

La arquitectura del data fabric ofrece 3 beneficios inmediatos

Integre y aproveche todos los datos de
su organización para obtener mejores
resultados de negocio

Contenidos

- 3** Los desafíos de los datos actualmente
- 5** Beneficios de negocio de un data fabric
- 6** Cómo IBM ofrece un data fabric
- 7** Data fabric en acción
- 7** El data fabric es el siguiente paso transformador para su empresa

Los desafíos de los datos actualmente

Los datos son un elemento integral de la transformación digital de las empresas.

Pero a medida que las organizaciones tratan de aprovechar sus datos, se encuentran con desafíos provocados por la diversidad de fuentes, tipos, estructuras, entornos y plataformas de datos. Este predicamento de datos multidimensionales se complica aún más cuando las organizaciones adoptan arquitecturas híbridas y multinube. Para muchas empresas actualmente, los datos operativos han permanecido en gran medida aislados y ocultos, lo que ha dado lugar a una enorme cantidad de datos oscuros.

Tomemos el ejemplo de una empresa norteamericana de energía que buscaba reinventarse como una organización impulsada por los datos para convertirse en una empresa donde las capacidades de la ciencia de datos fueran fácilmente accesibles en todas las unidades de negocio. Pronto se dieron cuenta de que su transformación digital se veía obstaculizada por datos aislados, herramientas incoherentes y diversos niveles de habilidades, todo eso provocando brechas críticas entre las competencias de los datos.

El problema al que se enfrentaban no era exclusivo de su negocio; de hecho, es una consecuencia común de los paisajes de datos que han superado sus arquitecturas de gestión de datos.

Lo que finalmente llevó a la empresa de energía de vuelta a un camino exitoso de transformación digital fue emplear un nuevo concepto de arquitectura de datos identificado como data fabric.

¿Qué es exactamente un data fabric, en qué se diferencia de las arquitecturas anteriores, qué puede lograr para las empresas y cuál es el papel de IBM en su implementación? Este documento responderá a esas preguntas.

¿Qué es un data fabric?

En el pasado, las organizaciones han intentado solucionar los problemas de acceso a datos, ya sea a través de la integración punto a punto o de la introducción de hubs de datos. Ninguno de ellos es adecuado cuando los datos están muy distribuidos y aislados. Las integraciones punto a punto añaden un costo exponencial para cualquier punto final adicional que haya que conectar, lo que significa que es un enfoque no escalable. Los hubs de datos permiten una integración más fácil de las aplicaciones y fuentes, pero aumentan el costo y la complejidad para mantener la calidad y la confianza de los datos dentro del hub.

El data fabric es una arquitectura emergente que pretende sabordar los desafíos de datos que surgen de un escenario de datos híbrido. Su idea fundamental es lograr un equilibrio entre la descentralización y la globalización actuando como el tejido conectivo virtual entre los puntos finales de datos (vea la página 4).

A través de tecnologías como la automatización y el aumento de la integración, la gobernanza federada, así como la activación de metadatos, una arquitectura de data fabric permite la orquestación dinámica e inteligente de datos a través de un escenario distribuido, creando una red de información disponible al instante para impulsar un negocio.

Un data fabric es agnóstico a plataformas de implementación, procesos de datos, ubicaciones geográficas y enfoque arquitectónico. Facilita la utilización de datos como un activo de la empresa. Un data fabric garantiza que sus distintos tipos de datos puedan combinarse, accederse y gobernarse con éxito de forma eficiente y eficaz.

Capacidades y principios de un data fabric

El núcleo de la arquitectura data fabric es una plataforma de gestión de datos que permite toda la amplitud de funcionalidades integradas de gestión de datos, que incluyen descubrimiento, gobernanza, curadoría y orquestación.

Sin embargo, un data fabric avanza y evoluciona desde los conceptos tradicionales de gestión de datos como DataOps, el cual solo se centra en establecer prácticas, para incrementar el nivel de operatividad de los datos. Está desarrollado sobre una arquitectura distribuida y una tecnología avanzada capaz de atender las necesidades que surgen de la extrema diversidad y la distribución de los activos de datos.

Un data fabric podría dividirse lógicamente en cuatro funcionalidades (o componentes):

Conocimientos, insights y semántica

- Proporciona un mercado de datos y experiencia de compra
- Enriquece automáticamente los activos de datos descubiertos con conocimientos y semántica, permitiendo a los consumidores encontrar y entender los datos

Gobernanza y cumplimiento unificados

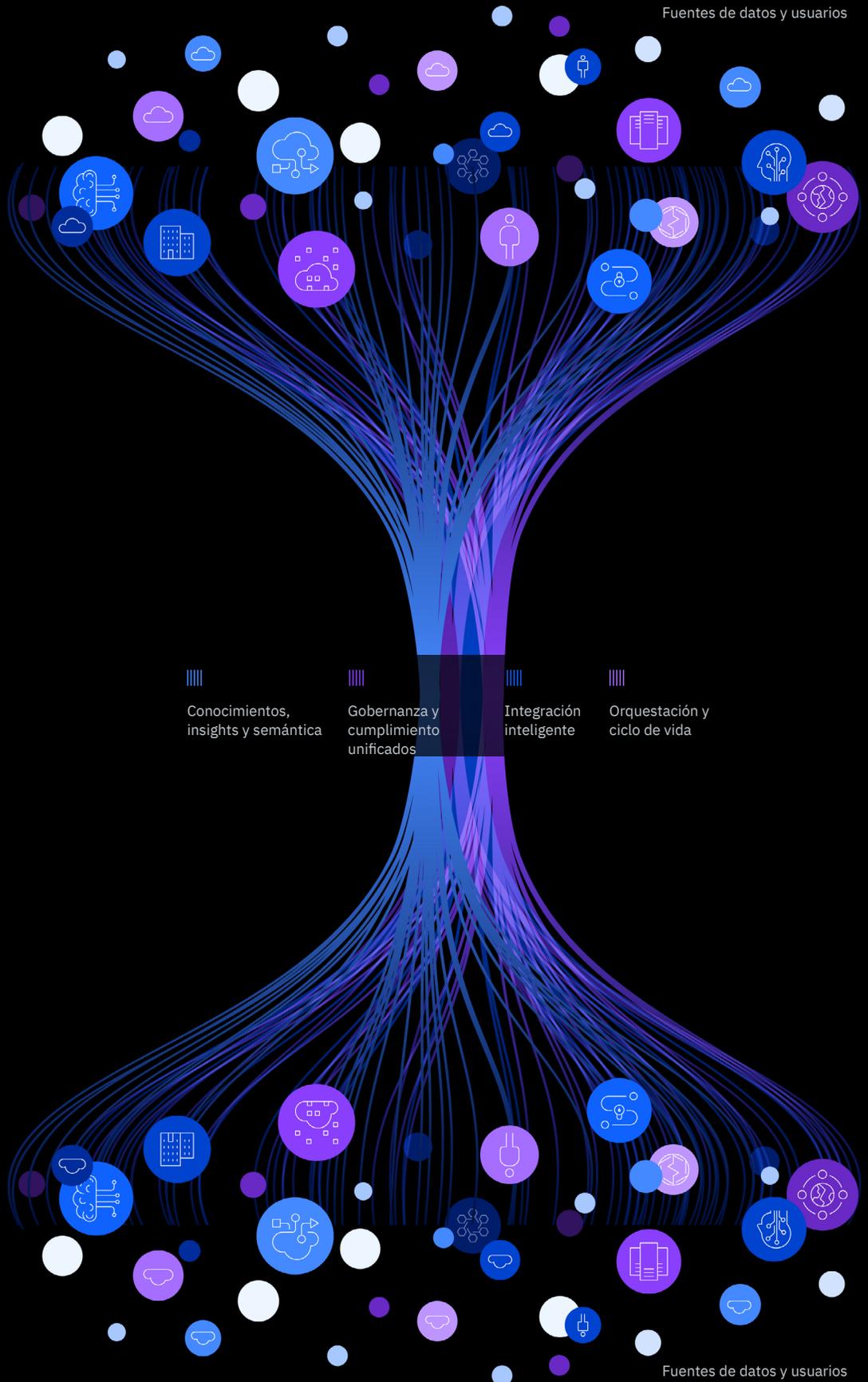
- Permite la gestión local y la gobernanza de metadatos, pero admite una visión global unificada y la aplicación de políticas
- Aplica automáticamente las políticas sobre los activos de datos de acuerdo con las normas globales y locales
- Utiliza funcionalidades avanzadas para automatizar la clasificación y curación de los activos de datos
- Establece automáticamente rutas de acceso consultables para cualquier activo catalogado para una mayor activación de los datos

Integración inteligente

- Acelera las tareas de los ingenieros de datos mediante el flujo automático y la creación de canalizaciones en fuentes de datos distribuidas
- Permite el autoservicio en ingesta y acceso de datos sobre cualquier dato con una aplicación profunda de las políticas de protección de datos a nivel local y global
- Determina automáticamente la ejecución que mejor funciona a través de la distribución optimizada de la carga de trabajo y el autoajuste y la corrección de las desviaciones del esquema

Orquestación y ciclo de vida

- Permite la composición, prueba, operación y monitoreo de los canales de datos
- Incorpora capacidades de IA en el ciclo de vida de los datos para automatizar tareas, autoajustar, autoreparar y detectar cambios en los datos de origen, todo lo cual facilita las actualizaciones automáticas



Beneficios de negocio de un data fabric

Los datos solo entregan valor de negocio cuando están contextualizados y son accesibles para cualquier usuario o aplicación de la organización. Cuando se implementa correctamente, un data fabric ayuda a asegurar que esos valores estén disponibles en toda la organización de la manera más eficiente y automática posible. Por eso, el fabric posee tres beneficios clave:

1. Habilitar el autoservicio en el consumo de datos y la colaboración
2. Automatizar la gobernanza, la protección y la seguridad habilitada por metadatos activos
3. Automatizar las tareas de ingeniería de datos y aumentar la integración de datos a través de los recursos de nube híbrida

Habilitar el autoservicio en el consumo de datos y la colaboración

Al integrar datos de varias fuentes y analizar una mayor fracción de la enorme cantidad de datos que se generan a diario, las organizaciones obtienen mejores insights y responden con mayor velocidad a las cambiantes demandas del negocio. Un data fabric pone rápidamente los datos en manos de quienes los necesitan. El autoservicio permite a la organización como un todo encontrar los datos apropiados más rápidamente y dedicar más tiempo a utilizar esos datos para proporcionar insights tangibles.

Beneficios del data fabric para el autoservicio en el consumo de datos:

- Los usuarios de la empresa tienen un único punto de acceso para encontrar, entender, dar forma y consumir datos en toda la organización.
- Una gobernanza centralizada de los datos y su linaje ayuda a los usuarios a entender qué significan los datos, de dónde proceden y cómo se relacionan con otros activos.
- La gestión de metadatos amplia y personalizable se escala fácilmente y es accesible a través de las API.
- El acceso de autoservicio a datos confiables y gobernados permite la colaboración de la línea de negocio con otros usuarios.

Un estudio de Forrester sobre el Impacto Económico total¹ reveló que estas funcionalidades pueden significar:

USD
5.8M en beneficios
ROI
459 %

Automatizar la gestión, la protección de datos y la seguridad gracias a los metadatos activos

Una capa de gestión activa distribuida para todas las iniciativas de datos reduce los riesgos de cumplimiento y regulatorios al proporcionar confianza y transparencia. Permite el cumplimiento automático de las políticas para cualquier acceso a los datos, proporcionando un alto nivel de protección y cumplimiento normativo de los datos.

La utilización de tecnologías de IA y machine learning permite a los usuarios de data fabric incrementar su nivel de automatización, por ejemplo extrayendo automáticamente reglas de gobernanza de datos basadas en el idioma y las definiciones de los documentos regulatorios. Esto permite a las organizaciones aplicar reglas de gobernanza específicas de la industria en cuestión de minutos para ayudar a evitar costosas multas y asegurar un uso ético de los datos dondequiera que estos residan.

Beneficios de un data fabric para la virtualización gobernada:

- La agilidad, la seguridad y la productividad aumentan para los ingenieros de datos, los científicos de datos y los analistas de negocio.
- Múltiples fuentes de datos globales aparecen como una sola base de datos.
- Un nuevo descubrimiento, líder en la industria, de información personal identificable (PII) y elementos de datos críticos, es posible en escala masiva.

Estas funcionalidades pueden significar:

USD
2.4M en beneficios¹
430 % mejora del rendimiento²

Automatizar las tareas de ingeniería de datos y aumentar la integración de datos

La ingeniería de datos avanzada significa que prácticamente cualquier proceso de acceso o entrega de datos es automático y no requiere ningún proceso de codificación tedioso o propenso a errores. El aumento de la integración utiliza datos de metadatos para optimizar la entrega y el acceso a los datos.

Beneficios de un data fabric para la ingeniería e integración de datos:

- La integración de datos optimizada automáticamente ayuda a acelerar la entrega de datos.
- El equilibrio automático de las cargas de trabajo y la escala flexible significa que los trabajos están preparados para cualquier entorno y cualquier volumen de datos.
- La resiliencia y la automatización CI/CD están integradas.
- El proceso automatizado para capturar los cambios en tiempo real apoya la entrega de datos de calidad para los procesos de negocio.
- El machine learning puede automatizar y ampliar los procesos de descubrimiento, clasificación y curadoría de datos personalizados, lo que permite crear valor más rápidamente.
- El análisis continuo puede realizarse automáticamente en tiempo real, dondequiera que estén los datos.

Los resultados de un minorista líder³ muestran:

60 veces aceleración en el tiempo de entrega de datos
20 veces análisis más rápidos de la afinidad de la clientela

Cómo IBM ofrece un data fabric

IBM Cloud Pak® for Data hace posible este concepto de data fabric. IBM Cloud Pak for Data es una plataforma de insights que simplifica y automatiza la recopilación, organización y análisis de datos y acelera la infusión de IA en todo su negocio.

Con sus funcionalidades para conectar datos en cualquier lugar; ejecutar cargas de trabajo en cualquier lugar; y construir, implementar y administrar IA a escala en entornos de nube híbrida, IBM Cloud Pak for Data es el habilitador para una transformación digital del negocio.

La plataforma ofrece integración sin interrupciones en una empresa híbrida para:

- IBM Cloud Pak for Data services
- Aplicaciones externas y fuentes de datos
- Funcionalidades avanzadas basadas en IA para la gestión y la gobernanza de los datos

Esta base pone los datos curados a disposición de los consumidores con el equilibrio óptimo de costo, rendimiento y cumplimiento normativo, y con la inteligencia para orquestar y optimizar el procesamiento de los datos en función de las cargas de trabajo, la localización de los datos y las políticas de datos.

IBM Cloud Pak for Data aprovecha las siguientes funcionalidades para entregar los datos listos para el negocio que exige un data fabric. Todas estas funcionalidades desempeñan un papel en el soporte a la arquitectura del data fabric.

1. Núcleo de conocimiento basado en metadatos

Facilita el descubrimiento de fuentes y catálogos de datos; enriquece los activos de datos; y realiza análisis para extraer insights para una mayor automatización utilizando IA. El núcleo de conocimiento se utiliza para dar poder al mercado con la búsqueda semántica.

2. Autoservicio del mercado de datos

El catálogo de datos de última generación que ayuda a los consumidores de datos, como los analistas de negocio, a recuperar datos de todo el panorama de datos de la empresa.

3. Integración inteligente

Permite el consumo de datos mediante la extracción, virtualización, transformación y transmisión. Se integra con el núcleo de conocimiento para automatizar la integración de datos, y tiene la inteligencia para decidir qué enfoque de integración es el más adecuado en función de las cargas de trabajo y las políticas de datos. También puede utilizarse para la preparación de datos como rol de cargas de trabajo de ingeniería de datos o para crear productos de datos. Por último, ofrece la posibilidad de publicar las actualizaciones de los productos de datos.

4. La gobernanza

cataloga y conserva los metadatos, define las políticas de privacidad de los datos, realiza la curaduría, captura su linaje y realiza otras tareas relacionadas con la seguridad y el cumplimiento normativo. Esta capa entiende el formato de los datos (por ejemplo, estructurados versus no estructurados) y el significado de los datos (por ejemplo, públicos versus protegidos) y aplica las políticas correctas a cada bit de datos y a cada posible usuario. En lugar de aplicar manualmente las normas y reglas a los datos, esta capacidad de la plataforma integrada permite aplicarlas a nivel organizacional y propagarlas a través de los distintos recursos de datos según sea necesario. Los modelos analíticos de diferentes herramientas pueden dialogar entre sí; la aplicación de políticas de datos a nivel granular puede ser automatizada en gran medida.

5. Operaciones y desarrollo unificados

Permite un ciclo de vida unificado para configurar y ejecutar todos los aspectos de la plataforma de datos en producción.



Figura 1. Las funcionalidades de la plataforma IBM Cloud Pak for Data, las cuales soportan el desarrollo de IA y del data fabric

Data fabric en acción

La mejor manera de entender el valor de un data fabric es ver los beneficios de negocio que organizaciones reales han logrado implementando un data fabric sobre la base de una plataforma de insights.

Empresa de energía

En la sección de desafíos, al principio de este documento, se presentó el caso de una empresa de energía norteamericana que está experimentando una transformación digital.

Este cliente se comprometió con IBM, que implementó una arquitectura de data fabric basada en IBM Cloud Pak for Data.

Con una estructura de datos flexible e integrada, la empresa ha podido poner en marcha una serie de importantes proyectos de datos en diferentes unidades, entre ellos:

- eMobility
- Descubrimiento de documentos de operaciones de gas, incluyendo la extracción de texto escrito a mano.
- Segmentación de clientes del área eléctrica y previsión de cargas
- Gestión de activos
- Impactos de la carga asociada al COVID-19
- Modelo de riesgo de reincorporación al trabajo

Con un data fabric, esta organización es capaz de:

- Proporcionar a varias unidades de negocio el acceso directo a los datos a través de una plataforma de insights de autoservicio.
- Utilizar la capacidad de computación adecuada para ejecutar modelos de manera eficiente en miles de millones de filas de datos.
- Desarrollar modelos en colaboración e implementar fácilmente esos modelos para integrar insights en toda la empresa.

Reducir los desplazamientos y mejorar la supervisión

Un gran proveedor de servicios industriales mejoró la gobernanza de los datos y facilitó el cumplimiento normativo.

Esta organización tuvo problemas para trasladar grandes cantidades de datos a su lago de datos basado en la nube. Tenían los requisitos gemelos de aplicar una sólida capa de gobernanza a cada instancia de acceso a los datos y de determinar la calidad de los datos antes de proporcionar acceso al usuario de la empresa.

Mediante la implementación de un data fabric construido sobre IBM Cloud Pak for Data, esta organización logró enormes mejoras en sus procesos de gobernanza, cumplimiento normativo y transformación de datos. Con una estructura de datos establecida, pueden proporcionar a los usuarios del negocio un acceso fácil y seguro a cientos de fuentes de datos en su lago de datos basado en la nube y en las fuentes de datos de SAP.

También están aprovechando las herramientas del acelerador de la industria regulatoria para explorar las fuentes de datos en busca de posibles datos PII que estarían sujetos a la regulación del GDPR y la CCPA. El marcado automático de datos ayuda a identificar cualquier fuente de datos que necesite ser gestionada para el manejo de PII tanto on premises como en la nube.

Con un mejor acceso y una selección más rápida de la información personal de identidad, los usuarios pueden extraer datos relevantes para obtener insights importantes sin esperar mucho tiempo para acceder a los datos o correr el riesgo de exponer datos protegidos.

Con un data fabric, esta organización es capaz de:

- Garantizar la gobernanza adecuada de los datos, al mismo tiempo que se aprovechan los datos de toda la empresa.
- Obtener datos confiables y reducir la cantidad de preparación de datos.
- Ayudar en el cumplimiento de las normativas de privacidad como GDPR y CCPA.

El data fabric es el siguiente paso transformador para su empresa

Con un data fabric construido sobre la tecnología IBM Cloud Pak for Data, usted puede hiper-automatizar el descubrimiento de datos, la gobernanza de datos y el consumo de datos en un escenario de datos híbrido y multinube. Emplear un data fabric para habilitar una creación de valor más rápida para los usuarios del negocio, mayor productividad para la ingeniería de datos y las operaciones, y mayor gobernanza y fidelidad de cumplimiento normativo.

[Para empezar, pruebe gratis IBM Cloud Pak for Data.](#)

[Agende una consulta individual gratis con un representante de datos e IA de IBM.](#)



© Copyright IBM Corporation 2021

IBM Argentina
Pje. Ing. Enrique Butty 275
C.A.B.A – Argentina

IBM, el logotipo de IBM, e IBM Cloud son marcas comerciales o marcas comerciales registradas de International Business Machines Corporation, en los Estados Unidos y/o en otros países. Otros nombres de productos y servicios pueden ser marcas registradas de IBM o de otras compañías. Puede consultarse una lista actualizada de las marcas registradas de IBM en ibm.com/trademark.

Este documento está actualizado conforme a la fecha inicial de la publicación y puede ser modificado por IBM en cualquier momento. No todas las ofertas están disponibles en todos los países donde opera IBM.

Los ejemplos de clientes y los datos de rendimiento citados se presentan solo para fines ilustrativos. Los resultados de rendimiento real pueden variar en función de las configuraciones específicas y las condiciones de funcionamiento. Es responsabilidad del usuario evaluar y verificar el funcionamiento de cualquier otro producto o programa con productos y programas de IBM. LA INFORMACIÓN CONTENIDA EN ESTE DOCUMENTO SE PROPORCIONA "TAL CUAL" SIN NINGUNA GARANTÍA, EXPRESA O IMPLÍCITA, NO INCLUYE CUALQUIER GARANTÍA DE CAPACIDAD DE COMERCIALIZACIÓN, IDONEIDAD PARA UN PROPÓSITO EN PARTICULAR NI CUALQUIER GARANTÍA O CONDICIÓN DE NO VIOLACIÓN. Los productos de IBM están garantizados de acuerdo con los términos y condiciones de los acuerdos bajo los cuales se proporcionan.

Declaración de buenas prácticas de seguridad: la seguridad de los sistemas de TI implica la protección de los sistemas y la información mediante la prevención, detección y respuesta al acceso inadecuado desde dentro y fuera de su empresa. El acceso inadecuado puede tener como consecuencia la alteración, destrucción, apropiación indebida o uso indebido de la información, o puede traducirse en daños o uso indebido de sus sistemas, incluso para atacar a terceros. Ningún sistema o producto de TI debe considerarse completamente seguro y ningún producto, servicio o medida de seguridad puede ser completamente eficaz para evitar el uso o acceso indebido. Los sistemas, productos y servicios de IBM están diseñados para ser parte de un enfoque de seguridad integral y legal, que necesariamente implicará procedimientos operativos adicionales, y pueden requerir que otros sistemas, productos o servicios sean más efectivos. IBM NO GARANTIZA QUE LOS SISTEMAS, PRODUCTOS O SERVICIOS SEAN INMUNES A LA CONDUCTA MALICIOSA O ILEGAL DE TERCEROS, NI QUE LOS MISMOS OTORGUEN INMUNIDAD A SU EMPRESA DE ESTAS AMENAZAS.

El cliente es responsable de garantizar el cumplimiento de las leyes y reglamentaciones que le sean aplicables. IBM no proporciona asesoramiento legal ni declara o garantiza que sus servicios o productos garantizarán que el cliente cumpla con cualquier ley o reglamento.

- 1 New Technology: Projected Total Economic Impact™, un estudio de Forrester a pedido de IBM, febrero de 2020
- 2 Las mediciones de rendimiento se recopilaron en un entorno de prueba controlado en los laboratorios de IBM de Silicon Valley, utilizando la virtualización de datos de IBM contra varias fuentes de datos de 100TB. Las mediciones fueron realizadas en mayo de 2019 y las ganancias de rendimiento se comparan con la federación IBM.
- 3 Basado en la experiencia de un cliente de IBM.