

# La arquitectura de datos Data Fabric ofrece 3 beneficios inmediatos

Integre y relacione todos los datos de su organización para obtener mejores resultados empresariales

## **Contenido**

- 3** Los retos actuales de los datos
- 5** Beneficios empresariales de un data fabric
- 6** Cómo IBM ofrece un Data Fabric
- 7** Data Fabric en acción
- 7** Data Fabric es el siguiente paso para transformar su empresa

# Los retos actuales de los datos

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

Pero a medida que las organizaciones tratan de optimizar sus datos, se enfrentan a desafíos derivados de la diversidad de orígenes, tipos, estructuras, entornos y plataformas de datos. Esta dificultad de los datos multidimensionales se complica aún más cuando las organizaciones adoptan arquitecturas híbridas y multicloud. En la actualidad, para muchas empresas los datos operativos han permanecido en gran medida compartimentados y ocultos, lo que ha generado una enorme cantidad de datos oscuros.

Tomemos el ejemplo de una empresa energética norteamericana que buscaba reinventarse como una organización impulsada por los datos, para convertirse en una empresa en la que las prestaciones de la ciencia de los datos fueran fácilmente accesibles en múltiples unidades de negocio. Pronto se dieron cuenta de que su transformación digital se veía obstaculizada por la compartimentación de datos, herramientas incoherentes y distintos niveles de formación, todo esto provocaba unas lagunas importantes en las competencias en materia de datos.

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

Lo que finalmente devolvió a la empresa energética a la senda del éxito en la transformación digital fue el empleo de un nuevo concepto de arquitectura de datos conocido como data fabric.

¿Qué es exactamente un data fabric, en qué se diferencia de arquitecturas anteriores, qué beneficios representa para las empresas y cuál es el papel de IBM en su implementación? Este libro blanco responderá a estas preguntas.

## ¿Qué es un data fabric?

En el pasado, las organizaciones habían intentado resolver los problemas de acceso a los datos mediante la integración punto a punto o la introducción de concentradores de datos. Ninguna de estas soluciones es adecuada cuando los datos están altamente distribuidos y compartimentados. Las integraciones punto a punto añaden un coste exponencial por cada punto final adicional que deba conectarse, lo que significa que es un enfoque no escalable. Los concentradores de datos permiten una integración más fácil de las aplicaciones y los orígenes, pero exacerban el coste y la complejidad necesarios para mantener la calidad y la confianza de los datos del concentrador.

El data fabric es una arquitectura emergente que pretende dar respuesta a los retos de los datos que surgen de un panorama de datos híbridos. Su idea fundamental es lograr un equilibrio entre la descentralización y la globalización, actuando como tejido conectivo virtual entre los puntos finales de datos (véase la página 4).

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

Un data fabric es agnóstico con respecto a las plataformas de despliegue, los procesos de datos, las ubicaciones geográficas y el enfoque arquitectónico. Facilita el uso de los datos como activo empresarial. Un data fabric garantiza que puedan combinarse con éxito diversos tipos de datos y se pueda acceder a ellos y gobernarlos de forma eficiente y también eficaz.

## Capacidades y principios de un data fabric

El núcleo de la arquitectura de data fabric es una plataforma de gestión de datos que permite la gama completa de las prestaciones de gestión de datos integradas, incluidos el descubrimiento, el gobierno, la conservación y la orquestación

Además, un data fabric representa un avance y una evolución con respecto a los conceptos tradicionales de gestión de datos, como DataOps, que sólo se centra en establecer prácticas para aumentar el nivel de operatividad de los datos. Se basa en una arquitectura distribuida y una tecnología avanzada capaces de responder a las necesidades derivadas de la extrema diversidad y distribución de los activos de datos.

Un data fabric puede dividirse lógicamente en cuatro prestaciones (o componentes):

### Conocimientos, información importante y semántica

- Proporciona un mercado de datos y una experiencia de compra
- Enriquece automáticamente los activos de datos descubiertos con conocimientos y semántica, lo que permite a los consumidores encontrar y comprender los datos

### Gobierno y conformidad unificados

- Permite la gestión local y el gobierno de los metadatos, pero admite una visión global unificada y la aplicación de políticas
- Aplica automáticamente políticas a los activos de datos de acuerdo con las normas globales y locales
- Utiliza prestaciones avanzadas para automatizar la clasificación y la conservación de los activos de datos
- Establece automáticamente rutas de acceso consultables para cualquier activo catalogado para aumentar la activación de los datos

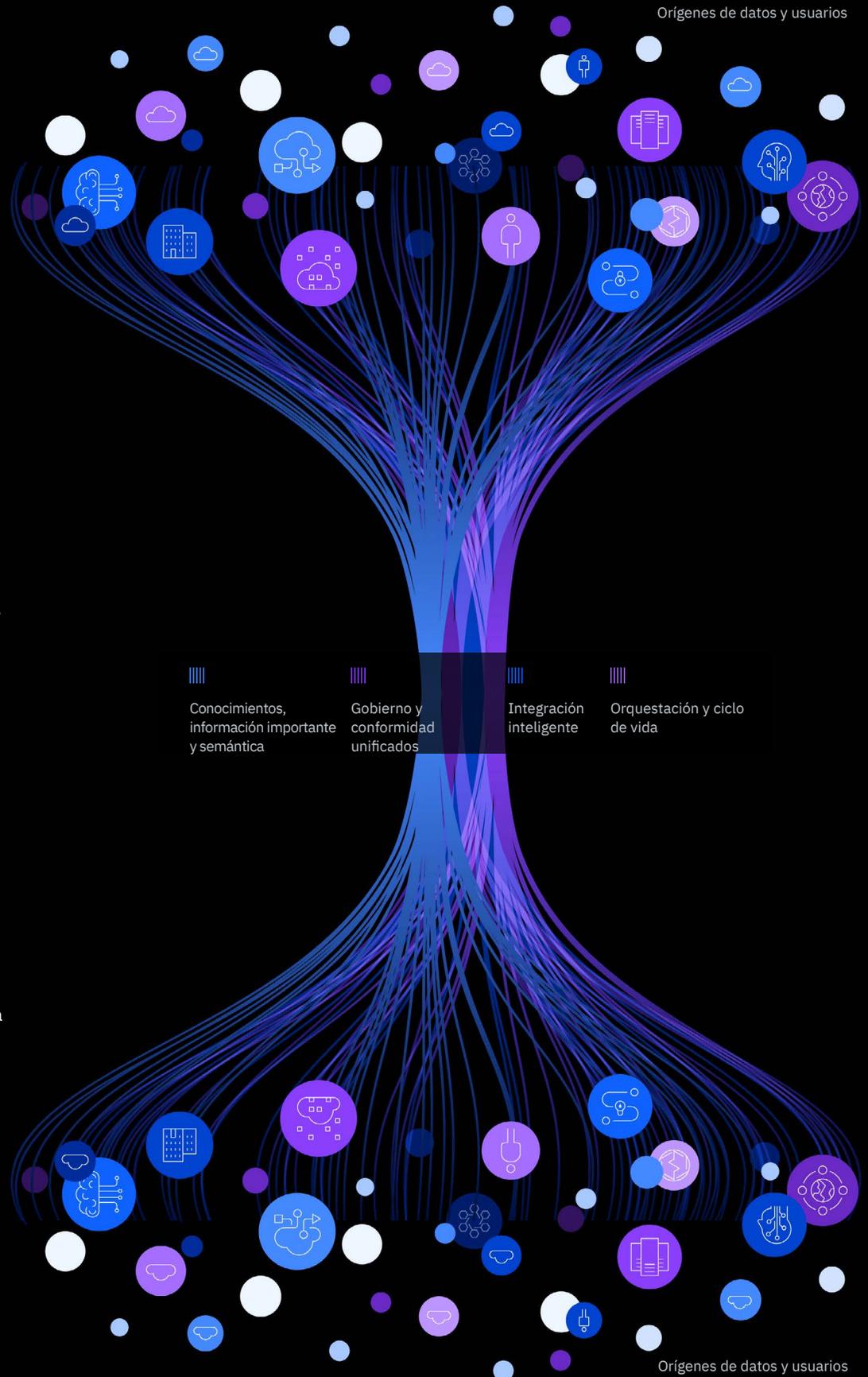
### Integración inteligente

- Acelera las tareas de los ingenieros de datos mediante el flujo automatizado y la creación de interconexiones entre orígenes de datos distribuidos
- Permite la ingesta de autoservicio y el acceso a cualquier dato con una aplicación profunda de las políticas de protección de datos a nivel local y global
- Determina automáticamente la ejecución más adecuada mediante 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, las pruebas, la operación y la supervisión de los conductos de datos
- Introduce prestaciones 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 automatizadas

Orígenes de datos y usuarios



# Beneficios empresariales de un data fabric

Los datos sólo aportan valor empresarial cuando están contextualizados y son accesibles para cualquier usuario o aplicación de la organización. Si se implementa correctamente, un data fabric ayuda a garantizar que esos valores estén disponibles en toda la organización de la manera más eficiente y automatizada posible. Por ello, el fabric aporta tres beneficios clave:

1. Permitir el consumo de datos en autoservicio y la colaboración
2. Automatizar el gobierno, la protección y la seguridad, gracias a los metadatos activos
3. Automatizar las tareas de ingeniería de datos y aumentar la integración de datos entre recursos de nube híbrida

## Permitir el consumo de datos en autoservicio y la colaboración

Al integrar datos procedentes de múltiples orígenes y analizar una mayor fracción de la enorme cantidad de datos que se generan a diario, las organizaciones obtienen un mayor conocimiento y responden más rápidamente a las cambiantes demandas del negocio. Un tejido de datos pone rápidamente los datos en manos de quienes los necesitan. El autoservicio permite a la organización en su conjunto encontrar los datos apropiados con mayor rapidez y dedicar más tiempo a utilizar esos datos para proporcionar información tangible.

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

- Los usuarios de la empresa tienen un único punto de acceso para encontrar, comprender, dar forma y consumir datos en toda la organización.
- Un gobierno y un linaje de datos centralizados ayudan 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 amplía fácilmente y es accesible por medio de las API.
- El acceso de autoservicio a datos gobernados y de confianza permite la colaboración de la línea de negocio con otros usuarios.

Un estudio de Forrester sobre el impacto económico total<sup>1</sup> reveló que estas prestaciones pueden significar:

USD  
**5,8M** de beneficios      **459 %** de rendimiento de la inversión

## Automatizar el gobierno, la protección y la seguridad, gracias a los metadatos activos

Un nivel de gobierno activo distribuido para todas las iniciativas de datos reduce los riesgos de incumplimiento y normativos al proporcionar confianza y transparencia. Permite la aplicación automática de políticas para cualquier acceso a datos, proporcionando a un alto nivel de protección de datos y de cumplimiento.

La utilización de tecnologías de IA y aprendizaje automático permite a los usuarios del data fabric aumentar su nivel de automatización, por ejemplo, extrayendo automáticamente reglas de gobierno de datos basadas en el idioma y las definiciones de documentos normativos. Esto permite a las organizaciones aplicar reglas de gobierno específicas del sector en cuestión de minutos para ayudar a evitar costosas multas y garantizar un uso ético de los datos dondequiera que residan.

Ventajas de un data fabric para la virtualización gobernada:

- Se aumenta la agilidad, seguridad y productividad para los ingenieros de datos, los científicos de datos y los analistas de negocio.
- Varios orígenes de datos globales aparecen como una sola base de datos.
- Es posible realizar nuevos descubrimientos, líderes del sector, de información de identificación personal (PII) y elementos de datos críticos a escala masiva.

Estas prestaciones pueden significar:

USD  
**2,4M** de beneficios<sup>1</sup>      **430 %** de mejora del rendimiento<sup>2</sup>

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

La ingeniería de datos avanzada implica que prácticamente cualquier proceso de acceso a datos o entrega de los mismos está automatizado 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 de datos y su acceso.

Ventajas 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 la carga de trabajo y el escalado elástico permiten que los trabajos estén preparados para cualquier entorno y volumen de datos.
- La resiliencia y la automatización de CI/CD están integradas.
- El proceso automatizado de captura de cambios en tiempo real favorece la entrega de datos de calidad para los procesos empresariales.
- El aprendizaje automático puede automatizar y ampliar los procesos de descubrimiento, clasificación y conservación de datos personalizados, lo que permite obtener valor más rápidamente.
- El análisis continuo puede realizarse automáticamente en tiempo real, dondequiera que estén los datos.

Los resultados de un minorista líder<sup>3</sup> muestran:

**60X** aceleración en el tiempo de entrega de datos      **20X** análisis más rápidos de la afinidad de los clientes

# 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 conocimiento que simplifica y automatiza la recopilación de datos, la organización y el análisis de los mismos y acelera la introducción de IA en toda la empresa.

Con sus prestaciones para conectar datos en cualquier lugar; ejecutar cargas de trabajo en cualquier lugar; y crear, desplegar y gestionar la IA a escala en entornos de nube híbrida, IBM Cloud Pak for Data es el facilitador para una transformación digital empresarial.

La plataforma ofrece una integración perfecta en una empresa híbrida para:

- Servicios de IBM Cloud Pak for Data
- Aplicaciones y orígenes de datos externos
- Capacidades avanzadas basadas en la IA para la gestión y el gobierno de los datos

Esta base pone los datos recuperados a disposición de los consumidores con un equilibrio óptimo entre coste, rendimiento y cumplimiento y con la inteligencia necesaria para orquestar y optimizar el proceso de 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 prestaciones para ofrecer los datos listos para el negocio que exige un data fabric. Todas estas prestaciones desempeñan un papel en el apoyo a la arquitectura del data fabric.

**1. Núcleo de conocimiento basado en metadatos** Facilita el descubrimiento de orígenes y catálogos de datos; enriquece los activos de datos; y realiza análisis para extraer información importante para una mayor automatización mediante la IA. El núcleo de conocimiento se utiliza para potenciar el mercado con la búsqueda semántica.

2. **Mercado de datos de autoservicio** El catálogo de datos de próxima 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 su 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 parte de las cargas de trabajo de ingeniería de datos o para crear productos de datos. Por último, ofrece la posibilidad de publicar actualizaciones de productos de datos.
4. **Gobierno** Cataloga y conserva los metadatos, define las políticas de privacidad de los datos, conserva los datos, captura el linaje de los datos y realiza otras tareas relacionadas con la seguridad y el cumplimiento. Esta nivel entiende el formato de los datos (por ejemplo, estructurados frente a no estructurados) y la importancia de los datos (por ejemplo, públicos frente a protegidos) y aplica las políticas correctas a cada bit de datos y a cada posible usuario. En lugar de aplicar manualmente los estándares y las reglas a los datos, esta capacidad de la plataforma integrada permite aplicarlas a nivel organizativo y propagarlas por los distintos recursos de datos, según sea necesario. Los modelos analíticos de diferentes herramientas pueden comunicarse entre sí; la aplicación de políticas de datos a nivel granular se puede automatizar en gran medida.
5. **Desarrollo y operaciones unificadas** Permite un ciclo de vida unificado para configurar y ejecutar todos los aspectos de la plataforma de datos en producción.



Figura 1. Las prestaciones de la plataforma IBM Cloud Pak for Data, que dan soporte al desarrollo de la IA y al tejido de datos

# Data fabric en acción

La mejor manera de entender el valor de un data fabric es ver los beneficios empresariales que las organizaciones reales han conseguido gracias a la implementación de un data fabric sobre una plataforma de conocimiento.

## Empresa energética

En la sección de retos al principio de este libro blanco se ha presentado el caso de una empresa energética norteamericana que estaba 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 un data fabric e integrado, la empresa pudo poner en marcha una serie de importantes proyectos de datos en diferentes unidades, entre ellos:

- eMobility
- Descubrimiento de documentos de operaciones de gas, incluida la extracción de escritura manuscrita
- Segmentación de clientes eléctricos y predicción de cargas
- Gestión de activos
- Impactos en la carga por el COVID-19
- Modelo de riesgo de reincorporación al trabajo

### Con un data fabric, esta organización es capaz de:

- Proporcionar a múltiples unidades de negocio un acceso directo a los datos a través de una plataforma de conocimiento de autoservicio.
- Utilizar la capacidad de cálculo adecuada para ejecutar modelos de forma eficiente en miles de millones de filas de datos.
- Desarrollar modelos de forma colaborativa y desplegarlos fácilmente para introducir información importante en toda la empresa.

## Reducir los desplazamientos y mejorar la supervisión

Un gran proveedor de servicios industriales mejoró el gobierno de los datos y facilitó el cumplimiento de la normativa.

Esta organización tuvo problemas para trasladar grandes cantidades de datos a su lago de datos basado en la nube. Tenían el doble requisito de aplicar un estricto nivel de gobierno a cada instancia de acceso a los datos y determinar la calidad de los datos antes de proporcionar acceso a los usuarios empresariales.

Mediante la implementación de un data fabric basado en IBM Cloud Pak for Data, esta organización logró enormes mejoras en sus procesos de gobierno de datos, cumplimiento de datos y transformación de datos. Con un data fabric ya en marcha, pueden proporcionar a los usuarios empresariales un acceso fácil y seguro a cientos de orígenes de datos en su lago de datos basado en la nube y en orígenes de datos SAP locales.

También están aprovechando las herramientas de aceleración normativa del sector para examinar los orígenes de datos en busca de posibles datos PII que estarían sujetos a la regulación del GDPR y la CCPA. El etiquetado automático de los datos ayuda a identificar cualquier origen de datos que deba ser gestionado para el tratamiento de la PII, tanto en las instalaciones locales como en la nube.

Con un mejor acceso y una exploración más rápida de la información personal, los usuarios empresariales pueden extraer datos relevantes para obtener información importante sin tener que sufrir largos tiempos de espera para acceder a los datos o correr el riesgo de exponer datos protegidos.

### Con un data fabric, esta organización es capaz de:

- Garantizar un gobierno de datos adecuado, al mismo tiempo que se aprovechan los datos de toda la empresa.
- Obtener datos de confianza y reducir la cantidad de preparación de datos.
- Ayudar en el cumplimiento de las regulaciones de privacidad como, por ejemplo, el GDPR y la CCPA.

## El tejido de datos es el siguiente paso para transformar su empresa

Con un tejido de datos basado en la tecnología de IBM Cloud Pak for Data, puede hiperautomatizar el descubrimiento de datos, el gobierno de datos y el consumo de datos en un entorno de datos híbrido y multicloud. Emplee un tejido de datos para conseguir generar valor más rápidamente para los usuarios empresariales, una mayor productividad para las operaciones y la ingeniería de datos y una mayor fidelidad con el gobierno y el cumplimiento.

[Para empezar, pruebe IBM Cloud Pak for Data de forma gratuita.](#)

[Programa una consulta personal gratuita con un representante de IBM Data e IA.](#)



© Copyright IBM Corporation 2021

IBM España, S.A  
Tel.: +34-91-397-6611  
Santa Hortensia, 26-28  
28002 Madrid Spain

Produced in the United States of America  
Mayo de 2021

IBM, el logotipo de IBM e IBM Cloud son marcas comerciales o marcas registradas de International Business Machines Corporation, en los Estados Unidos y/o en otros países. Otros nombres de productos y de servicios pueden ser marcas registradas de IBM o de otras empresas. Una lista actual de las marcas registradas de IBM está disponible en [ibm.com/trademark](http://ibm.com/trademark).

Este documento es vigente en la fecha de publicación inicial y puede ser modificado en cualquier momento por IBM. No todas las ofertas están disponibles en todos los países en los que IBM opera.

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 los productos y programas de IBM. LA INFORMACIÓN PRESENTADA EN ESTE DOCUMENTO SE PROVEE "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO, NI EXPRESA NI IMPLÍCITA, INCLUYENDO, PERO NO LIMITADO A, LAS GARANTÍAS IMPLÍCITAS DE COMERCIO, CONVENIENCIA PARA UN PROPÓSITO PARTICULAR, O NO INFRACCIÓN. Los productos de IBM están garantizados según los términos y condiciones de los acuerdos bajo los que se proporcionan.

Declaración de buenas prácticas de seguridad: la seguridad del sistema de TI incluye la protección de sistemas e información a través de la prevención, detección y respuesta de acceso indebido desde el interior y exterior de su empresa. El acceso incorrecto puede tener como resultado que la información sea alterada, destruida, sustraída o mal utilizada o puede tener como resultado el daño o el mal uso de sus sistemas, incluyendo que sea utilizada en ataques hacia otros. Ningún producto o sistema de TI debería considerarse completamente seguro y ningún único producto, servicio o medida de seguridad puede ser completamente efectivo al prevenir el uso o acceso incorrecto. Los sistemas, productos y servicios de IBM están diseñados para ser parte de un enfoque de seguridad integral y legal, que necesariamente involucrará procedimientos operativos adicionales, y puede requerir otros sistemas, productos o servicios para ser más efectivo. IBM NO GARANTIZA QUE LOS SISTEMAS, PRODUCTOS O SERVICIOS SON INMUNES O HARÁN QUE SU EMPRESA SEA INMUNE A LA CONDUCTA MALINTENCIONADA O ILEGAL DE OTRA PARTE.

El cliente es responsable por garantizar el cumplimiento de las leyes y las regulaciones correspondientes. IBM no brinda asesoría legal o representa o garantiza que sus servicios o productos garantizarán que el cliente esté en conformidad con cualquier ley o regulación.

- 1 A Forrester New Technology: Projected Total Economic Impact™, Estudio encargado por IBM, Febrero de 2020
- 2 Las mediciones de rendimiento se recogieron en un entorno de pruebas controlado en los laboratorios de IBM en Silicon Valley, utilizando la virtualización de datos de IBM con varios orígenes de datos de 100 TB. Las mediciones realizadas en mayo de 2019 y las ganancias de rendimiento se comparan con la federación de IBM.
- 3 Basado en la experiencia de un cliente de IBM.