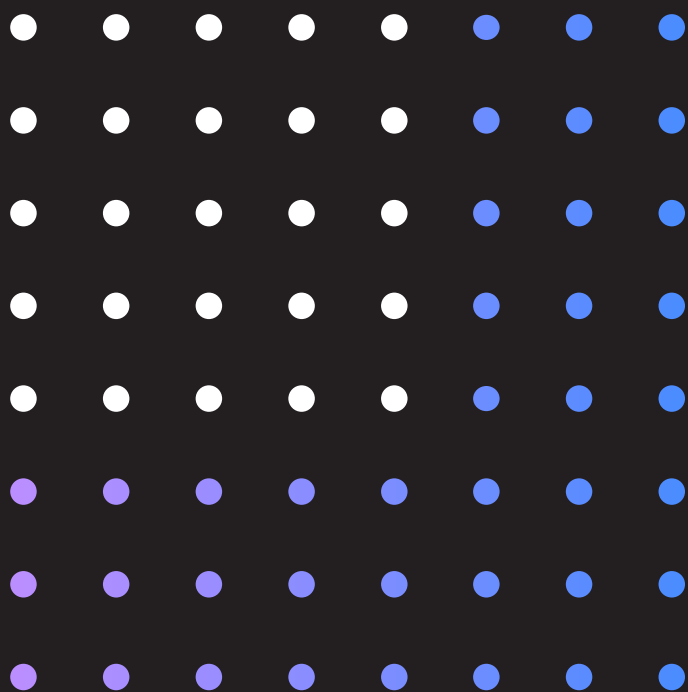


# Aporte datos listos para usarse en los negocios mediante la catalogación de información y la gobernanza del lago de datos

IBM Watson Knowledge Catalog ofrece una plataforma de gestión de datos basada en aprendizaje automático que ayuda con los desafíos del lago de datos



# Contenido

## 03

Solucione los desafíos del lago de datos con un enfoque de DataOps

## 03

Los desafíos de usar lagos de datos empresariales

## 05

IBM Watson Knowledge Catalog

## 06

Una única fuente de información y un único punto de acceso

## 08

Cuatro beneficios de crear un lago de datos gestionado para la IA

## 09

Conclusión

# Conclusiones clave

- Pocas organizaciones ven el valor que esperan de los lagos de datos que han creado para almacenar y analizar sus datos a fin de obtener conclusiones confiables.
- DataOps soluciona los desafíos que enfrentan las organizaciones en cuanto a las deficiencias para acceder a los datos, prepararlos, integrarlos y ponerlos a disposición de los consumidores, mientras cumplen con las políticas corporativas y reglamentarias.
- Los desafíos comunes del lago de datos incluyen la dificultad y el costo de importar nuevas fuentes de datos al lago de datos, la incapacidad de integrar conjuntos de datos internos y externos, la falta de confianza sobre la gestión de datos, la falta de acceso a herramientas de preparación de datos de autoservicio, y la incapacidad de buscar y comprender los datos que están en el lago de datos.
- Una plataforma empresarial de gestión de datos con catalogación, calidad y detección de datos puede transformar un proyecto fallido de lago de datos en una verdadera fuente de valor comercial.
- [IBM Watson® Knowledge Catalog](#), basado en IBM Cloud Pak™ for Data, ofrece un catálogo de aprendizaje automático o machine learning (ML) para la detección, catalogación, calidad y gestión de datos. Ayuda a los usuarios de datos a descubrir, depurar, categorizar y compartir rápidamente activos de datos, conjuntos de datos y modelos analíticos.
- Si las organizaciones carecen de un conocimiento profundo de sus datos, es más difícil confiar y usar esta información con todas las formas de inteligencia artificial (IA), incluido el ML y el aprendizaje profundo.

## Solución de desafíos del lago de datos con un enfoque de DataOps

Hace diez años, el camino comenzó para encontrar un enfoque flexible y versátil para crear un almacén de datos central en el que pudiera residir toda la información empresarial. La solución era el lago de datos: un entorno de almacenamiento de información de uso general que almacenaría casi cualquier tipo de datos. También permitiría a los analistas de negocios y a los científicos de datos aplicar los motores y las herramientas de análisis más adecuadas a cada conjunto de datos, en su ubicación original.

En general, estos lagos de datos se crearon con Apache Hadoop y el sistema de archivos distribuidos de Hadoop (HDFS), combinados con motores tales como Apache Hive y Apache Spark. A medida que estos lagos de datos comenzaron a crecer, una serie de problemas se hizo evidente. Mientras la tecnología era físicamente capaz de escalar para capturar, almacenar y analizar amplias y variadas colecciones de datos estructurados y no estructurados, se prestó poca atención a los detalles prácticos sobre cómo insertar estas capacidades en los flujos de trabajo comerciales.

Para 2022, más del 80 % de los proyectos de lago de datos no aportarán valor, ya que buscar, inventariar y depurar datos demostrará ser el mayor inhibidor para el éxito del análisis y la ciencia de datos.<sup>1</sup> En consecuencia, preguntas tales como: “¿Qué datos debemos colocar en el lago de datos?”, “¿Quién los usará?”, “¿Cómo facilitamos la búsqueda?”, “¿De dónde provinieron estos datos?” y “¿Cómo evitamos que los datos se usen incorrectamente?”, con frecuencia quedaban sin respuesta. Estas limitaciones críticas para abordar los problemas con personas, procesos y tecnología efectivamente conllevaron implementaciones de lago de datos sin éxito.

Actualmente, muchas organizaciones han reconocido sus fallas, han cambiado los equipos de liderazgo para la implementación del lago de datos y están lanzando un segundo, tercer o incluso cuarto intento de implementar un lago de datos con éxito, esta vez liderando con operaciones de datos [DataOps](#).

En este informe técnico se hace una evaluación de los desafíos comunes que enfrentan los lagos de datos y brinda nuevos enfoques, como DataOps, que pueden ayudar a transformarlos de un pantano de datos a la atracción principal del canal de datos listos para usarse en los negocios de una organización.

---

DataOps es una práctica colaborativa de administración de datos enfocada en mejorar la comunicación, integración y automatización de los flujos de datos entre los administradores y los consumidores de datos en una organización.

---

### Presentación de DataOps

DataOps incorpora las mejores prácticas de DevOps, la administración de datos y gestión de datos a un marco común, con una manera colaborativa de desarrollar y mantener flujos de datos entre varias partes interesadas. DataOps se diseñó para solucionar los desafíos que enfrentan las organizaciones y que están relacionados con las deficiencias para acceder a los datos, prepararlos, integrarlos y ponerlos a disposición de los consumidores, mientras cumplen con las políticas

corporativas y regulatorias. Estas mejoras se pueden encontrar en una unidad comercial, un equipo analítico o, incluso, un proceso operativo.

Seguir esta metodología requiere atender problemas con personas, procesos y tecnología que se refieren a la diferencia entre las implementaciones de lago de datos con o sin éxito. Del lado de la tecnología, DataOps destaca la importancia de usar una plataforma completa y totalmente integrada para la ingesta, la integración, la gestión y el consumo de datos para crear un lago de datos gestionado. Las reglas de validación de calidad de datos deben ejecutarse automáticamente como parte del proceso de ingesta para sostener un canal de datos continuos en la empresa. El proceso de ingesta debe integrarse completamente con el catálogo de datos, el cual se convierte en el centro de su canal. Los consumidores de datos deben ser capaces de acceder a los niveles de calidad de información y a los resultados de perfiles de datos del catálogo de datos y confiar en que la organización trabaje con los mismos datos en contexto.

El aumento de los datos está superando la capacidad de las organizaciones de obtener valor de ello. Cuando se les preguntó a las organizaciones cuáles eran los desafíos más grandes de usar sistemas de conocimientos, respondieron lo siguiente: 1) el 40 % fusiona los procesos comerciales existentes con los datos fuente para analizarlos y 2) el 39 % obtiene, recolecta, administra y gestiona los datos a medida que aumentan.<sup>2</sup> Actualmente, no se trata solo de un caso de protección de las enormes inversiones de tiempo y recursos que ya se han hecho en las tecnologías de lago de datos: se trata del hecho de que no hay alternativa. Desde la implementación de IA o incluso para efectuar un análisis extenso, es fundamental tener una visión completa de la mayor cantidad de datos posible, lo cual significa que usted necesita una arquitectura que sea capaz de contener, analizar y gestionar todos esos datos en un solo lugar. En muchos casos, un lago de datos gestionado es la única opción razonable para cumplir con estos requisitos.

---

Las empresas actuales pueden (y deben) buscar la manera de extraer valor de su lago de datos al asegurarse de que respalde un canal de datos listos para usarse en los negocios para DataOps.

---

### Los desafíos de usar lagos de datos empresariales

#### Uso compartido de datos

Cuando un equipo dentro de una empresa adquiere o crea un nuevo conjunto de datos, es probable que tenga un fuerte sentido del valor de la información y de las susceptibilidades que la rodean. Si contiene información comercialmente confidencial, información de identificación personal (IIP) o datos del cliente, por ejemplo, el equipo sabrá cómo se debe usar dicha información y tomará las precauciones para asegurarse de que nadie la use incorrectamente.

También sabrá que, fuera de su equipo, otros posibles usuarios de los datos no podrán tener la misma comprensión del valor de los datos ni de los riesgos relacionados con su uso incorrecto. Estos riesgos harán, naturalmente, que el equipo sea muy precavido al compartir los datos o almacenarlos en ubicaciones que no estén bajo su control.

Esto es una mala noticia para los lagos de datos. Si la empresa ve el lago de datos simplemente como un basurero de datos descontrolado, será muy reacia a confiarle sus valiosos datos. En consecuencia, otras partes de la empresa no podrán beneficiarse de esos datos, y se desmoronará todo el concepto de usar el lago de datos como un depósito de autoservicio para compartir datos empresariales.

### Integración de datos

Incluso cuando un equipo acepta que sus datos se integren al lago de datos, puede ser un proceso complejo. El concepto original del lago de datos es importar datos en su formato en bruto, sin requerir los complejos procesos de extracción, transformación y carga (ETC) de un almacén de datos tradicional. Sin embargo, la realidad es que casi todas las fuentes de datos requieren cierto grado de preprocesamiento antes de que puedan ser útiles para cualquier tipo de análisis significativo.

En consecuencia, integrar una nueva fuente de información en un lago de datos con frecuencia puede demorar varios meses. Y, debido a que gran parte de estos datos anteriormente se han mantenido en pequeños silos operativos, podría haber docenas o incluso cientos de fuentes para integrar en total.

Esto significa que, en muchos casos, la información que necesitan los analistas de negocios o los científicos de datos aún no se ha agregado al lago de datos, y no se podría agregar durante meses o incluso años. Nuevamente, esto puede ser un obstáculo significativo para la adopción.

### Almacenamiento de datos

Mientras el costo de los recursos de almacenamiento y recursos informáticos han disminuido considerablemente en los últimos años, los conjuntos de Hadoop no son gratuitos. Almacenar cantidades enormes de datos en un lago de datos es mucho más rentable que hacerlo en un dispositivo de almacenamiento de datos de alto rendimiento, pero el costo aún puede ser significativo.

Además, a diferencia de los datos tradicionalmente almacenados en almacenes de datos, el índice de valor a volumen de los datos masivos mantenidos en un lago de datos es relativamente bajo. Es posible que necesite almacenar un pajar muy grande para ubicar un puñado de agujas de alto valor allí dentro.

Si no sabe qué conjuntos de datos realmente serán útiles y valiosos para sus científicos de datos, podría invertir sumas considerables en la integración y el almacenamiento de datos que están destinados a hundirse en el fondo de su lago de datos y nunca usarse.

### Búsqueda de datos

Suponiendo que haya identificado los conjuntos de datos más valiosos para almacenar, haya convencido a sus accionistas de que los compartan y haya podido integrarlos en su lago

## Los desafíos de usar lagos de datos empresariales

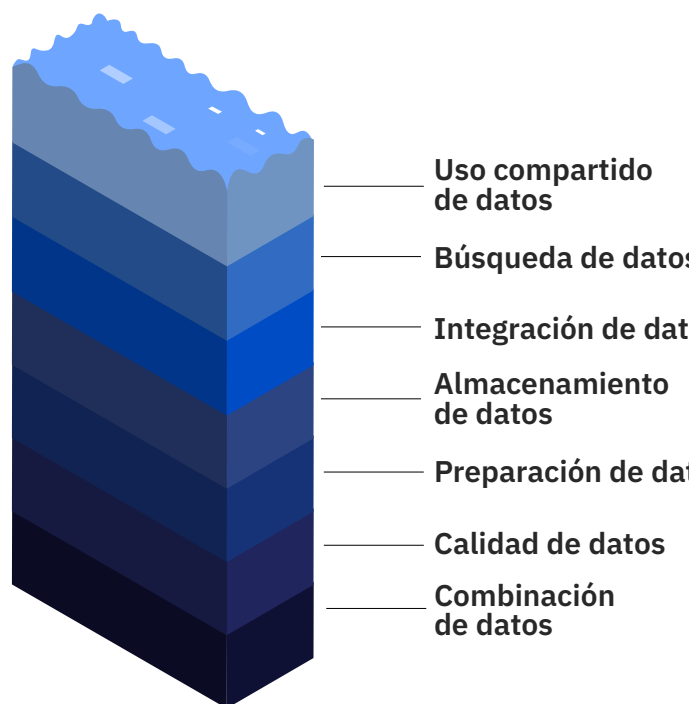


Figura 1. Las empresas que han adoptado tecnologías de lago de datos podrían tener uno o más de estos problemas comunes.

de datos, aún deberá posibilitar que los demás usuarios los encuentren, comprendan y usen correctamente. La calidad de los datos en el lago de datos también es otro desafío. Usted no está seguro de si los datos son de alta o baja calidad, pero se introducen en el lago.

Lamentablemente, en la mayoría de los lagos de datos esto no es fácil de lograr. Con frecuencia, los datos se almacenan sin contexto, lo que hace que sea difícil o imposible que un nuevo usuario los decodifique sin consultar al propietario original. La terminología suele ser tan específica del dominio que una métrica usada en un área de la empresa podría ser conocida con un nombre completamente diferente (o definido de manera sutilmente diferente) en otra. La posibilidad de confusión y malentendido puede ser tan grande que muchos conjuntos de datos son efectivamente inútiles, o incluso peligrosos, para un analista que no los conozca.

### Combinación de datos internos y externos

Por último, incluso el lago de datos más grande no debería contener todos los conjuntos de información posibles que los científicos de datos de una compañía deban usar. Por ejemplo, no tendría sentido importar una copia completa de Google Maps, Weather.com® o Bloomberg a su lago de datos solo porque uno de sus científicos de datos quiere realizar un análisis geoespacial, o integrar datos meteorológicos o precios de acciones a un algoritmo.

Debido a que su lago de datos no contendrá todos los datos que sus analistas de negocios necesitan para el análisis, deberán dedicar tiempo a buscarlos en varias aplicaciones. Debido a que es probable que una proporción muy grande de análisis útiles impliquen la combinación de conjuntos de datos

internos y externos, una vez más esto aumenta la barrera de ingreso y, desde la perspectiva del usuario, reduce el valor percibido del lago de datos.

### Preparación de datos

Hay muchos factores que hacen que la [preparación de datos](#) sea un desafío: desde saber dónde buscar los datos hasta darles formato. Preparar los datos para usar en el análisis es la tarea más ineficiente y demandante para los usuarios de datos. Los usuarios de datos dedican la mayoría del tiempo a buscar, limpiar y dar formato a la información en lugar de enfocarse en el análisis de datos, modelar y extraer conocimientos para lograr un impacto comercial.

La accesibilidad limitada a los conjuntos de datos gestionados también ha generado una dependencia excesiva de la TI durante la fase de preparación. Este acceso limitado señala la necesidad de mejorar las capacidades de autoservicio y habilidades de alfabetización de datos en la empresa para reducir este obstáculo.

### Calidad de los datos

Tirar los datos a un lago de datos puede convertirlos en inutilizables. Dado que no hay reglas de calidad o validación de datos que se apliquen a ellos antes de introducirlos en un lago de datos, este no proporciona información que se pueda usar o en la que se pueda confiar. Los datos de alta calidad son una característica esencial que determina la confiabilidad de los datos para tomar decisiones. Los datos son activos valiosos que se deben administrar a medida que se mueven por una organización. A medida que las fuentes de información se vuelven más numerosas y diversas, y las iniciativas de cumplimiento regulatorio están más enfocadas, es crítica la necesidad de integrar y acceder a la información desde estas fuentes dispares de forma coherente, confiable y reutilizable.

## Un enfoque integral para crear lagos de datos gestionados

La mayoría de los lagos de datos aprovechan Apache Hadoop y su amplio ecosistema de proyectos de código abierto para sus capas de almacenamiento de datos y motores de análisis. Como se esperaba, la comunidad de código abierto en torno a Hadoop ha reconocido los problemas que enfrentan las implementaciones actuales del lago de datos, y recientemente han surgido muchos proyectos con el objetivo de abordar diversos problemas de forma individual. Del mismo modo, hay varias herramientas patentadas en el mercado que pretenden solucionar los mismos problemas.

Puede ser tentador reparar los problemas de su lago de datos gradualmente, a medida que surjan. Si la cantidad de conjuntos de datos aumenta demasiado para poder manejarla, agregue una herramienta de catalogación. Si los usuarios se quejan de que no pueden encontrar los datos que necesitan, requieren un ajuste en la interfaz de usuario con una función de búsqueda. Y si los administradores de sus datos ya no pueden controlar de dónde vinieron sus datos o quiénes los usan, implemente herramientas de linaje de datos y un marco de gestión de datos.

Parece simple; sin embargo, en la práctica este enfoque gradual tiende a aparecer al costo de aumentar la complejidad y reducir la capacidad de mantenimiento enormemente, en especial a medida que aumenta la escala y el alcance del lago de datos. De

la misma manera que agregar nuevas fuentes de datos a un lago de datos aumenta la complejidad de sus requisitos ETL, agregar nuevas herramientas tiende a aumentar la complejidad de los requisitos no funcionales del lago de datos.

En lugar de tener una plataforma integrada y completa que pueda integrar datos, realizar operaciones de calidad sobre sus datos y catalogarlos para el uso eficaz de sus analistas de negocios, en general descubrirá que cada herramienta tiene sus propias formas de manejar las fallas y su propio enfoque para registrarse. En consecuencia, la localización de fallas y resolución de problemas puede ser muy demandante.

Otra limitación importante del enfoque gradual se hace evidente si adopta una perspectiva menos técnica y más conceptual de los problemas que normalmente experimentan los lagos de datos. La idea clave es que la escalabilidad, capacidad de búsqueda, integración, calidad de datos y gestión no sean problemas separados: están estrechamente interrelacionados. Solucionarlos requerirá un enfoque mucho más integral.

---

La escalabilidad, capacidad de búsqueda, integración, calidad de datos y gestión no son problemas separados: están estrechamente interrelacionados. Solucionarlos requerirá un enfoque integral sobre la administración de información.

---

## IBM Watson Knowledge Catalog Detección, catalogación y calidad de datos

El [IBM Watson Knowledge Catalog](#) impulsado por IBM Cloud Pak for Data ayuda a los usuarios de datos a descubrir, depurar, categorizar y compartir rápidamente activos de datos, conjuntos de datos, modelos analíticos y su relación con otras áreas de la organización. Ayuda a los equipos de gestión de datos a definir el glosario, las políticas y las reglas comerciales, y ofrece flujos de trabajo avanzados para gestión. El catálogo sirve como una única fuente de información para los ingenieros, administradores de datos, científicos de datos y analistas de negocios, a fin de que obtengan acceso de autoservicio a los datos en los que pueden confiar y usar con confianza.

Soluciones tales como IBM Watson Knowledge Catalog impulsado por IBM Cloud Pak for Data puede brindar todas las capacidades requeridas para solucionar los problemas principales de los lagos de datos actuales en una plataforma única y completa. El catálogo ayuda a abordar la causa raíz de estos problemas interrelacionados: la falla generalizada de los lagos de datos para ofrecer las herramientas eficaces para capturar, almacenar y administrar los metadatos y controlar el linaje de datos.

En muchos sentidos, el valor de un lago de datos depende de los metadatos que contiene, tanto como depende de los datos en sí. Sin los metadatos que expliquen de dónde proviene un conjunto de datos, quién lo creó, qué contiene, quién tiene permitido usarlo y cómo se usa, los datos en sí son prácticamente inútiles. Los usuarios no podrán encontrarlo e, incluso si lo hacen, no comprenderán lo que significa, ni lo usarán con confianza ni sabrán cómo pueden usarlo.

# Watson Knowledge Catalog

## Brindar datos confiables y significativos

### Organice sus datos



#### Conozca

Los datos deben estar completos, ser vigentes y se debe poder acceder a ellos en todas partes. Descubra, clasifique y comprenda todos los tipos de datos.

### Gestione sus datos



#### Confíe

Los datos deben ser seguros, estar limpios y ser fáciles de encontrar para incentivar el acceso de autoservicio confiable. Comprenda de dónde vinieron los datos y su calidad.

### Democratice sus datos



#### Consuma

Capacidad de impulsar la detección de autoservicio y automatizar la toma de decisiones para desarrollar el negocio. Ofrezca una visión de toda la información a aquellos que la necesitan y permítales acceder a ella.

Figura 2. IBM Watson Knowledge Catalog ofrece una amplia gama de capacidades para detección, catalogación, calidad y gestión de datos.

## Una única fuente de información y un único punto de acceso

IBM Watson Knowledge Catalog basado en IBM Cloud Pak for Data aborda estos problemas al hacer que los metadatos sean una prioridad clave. En el fondo es un poderoso motor de catalogación que indexa todos los conjuntos de datos y activos analíticos a los que tiene acceso su empresa, sin importar donde esté esta información, como en un lago de datos, almacén de datos o sistema transaccional, o incluso en un conjunto de hojas de cálculo. Independientemente de si están estructurados o no, o de si están almacenados en el sitio o alojados en la nube. Además, el catálogo también puede incluir conjuntos y fuentes de datos externos, como los servicios de datos patentados a los que su compañía esté suscrita, o API de datos abiertos.

Además de proporcionar una fuente única de información sobre todos sus conjuntos de datos, el catálogo de datos también ofrece un punto único de acceso. Las capacidades de búsqueda y sugerencia impulsadas por la IA ayudan a los analistas de negocios, científicos de datos, ingenieros de calidad de datos y equipos de gestión de datos a buscar activos más fácilmente, y presentar los metadatos disponibles para ayudar a los usuarios a entender lo que encontraron y evaluar si es útil para ellos.

Las capacidades integradas de preparación de datos de autoservicio aceleran el tiempo que lleva transformar los datos para un uso productivo en el análisis y las aplicaciones de IA. Los analistas de negocios y científicos de datos no deben perder tiempo en preparar y analizar los datos. La integración con una solución de preparación de datos a nivel empresarial, como [IBM® InfoSphere® Advanced Data Preparation](#), ayuda a garantizar que los conjuntos de datos gestionados que se crearon a través del catálogo aparezcan ante aquellos con el mayor contexto posible para impulsar los conocimientos y las acciones comerciales para usuarios comerciales. Esta integración promueve la colaboración en el canal de datos.

---

La escalabilidad, capacidad de búsqueda, integración, calidad de datos y gestión no son problemas separados: están estrechamente interrelacionados. Solucionarlos requerirá un enfoque integral sobre la administración de información.

---

El catálogo también ayuda a los administradores de datos en la oficina del director de datos (CDO) al etiquetar y clasificar los conjuntos de datos y rastrear automáticamente su linaje y uso, y al aprovechar el glosario comercial incorporado para estandarizar la terminología comercial entre los datos. En consecuencia, es más fácil para los administradores comprender qué contiene cada conjunto de datos, dónde se encuentra la información confidencial o la IIP, y a quiénes se les debe permitir acceder.



# Un catálogo único para varias fuentes de datos dentro y fuera de la organización

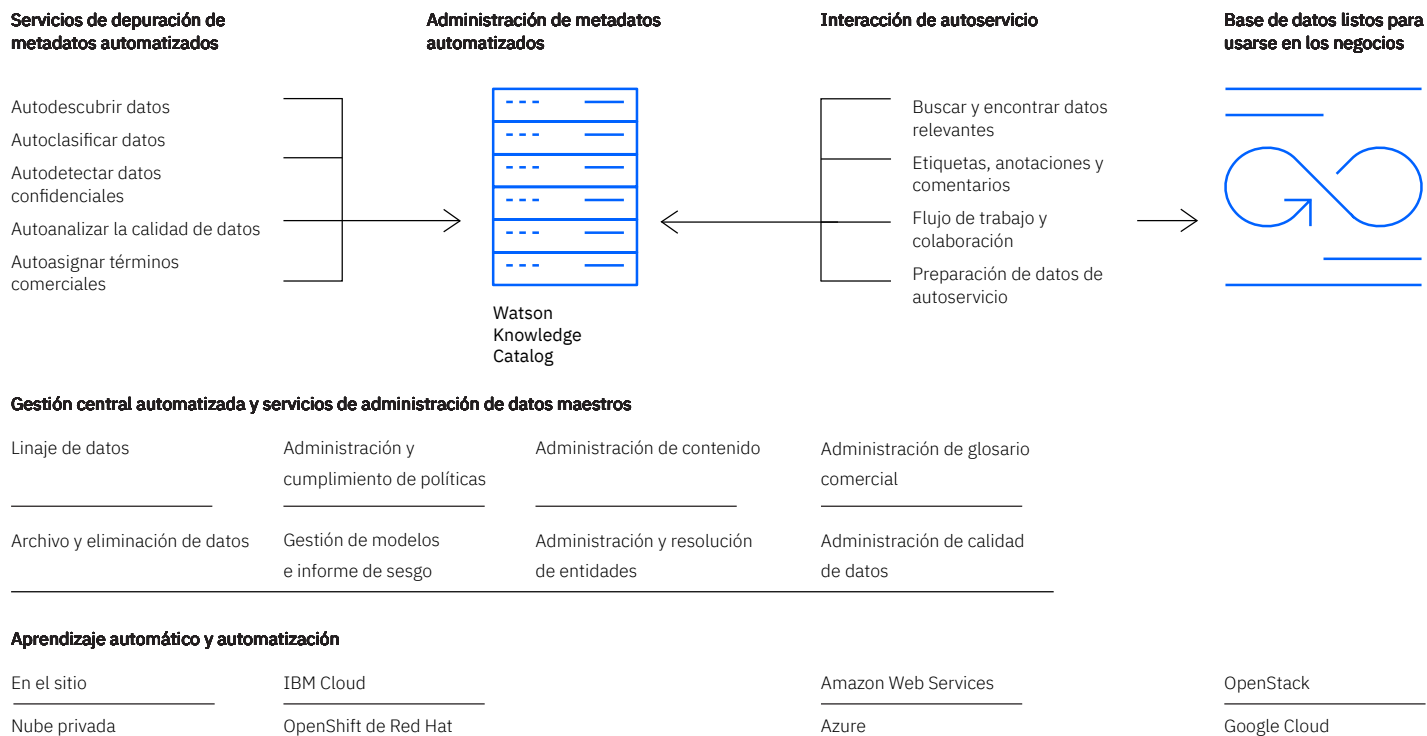


Figura 3. Con el índice de metadatos inteligente de IBM Watson Knowledge Catalog, los datos (estructurados y no estructurados) pueden residir en sistemas originales, pero los usuarios pueden descubrirlos rápidamente para un análisis más inteligente.

IBM Watson Knowledge Catalog hace que los metadatos sean una prioridad clave, al proporcionar una fuente única de información y un punto único de acceso a todos los conjuntos de datos a los que tiene acceso su empresa.

## Detección incorporada de datos inteligentes

Para mejorar más la capacidad de búsqueda, el catálogo permite a los usuarios etiquetar y comentar los conjuntos de datos y activos analíticos, lo que mejora los metadatos y agrega contexto adicional para ayudar a los compañeros a encontrar los que necesitan. La solución incluye algoritmos de detección incorporada de datos que utiliza ML para clasificar automáticamente los contenidos de cada conjunto de datos. Al identificar tipos de campo comunes, como nombres, direcciones, códigos postales y números de seguro social, la solución reduce la necesidad de que los autores anoten los datos manualmente. Integra automatización y ML para automatizar la depuración de datos y la administración de metadatos. Con las funciones de calidad de datos incorporadas, la solución permite perfiles de datos, calidad de datos y reglas de validación.

Las operaciones de datos automatizadas ofrecen un canal de datos depurados, con calidad y gestión de datos, también ayuda a garantizar que haya un flujo continuo de datos gestionados de alta calidad en el lago de datos.

De manera similar, agregar un modelo de metadatos inteligente de sus activos brinda una forma única de hacer cumplir automáticamente, como el Reglamento General de Protección de Datos (GDPR) y Ley de Privacidad del Consumidor de California (CCPA).

IBM Watson Knowledge Catalog impulsado por IBM Cloud Pak for Data ayuda a brindar datos listos para usarse en los negocios, confiables y de alta calidad, a básicamente a todos los usuarios de datos.

Todos los componentes de la solución han sido diseñados como microservicios, con un conjunto único de principios de diseño y un enfoque común a los requisitos no funcionales, como la escalabilidad, la gestión de errores, la seguridad y el registro.

IBM Watson Knowledge Catalog ofrece una plataforma de gestión empresarial de ML, de modo que está lista para la IA a escala.

En lugar de errores confusos y congestión de rendimiento que posiblemente deriven de un enfoque gradual de autogestión, IBM Watson Knowledge Catalog ofrece una plataforma de gestión empresarial de ML, de modo que está lista para la IA a escala.

IBM Watson Knowledge Catalog está disponible en tres variantes:

- Como una solución de software como servicio (SaaS) en IBM Cloud™
- En [IBM Cloud Pak for Data](#)
- Integrado a [IBM Watson Studio](#)

Las soluciones como IBM Watson Knowledge Catalog pueden desbloquear el valor que las iniciativas del lago de datos prometieron originalmente. Watson Knowledge Catalog con capacidades de catalogación y gestión inteligente ayuda a crear un lago de datos confiable y gestionado para la IA.

## Cuatro beneficios de crear un lago de datos gestionado para la IA

### 1. Cree confianza en los datos a través de la calidad y la gestión

- Las capacidades de calidad de datos lo ayudarán a mejorar la calidad de su información y a poner a disposición datos de alta calidad en su lago de datos.
- Las políticas de gestión se establecen y se hacen cumplir automáticamente, de modo que, cuando encuentre un conjunto de datos, sabrá si está autorizado a usarlo y cómo hacerlo.
- Puede depurar sus datos a medida que los usuarios agregan calificaciones, comentarios y otra información que ayudará a otras personas a determinar si un conjunto de datos será útil para ellas.

### 2. Autorice a sus usuarios de datos

- Sus equipos de línea de negocios (LDN) comparten sus datos de forma voluntaria porque confían en que serán gestionados y protegidos adecuadamente contra su uso incorrecto.
- Usted puede impulsar la colaboración y transformar los datos en activos empresariales confiables a través de las políticas de datos dinámicos y su cumplimiento.
- Sus datos se vuelven más localizables y reutilizables con el tiempo, a medida que los usuarios añaden etiquetas y metadatos relevantes para ayudar a otros a extraer valor de ellos.
- Una interfaz única le da acceso a cada conjunto de datos perteneciente a su organización, independientemente de la ubicación de almacenamiento.

### 3. Recupere su tiempo

- La detección automática de datos reduce el tiempo y el esfuerzo que debe dedicar a agregar metadatos para nuevos conjuntos de datos.
- La depuración automática de datos y la administración de metadatos reducen el tiempo que lleva descubrir metadatos y asignar términos, y también reducen el tiempo de creación de un glosario comercial.

- Con herramientas simples e intuitivas de preparación de datos de autoservicio, sus usuarios de información dedicarán menos tiempo a preparar datos y más tiempo a descubrir conocimientos.
- Usted libera a sus científicos de datos y a sus analistas de negocios para proporcionar mejores análisis en un espacio de tiempo más corto.
- La búsqueda inteligente basada en la IA lo ayuda a encontrar los datos que usted necesita en cuestión de segundos, en lugar de esperar semanas para que otro equipo los proporcione.

### 4. Administre los datos crecientes y el costo

- Usted puede optimizar los costos de almacenamiento al evitar el gasto de ingresar conjuntos de datos de bajo valor al lago de datos.
- También podrá ver todos los conjuntos de datos a los que su organización esté suscrita y reducir el riesgo de pagar más suscripciones de las que necesita.
- Usted puede priorizar la ingesta de nuevas fuentes de información en el lago de datos según la demanda de datos de los usuarios, lo que lo ayudará a integrar las fuentes más valiosas primero.

## Desbloquee el valor de sus datos

Si trabaja en la oficina del CDO, en el departamento de TI o como científico o analista de datos de LDN, usted y sus colegas comparten un objetivo en común. Si puede crear un lago de datos que realmente cumpla con lo prometido, usted no solo podría hacer sus propios trabajos de forma mucho más rápida y productiva. También podría cumplir una función clave al brindarle a su empresa una ventaja competitiva con la que pocas organizaciones actualmente puedan competir.

Si puede limpiar las aguas de su lago de datos mientras sus competidores aún se hunden en el pantano, abrirá las posibilidades con las que ellos solo pueden soñar. La ventaja genuina de ser pioneros espera a quienes son los primeros en desbloquear el valor de los datos previamente sin explotar.



# Conclusión

Conozca dónde residen sus datos, quién los usa y su valor para su empresa para el análisis.

Las iniciativas críticas para DataOps son los catálogos de datos, debido a que pueden ayudar a brindar administración de metadatos abiertos automatizados al integrar la gestión de datos, la calidad y la administración de políticas activas.

IBM Watson Knowledge Catalog con capacidades de catalogación y gestión inteligente ayuda a crear un lago de datos confiable y gestionado para la IA. El catálogo inserta la integración, calidad de datos y gestión en su entorno del lago de datos para brindar datos listos para usarse en los negocios para DataOps, y una fuente única de información.

## Más información

Para conocer más, visite:

[ibm.com/cloud/watson-knowledge-catalog](https://ibm.com/cloud/watson-knowledge-catalog)

© Copyright IBM Corporation 2019

IBM Corporation  
New Orchard Road  
Armonk, NY 10504

Producido en los Estados Unidos de América, en octubre de 2019 IBM, el logotipo de IBM, **ibm.com**, IBM Cloud, IBM Cloud Pak, IBM Watson y InfoSphere son marcas registradas de International Business Machines Corp., registradas en muchas jurisdicciones en todo el mundo.

Red Hat y OpenShift son marcas registradas de Red Hat, Inc. o sus sucursales en los Estados Unidos y otros países. Otros nombres de productos y servicios podrían ser marcas comerciales de IBM o de otras compañías. Hay una lista actualizada de las marcas comerciales de IBM disponible en la Web en "Copyright and trademark information" en [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Este documento está actualizado a la fecha inicial de su publicación y puede ser modificado por IBM en cualquier momento. No todas las ofertas están disponibles en todos los países donde opera IBM. La información de este documento se proporciona "tal cual" sin garantías de ningún tipo, ya sean expresas o implícitas, incluyendo cualquier garantía de comerciabilidad o de idoneidad para un propósito específico y cualquier garantía o condición de no violación. Los productos de IBM tienen garantía según los términos y condiciones de los acuerdos bajo los cuales se proporcionan. El cliente es responsable de garantizar el cumplimiento de las leyes y regulaciones correspondientes. IBM no proporciona asesoramiento legal ni declara o garantiza que sus servicios o productos asegurarán que el cliente cumpla con cualquier ley o regulación.

1. Augmented Data Catalogs: Now an Enterprise Must-Have for Data and Analytics Leaders, Gartner, sept. de 2019
2. The Forrester Wave: Machine Learning Data Catalogs, segundo trimestre de 2018

ASW12449-COES-03

