

A arquitetura de malha de dados oferece 3 benefícios imediatos

Integre e aproveite todos os dados da sua organização para obter melhores resultados de negócios

Índice

- 3** Os desafios dos dados hoje
- 5** Benefícios comerciais de uma malha de dados
- 6** Como a IBM entrega um malha de dados
- 7** Malha de dados em ação
- 7** A malha de dados é a próxima etapa transformadora para sua empresa

Os desafios dos dados atualmente

Os dados são um elemento integrante da transformação digital para empresas.

Mas, à medida que as organizações buscam alavancar seus dados, elas encontram desafios resultantes de diversas fontes de dados, tipos, estruturas, ambientes e plataformas. Esta situação de dados multidimensionais é ainda mais complicada quando as organizações adotam arquiteturas de nuvens híbridas e multinuvem. Para muitas empresas hoje, os dados operacionais permaneceram em grande parte isolados e ocultos, levando a uma enorme quantidade de dados obscuros.

Veja o exemplo de uma empresa de energia norte-americana que buscou se reinventar como uma organização orientada a dados para se tornar uma empresa na qual os recursos de ciência de dados estivessem prontamente acessíveis em várias unidades de negócios. Eles logo perceberam que sua transformação digital foi prejudicada por dados isolados, ferramentas inconsistentes e vários níveis de habilidade, o que causou lacunas críticas entre as competências de dados.

O problema que enfrentaram não era exclusivo de seus negócios; na verdade, é uma consequência comum de cenários de dados que ultrapassaram suas arquiteturas de gerenciamento de dados.

O que finalmente trouxe a empresa de energia de volta a um caminho de sucesso para a transformação digital foi o emprego de um novo conceito de arquitetura de dados conhecido como malha de dados.

O que exatamente é uma malha de dados, como ela difere das arquiteturas anteriores, o que ela pode alcançar para as empresas e qual é o papel da IBM em sua implementação? Este white paper responderá a essas perguntas.

O que é uma malha de dados?

No passado, as organizações tentaram resolver os problemas de acesso aos dados por meio da integração ponto a ponto ou da introdução de hubs de dados. Nenhum deles é adequado quando os dados são altamente distribuídos e isolados. As integrações ponto a ponto adicionam custo exponencial para qualquer terminal adicional que precise ser conectado, o que significa que esta é uma abordagem não escalonável. Os hubs de dados permitem uma integração mais fácil de aplicações e fontes, mas exacerbam o custo e a complexidade para manter a qualidade e a confiança dos dados no hub.

A malha de dados é uma arquitetura emergente que visa abordar os desafios de dados decorrentes de um cenário de dados híbrido. Sua ideia fundamental é encontrar um equilíbrio entre a descentralização e a globalização, agindo como o tecido conectivo virtual entre os terminais de dados (consulte a página 4).

Por meio de tecnologias como automação e aumento da integração, controle federado e ativação de metadados, uma arquitetura de malha de dados permite a orquestração de dados dinâmica e inteligente em um cenário distribuído, criando uma rede de informações disponíveis instantaneamente para impulsionar um negócio.

Uma malha de dados é independente de plataformas de implementação, processos de dados, localizações geográficas e abordagem arquitetônica. Facilita o uso de dados como um ativo corporativo. Uma malha de dados garante que seus vários tipos de dados possam ser combinados, acessados e controlados com êxito de forma eficiente e eficaz.

Capacidades e princípios de uma malha de dados

O núcleo da arquitetura da malha de dados é uma plataforma de gerenciamento de dados que permite toda a amplitude de recursos integrados de gerenciamento de dados, incluindo descoberta, controle, curadoria e orquestração.

No entanto, uma malha de dados avança e evolui de conceitos tradicionais de gerenciamento de dados, como DataOps, que se concentra apenas em estabelecer práticas, para aumentar o nível de operacionalização de dados. É construído sobre uma arquitetura distribuída e tecnologia avançada capaz de atender às necessidades que surgem da extrema diversidade e distribuição de ativos de dados.

Um malha de dados pode ser logicamente dividida em 4 capacidades (ou componentes):

Conhecimento, insights e semântica.

- Fornece um mercado de dados e experiência de compra.
- Enriquece automaticamente os dados descobertos ativos com conhecimento e semântica, permitindo ao consumidor encontrar e entender os dados.

Controle unificado e conformidade

- Permite gerenciamento local e governança de metadados, mas oferece suporte a uma visão global unificada e cumprimento de política.
- Aplica automaticamente políticas sobre dados ativos de acordo com as regras globais e locais.
- Utiliza capacidades avançadas para automatizar a classificação e curadoria de ativos de dados.
- Estabelece automaticamente rotas de acesso consultáveis para qualquer ativo catalogado para aumentar a ativação dos dados.

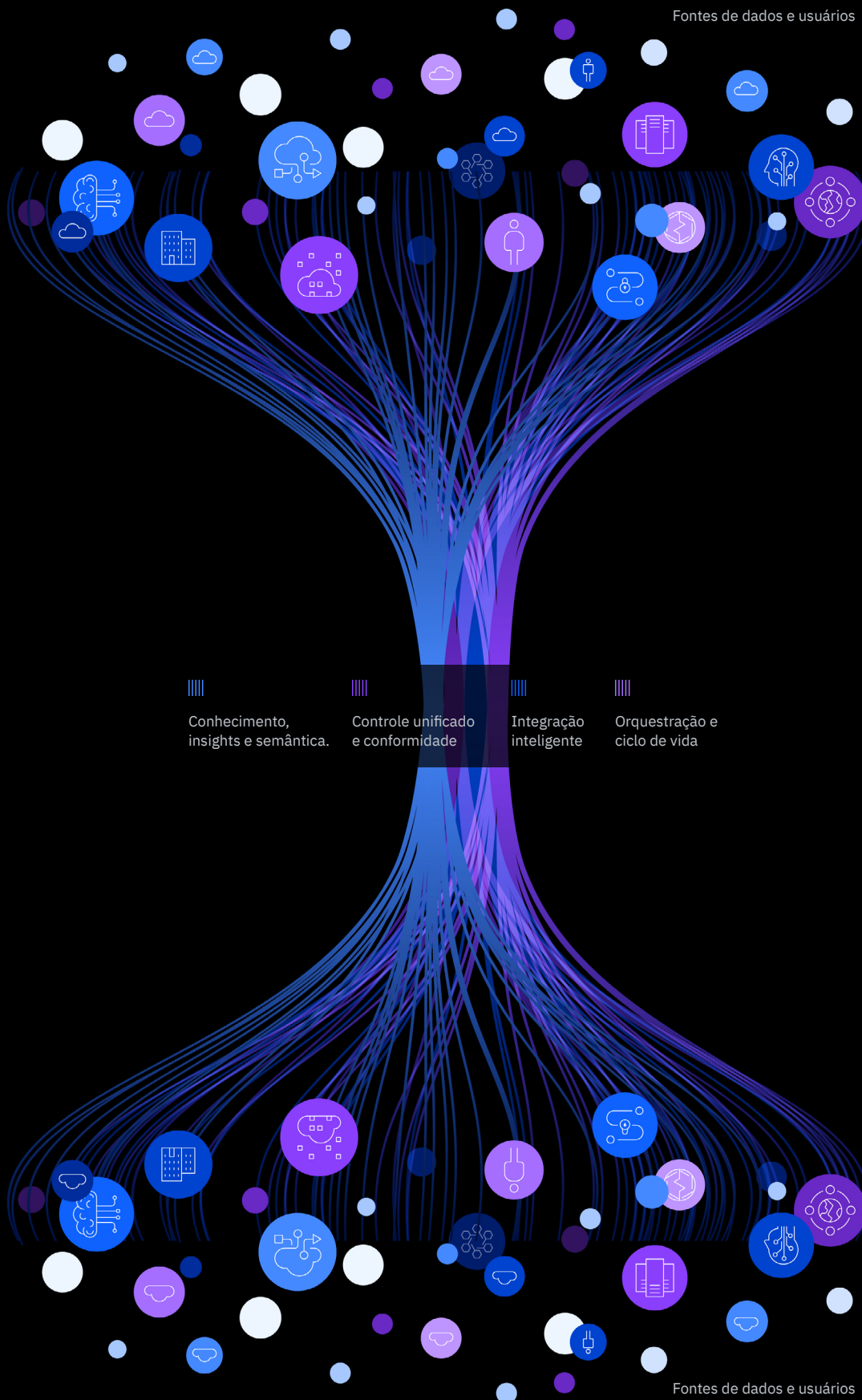
Integração inteligente

- Acelera as tarefas da equipe de engenharia de dados por meio de fluxo automatizado e criação de pipeline em fontes de dados distribuídas.
- Permite ingestão de autoatendimento e acesso a dados sobre quaisquer dados com aplicação profunda local e global de políticas de proteção de dados.
- Determina automaticamente a execução do melhor ajuste por meio da distribuição de carga de trabalho otimizada, autoajuste e correção de desvios de esquema.

Orquestração e ciclo de vida

- Permite composição, teste, operação e monitoramento de pipelines de dados.
- Desenvolve recursos de IA no ciclo de vida dos dados para automatizar tarefas, autoajuste, autocorreção e detecção de mudanças de dados de origem, tudo o que facilita as atualizações automatizadas

Fontes de dados e usuários



Benefícios comerciais de uma malha de dados

Os dados só fornecem valor de negócios quando são contextualizados e se tornam acessíveis a qualquer usuário ou aplicação na organização. Quando implementado corretamente, uma malha de dados ajuda a garantir que esses valores estejam disponíveis em toda a organização da maneira mais eficiente e automatizada possível. Como tal, a malha tem 3 benefícios principais:

1. Permitir consumo de dados de autoatendimento e colaboração
2. Automatize o controle, a proteção e a segurança; ativado por metadados ativos.
3. Automatize tarefas de engenharia de dados e aumente a integração de dados em recursos de nuvem híbrida.

Permitir consumo de dados de autoatendimento e colaboração

Ao integrar dados de várias fontes e analisar uma fração maior da enorme quantidade de dados gerados diariamente, as organizações obtêm melhores insights e respondem mais rapidamente às mudanças nas demandas de negócios. Um malha de dados entrega rapidamente os dados nas mãos de quem precisa deles. O autoatendimento permite que a organização como um todo encontre os dados apropriados com mais rapidez e gaste mais tempo usando esses dados para fornecer insights tangíveis.

Benefícios da malha de dados para consumo de dados de autoatendimento:

- Os usuários corporativos têm um único ponto de acesso para localizar, compreender, moldar e consumir dados em toda a organização.
- Um controle centralizado de dados e linhagem ajuda os usuários a entender o que os dados significam, de onde eles vêm e como estão relacionados a outros ativos.
- O gerenciamento de metadados extenso e customizável pode ser facilmente dimensionado e acessível por meio de APIs.
- O acesso de autoatendimento a dados confiáveis e controlados permite a colaboração de linha de negócios com outros usuários.

Um estudo Forrester Total Economic Impact¹ revelou que esses recursos podem significar:

USD
5,8 Mi 459%
em benefícios de ROI

Automatize a governança, proteção de dados e segurança; ativado por metadados ativos.

Uma camada de controle ativa distribuída para todas as iniciativas de dados reduz a conformidade e os riscos regulamentares ao fornecer confiança e transparência.

Ela permite a aplicação automática de políticas para qualquer acesso a dados, fornecendo um alto nível de proteção e conformidade de dados.

A utilização de tecnologias de IA e aprendizado de máquina permite que os usuários de malha de dados aumentem seu nível de automação, por exemplo, extraindo automaticamente as regras de governança de dados com base na linguagem e nas definições dos documentos regulatórios. Isso permite que as organizações apliquem regras de controle específicas do setor em questão de minutos para ajudar a evitar multas caras e garantir o uso ético dos dados onde quer que eles residam.

Benefícios de um malha de dados para uma virtualização controlada:

- A agilidade, segurança e produtividade é aumentada para as equipes de engenharia de dados, cientistas de dados e analistas de negócios;
- Várias fontes de dados globais aparecem como um banco de dados;
- A descoberta nova e líder do setor de informações pessoalmente identificáveis (PII) e elementos de dados críticos é possível em grande escala;

Estas capacidades podem significar:

USD
2,4 Mi 430%
em benefícios¹ melhoria de desempenho²

Automatizar tarefas de engenharia de dados e aumentar a integração de dados.

Engenharia de dados avançada significa que virtualmente qualquer acesso de dados ou processo de entrega é automatizado e não requer nenhum processo de codificação tedioso ou sujeito a erros. O aumento da integração utiliza dados de metadados para otimizar a entrega e o acesso aos dados.

Benefícios de uma malha de dados para integração e engenharia de dados:

- A integração de dados otimizada automaticamente ajuda a acelerar a entrega de dados;
- O balanceamento automático da carga de trabalho e o escalonamento elástico significam que os trabalhos estão prontos para qualquer ambiente e qualquer volume de dados;
- Resiliência e automação de CI/CD integradas;
- O processo automatizado de captura de mudanças em tempo real oferece suporte à entrega de dados de qualidade para processos de negócios;
- O aprendizado de máquina pode automatizar e estender os processos de descoberta, classificação e curadoria de dados customizados, levando a um tempo de maturação mais rápido;
- A análise contínua pode ser realizada automaticamente em tempo real, onde quer que os dados residam.

Os resultados de um varejista líder³ mostram:

60 vezes 20 vezes
maior aceleração no tempo de entrega de dados mais rápido ao realizar análises de afinidade do cliente

Como a IBM entrega um malha de dados

O IBM Cloud Pak® for Data torna este conceito de malha de dados possível. IBM Cloud Pak for Data é uma plataforma de insight que simplifica e automatiza a coleta de dados, organização, análise de dados e acelera a infusão de IA em toda a sua empresa.

Com seus recursos para conectar dados em qualquer lugar, execute cargas de trabalho em qualquer lugar; e para construir, implementar e gerenciar IA em escala em ambientes de nuvem híbrida, o IBM Cloud Pak for Data é o ativador para uma transformação digital de negócios.

A plataforma oferece integração perfeita em uma empresa híbrida para:

- Serviços IBM Cloud Pak for Data
- Aplicações externas e fontes de dados
- Recursos avançados baseados em IA para gerenciamento e controle de dados

Essa base disponibiliza os dados selecionados para os consumidores com o equilíbrio ideal de custo, desempenho, conformidade e com a inteligência para orquestrar e otimizar o processamento de dados com base em cargas de trabalho, localidade de dados e políticas de dados.

O IBM Cloud Pak for Data potencializa os recursos a seguir para fornecer os dados prontos para negócios que uma malha de dados exige. Todos esses recursos desempenham um papel no suporte à arquitetura da malha de dados.

1. Núcleo de conhecimento baseado em metadados

Facilita a descoberta de fontes de dados e catálogos; enriquece os ativos de dados e realiza análises para extrair insights para mais automação usando IA. O núcleo de conhecimento é usado para impulsionar o mercado com a pesquisa semântica.

2. Mercado de dados de autoatendimento

O catálogo de dados de última geração que ajuda os consumidores de dados, como analistas de negócios, a recuperar dados de todo o cenário de dados da empresa.

3. Integração inteligente

Permite o consumo de dados extraíndo, virtualizando, transformando e transmitindo dados. Ele é integrado ao núcleo de conhecimento para automatizar a integração de dados e tem inteligência para decidir qual abordagem de integração é mais adequada com base em cargas de trabalho e políticas de dados. Ele também pode ser usado para preparação de dados como parte de cargas de trabalho de engenharia de dados ou para criar produtos de dados. Finalmente, ele fornece a capacidade de publicar atualizações para produtos de dados.

4. Governança

Cataloga e organiza metadados, define políticas de dados para privacidade, organiza dados, captura a linhagem de dados e executa outras tarefas relacionadas à segurança e conformidade. Essa camada entende o formato dos dados (por exemplo, estruturado vs não estruturado) e a importância dos dados (por exemplo, público vs protegido) e aplica as políticas corretas a cada pedaço de dado e a cada usuário em potencial. Em vez de aplicar padrões e regras aos dados manualmente, esse recurso de plataforma integrada significa que eles podem ser aplicados no nível organizacional e se propagam por meio de vários recursos de dados conforme necessário. Modelos analíticos em diferentes ferramentas podem se comunicar entre si; a aplicação da política de dados em nível granular pode ser amplamente automatizada.

5. Desenvolvimento e operações unificados

Permite um ciclo de vida unificado para configurar e executar todos os aspectos da plataforma de dados em produção.



Figura 1. Os recursos da plataforma IBM Cloud Pak for Data - todos os quais suportam o desenvolvimento de IA e a malha de dados

Malha de dados em ação

A melhor maneira de entender o valor de uma malha de dados é ver os benefícios de negócios que as organizações reais alcançaram com a implementação de uma malha de dados sobre uma plataforma de insight.

Empresa de energia

A seção de desafios no início deste white paper apresentou o caso de uma empresa de energia norte-americana em transformação digital.

Este cliente se envolveu com a IBM, que implementou uma arquitetura de malha de dados baseada no IBM Cloud Pak for Data.

Com uma malha de dados flexível e integrada, a empresa foi capaz de implementar uma série de projetos de dados importantes em diferentes unidades, incluindo:

- eMobility;
- Descoberta de documentos de operações de gás, incluindo extração de caligrafia;
- Segmentação de clientes elétricos e previsão de carga;
- Gerenciamento de ativos;
- Impactos de carga da COVID-19;
- Modelo de risco de retorno ao trabalho.

Com uma malha de dados, esta organização é capaz de:

- Fornecer a várias unidades de negócios acesso direto aos dados por meio de uma plataforma de insight de autoatendimento;
- Usar a capacidade de computação adequada para a finalidade a fim de executar modelos com eficiência em bilhões de linhas de dados;
- Desenvolver modelos de forma colaborativa e implemente facilmente esses modelos para infundir percepções em toda a empresa.

Reduzindo o movimento e melhorando a supervisão

Um grande provedor de serviços industriais melhorou o controle de dados e facilitou a conformidade com as regulamentações.

Essa organização lutou para mover grandes quantidades de dados para seu data lake baseado em nuvem. Ela tinha os requisitos gêmeos de aplicar uma forte camada de governança a cada instância de acesso a dados e determinar a qualidade dos dados antes de fornecer acesso aos usuários corporativos.

Ao implementar uma malha de dados construída no IBM Cloud Pak for Data, essa organização fez enormes melhorias em seus processos de controle de dados, conformidade de dados e transformação de dados. Com uma malha de dados instalada, ela pode fornecer aos usuários de negócios acesso fácil e seguro a centenas de fontes de dados em seu data lake baseado em nuvem e fontes de dados SAP locais.

Ela também está aproveitando as ferramentas de aceleração regulamentar do setor para verificar as fontes de dados em busca de possíveis dados PII que estariam sujeitos à regulamentação GDPR e CCPA. A sinalização automática de dados ajuda a identificar quaisquer fontes de dados que precisam ser gerenciadas para manipulação de PII tanto no local quanto na nuvem.

Com melhor acesso e triagem de PII mais rápida, os usuários de negócios podem minerar dados relevantes em busca de insights importantes sem longas esperas pelo acesso aos dados ou correndo o risco de expor dados protegidos.

Com uma malha de dados, a organização é capaz de:

- Garantir o controle adequado de dados, ao mesmo tempo que aproveita os dados de toda a empresa;
- Obter dados confiáveis e reduzir a quantidade de preparação de dados;
- Ajudar na conformidade com os regulamentos de privacidade, como GDPR e CCPA.

A malha de dados é a próxima etapa transformadora para sua empresa

Com uma malha de dados desenvolvida com base na tecnologia IBM Cloud Pak for Data, é possível hiper-automatizar a descoberta, o controle e consumo de dados em um cenário de dados híbrido e multinuvel. Implemente uma malha de dados para permitir um tempo de retorno mais rápido para usuários de negócios, maior produtividade para engenharia de dados e operações e maior controle e fidelidade de conformidade.

[Para começar, experimente o IBM Cloud Pak for Data gratuitamente.](#)

[Agende uma consulta individual gratuita com um representante IBM Data and AI.](#)



© Copyright IBM Corporation 2021

IBM Brasil Ltda
Rua Tutóia, 1157
CEP 04007-900
São Paulo – SP
Brasil

Produzido nos Estados Unidos da América
Maio de 2021

IBM, o logotipo IBM e IBM Cloud Pak são marcas comerciais ou registradas da International Business Machines Corporation nos Estados Unidos e/ou em outros países. Outros nomes de produtos e de serviços podem ser marcas comerciais da IBM ou de outras empresas. Uma lista atual de marcas registradas da IBM em ibm.com/trademark.

Este documento estava atualizado na data de publicação inicial e pode ser alterado pela IBM a qualquer momento. Nem todas as ofertas estão disponíveis em todos os países onde a IBM opera.

Os dados de desempenho e exemplos de clientes citados são apresentados apenas para fins ilustrativos. Os resultados de desempenho reais poderão variar, dependendo das configurações e das condições operacionais específicas. É responsabilidade do usuário avaliar e verificar a operação de quaisquer outros produtos ou programas com produtos e programas IBM. AS INFORMAÇÕES NESTE DOCUMENTO SÃO OFERECIDAS NO ESTADO EM QUE SE ENCONTRAM ("AS IS") SEM QUALQUER GARANTIA, EXPLÍCITA OU IMPLÍCITA, INCLUINDO SEM QUALQUER GARANTIAS DE COMERCIALIZABILIDADE, ADEQUAÇÃO A UM PROPÓSITO ESPECIAL E QUALQUER GARANTIA OU CONDIÇÃO DE NÃO VIOLAÇÃO. Os produtos IBM são garantidos de acordo com os termos e condições dos acordos sob os quais são fornecidos.

Declaração de Boas Práticas de Segurança: a segurança do sistema de TI envolve a proteção dos sistemas e informações através da prevenção, detecção e resposta ao acesso indevido de dentro e de fora de sua empresa. O acesso incorreto pode resultar em alteração, destruição, desapropriação e mau uso de informações ou pode resultar em danos ou mau uso de seus sistemas, incluindo ataques a outras pessoas. Nenhum sistema ou produto de TI deve ser considerado completamente seguro e nenhum produto, serviço ou medida de segurança pode ser completamente efetivo para evitar uso ou acesso indevidos. Os sistemas, produtos e serviços IBM foram projetados para fazer parte de uma abordagem de segurança legítima e abrangente, a qual necessariamente envolve procedimentos operacionais adicionais e pode exigir que outros sistemas, produtos ou serviços sejam mais efetivos. A IBM NÃO GARANTE QUE TODOS OS SISTEMAS, PRODUTOS OU SERVIÇOS ESTEJAM LIVRES DE, OU QUE TORNARÃO A SUA EMPRESA LIVRE DE CONDUTA MALICIOSA OU ILEGAL DE QUALQUER PARTE.

O cliente é responsável por garantir a conformidade com as leis e regulamentações aplicáveis. A IBM não fornece conselhos jurídicos e não declara ou garante que seus serviços ou produtos irão assegurar que o cliente está em conformidade com qualquer lei ou regulamento.

- 1 A Forrester New Technology: estudo Projected Total Economic Impact™ encomendado pela IBM, fevereiro de 2020
- 2 As medições de desempenho foram coletadas no âmbito de um ambiente controlado de teste no IBM Silicon Valley Labs usando virtualização de dados IBM com relação a várias fontes de dados de 100TB. As medições feitas em maio de 2019 e os ganhos de desempenho são comparados com os da federação IBM.
- 3 Baseado na experiência de um cliente da IBM.