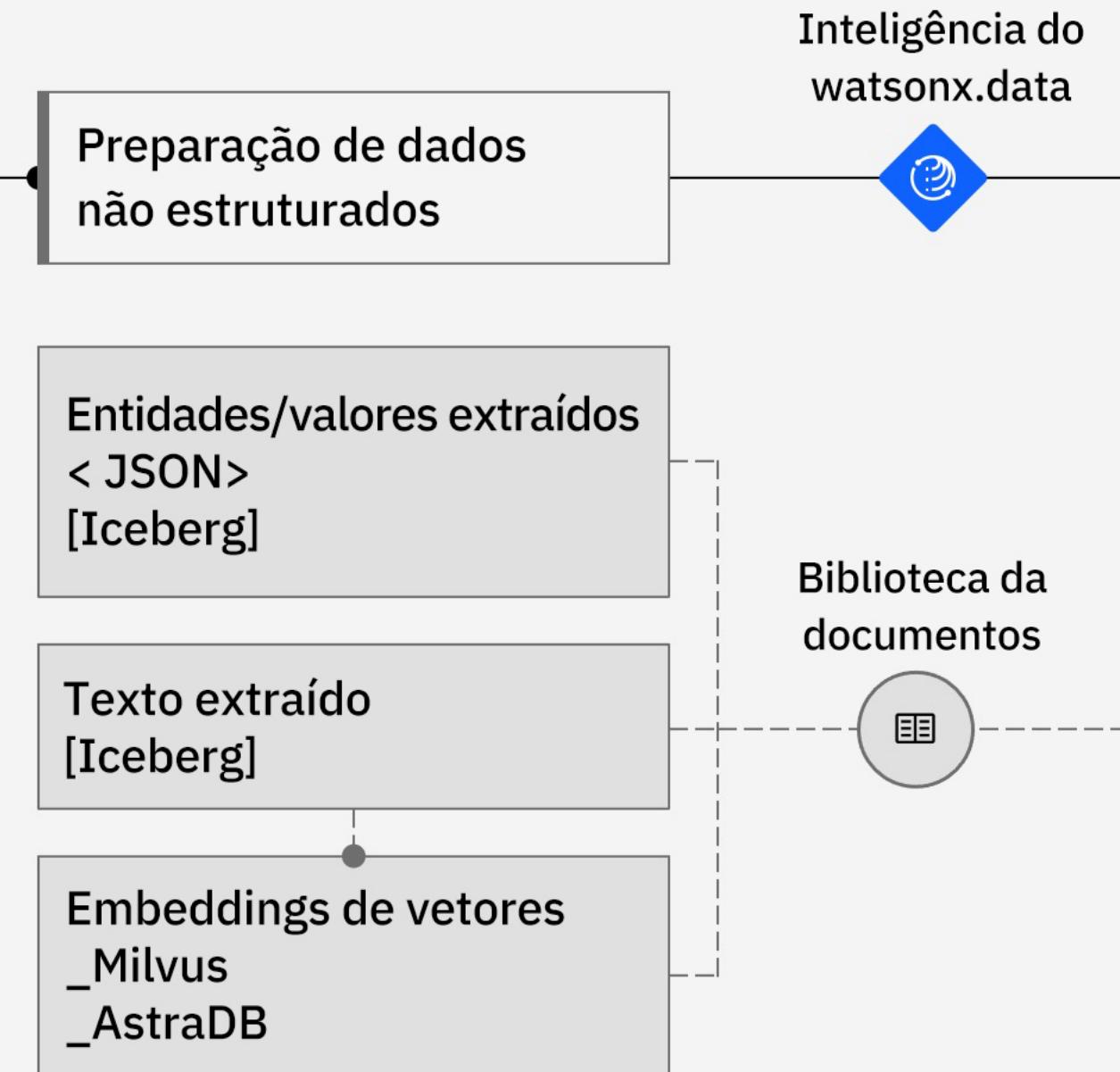
40%

maior precisão da IA do que o RAG convencional com o watsonx.data<sup>3</sup>

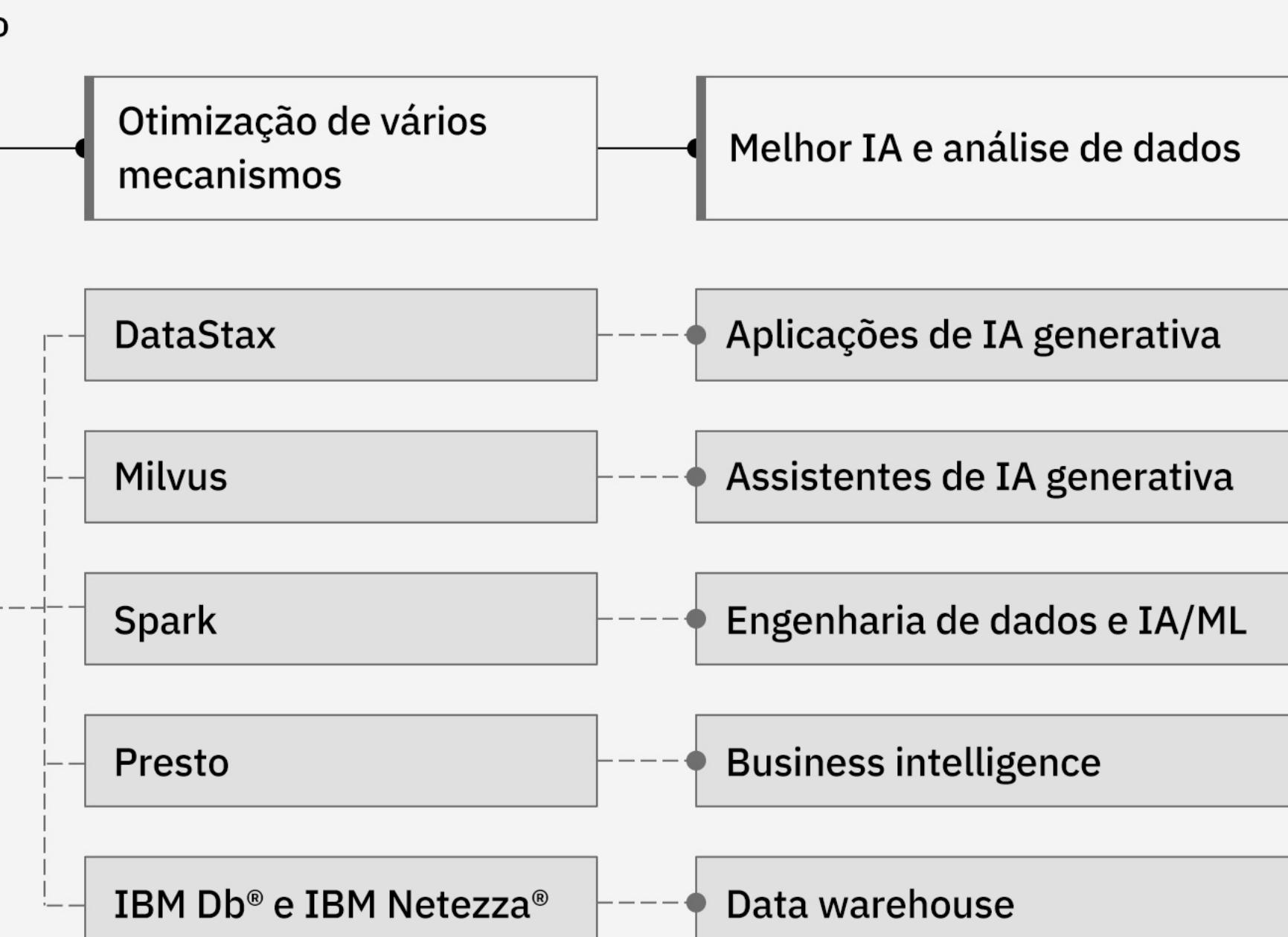
## Ingestão em escala



## Selecionar e governar



## Entregar à IA



Controles consistentes de acesso do usuário

## Por que escolher o watsonx.data

-  Para unificar dados estruturados e não estruturados para aumentar a precisão da IA e o ROI
-  Para liberar dados inexplorados para análise de dados conversacional
-  Para simplificar e escalar a integração, o gerenciamento e a governança de dados estruturados e não estruturados
-  Para capacitar todos os usuários com dados de alta qualidade e confiáveis
-  Para proteger seus investimentos atuais em dados com um stack de dados moderno e aberto

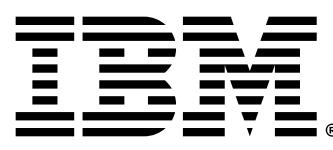
## Próximas etapas

Aproveite o conhecimento da IBM no gerenciamento e otimização de dados, acumulado ao longo de décadas lidando com as cargas de trabalho de dados mais exigentes do mundo. Veja com que rapidez você pode maximizar o ROI de seus dados com o watsonx.data.

Comece sua avaliação sem custo →

Solicite uma 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. Com base em testes internos que compararam a precisão das respostas geradas por modelos de IA usando a camada de recuperação do watsonx.data Premium Edition com RAG apenas com vetores, em três casos de uso comuns, com conjuntos de dados proprietários da IBM e o mesmo conjunto de modelos abertos selecionados para inferência, avaliação e embeddings, além de variáveis adicionais. Os resultados podem variar.

© Copyright IBM Corporation 2025

IBM, o logotipo da IBM, Db2, FileNet, Netezza, e watsonx.data são marcas comerciais ou marcas registradas da International Business Machines Corporation, nos Estados Unidos e/ou em outros países. Outros nomes de produtos e serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual de marcas comerciais da IBM está disponível em [ibm.com.br-pt/legal/copytrade](http://ibm.com.br-pt/legal/copytrade).

Este documento estava atualizado na data de sua publicação inicial e pode ser alterado pela IBM a qualquer momento.

Nem todas as ofertas estão disponíveis em todos os países nos quais a IBM opera.

É responsabilidade do usuário verificar a operação de quaisquer produtos ou programas não pertencentes à IBM com produtos e programas da IBM. A IBM não é responsável por produtos e programas que não sejam da IBM.

AS INFORMAÇÕES CONTIDAS NESTE DOCUMENTO SÃO FORNECIDAS NO ESTADO EM QUE SE ENCONTRAM, SEM QUALQUER GARANTIA, EXPRESSA OU IMPLÍCITA, INCLUSIVE SEM NENHUMA GARANTIA DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UMA DETERMINADA FINALIDADE NEM QUALQUER GARANTIA OU CONDIÇÃO DE NÃO INFRAÇÃO. Os produtos IBM têm garantia de acordo com os termos e condições dos contratos sob os quais são fornecidos.

Nenhum sistema ou produto de TI deve ser considerado completamente seguro e nenhum produto, serviço ou medida de segurança será completamente eficaz em impedir o uso ou acesso indevido. A IBM não garante que nenhum de seus sistemas, produtos ou serviços é imune, tampouco que imunizará a sua empresa, contra condutas maliciosas ou ilegais por parte de terceiros.

O cliente é responsável por garantir o cumprimento de todas as leis e regulamentações aplicáveis. A IBM não oferece orientação jurídica, não declara nem garante que seus serviços ou produtos farão com que o cliente esteja em conformidade com nenhuma lei ou regulamentação.