



White Paper

# Cree su ecosistema de datos listo para la IA

Crezca y prospere gracias a formatos de tabla abiertos y una gestión de datos inteligente

# Contenidos

<b>Resumen ejecutivo</b>	3
<b>Parte 1: Desbloqueo de datos preparados para la IA: El aumento de los formatos de tabla abiertos</b>	4
<b>Parte 2: Creación de un ecosistema de datos preparado para la IA</b>	7
<b>Parte 3: La estructura de la solución</b>	11
<b>Parte 4: Preparación para el futuro de los datos preparados para la IA con los OTF e Informatica IDMC</b>	14
<b>Recursos adicionales</b>	15

## Resumen ejecutivo

Los formatos de tabla abiertos (OTF, del inglés "Open Table Formats") como Apache Iceberg, Delta Lake y Apache Hudi se están convirtiendo rápidamente en la columna vertebral de los ecosistemas de datos empresariales modernos. Elevan el almacenamiento de objetos en el cloud sin procesar a capas de tabla inteligentes, versionadas e interoperables, lo que proporciona la base fundamental para las cargas de trabajo basadas en IA.

Sin embargo, a medida que las empresas escalan, los OTF por sí solos no pueden garantizar que los datos estén preparados para la IA. La preparación para la IA requiere una solución integral de gestión de datos para orquestar los datos y procesos de IA en diversas fuentes, formatos y motores informáticos, al tiempo que se mantiene la fiabilidad en entornos distribuidos.

**Intelligent Data Management Cloud (IDMC) de Informatica** ofrece esta base: combinar la compatibilidad nativa con los OTF con la automatización basada en CLAIRE, la inteligencia de metadatos y la compatibilidad de ecosistemas en entornos híbridos, multicloud y Edge.

Mientras se encuentra en la cúspide escalando los pilotos de IA a iniciativas estratégicas de nivel empresarial, este white paper describe cómo los OTF e IDMC pueden ofrecer juntos una estrategia preparada para el futuro para que pueda aprovechar los datos preparados para la IA a escala.

### IDMC Foundation



Compatibilidad nativa con los OTF



Inteligencia de metadatos



Automatización impulsada  
por CLAIRE



Compatibilidad de los  
ecosistemas

## Parte 1

### Desbloqueo de datos preparados para la IA: El aumento de los formatos de tabla abiertos

Las arquitecturas de datos empresariales siempre han evolucionado en respuesta a las demandas empresariales, las nuevas tecnologías y los casos de uso emergentes.

Cada ola de innovación amplió el alcance de lo que las organizaciones podían preguntar y responder con datos, y cómo podían hacerlo de la manera más fácil. Desde los data marts para el análisis departamental, hasta los data warehouses empresariales para la creación de informes consolidados y los lakes basados en Hadoop para el IoT y los datos de pulsaciones, la arquitectura de datos ha seguido avanzando para avanzar en la eficiencia y el rendimiento. (Figura 1)

**Las nuevas fronteras habilitadas para el cloud:** data lakes escalables para la experimentación con inteligencia artificial y aprendizaje automático (IA/ML), almacenes en la nube para business intelligence (BI) en tiempo real y vistas integrales de los clientes, y modelos de lakehouse que combinan lo mejor de ambos mundos.

Hoy en día, con el auge de la transmisión en tiempo real y las bases de datos vectoriales que impulsan la IA generativa, las arquitecturas de datos se están rediseñando una vez más. Con el crecimiento exponencial de las cargas de trabajo de IA en toda la empresa, la ineficiente escalabilidad, la falta de flexibilidad y la fragmentación continua de los datos están creando obstáculos para la preparación para la IA.

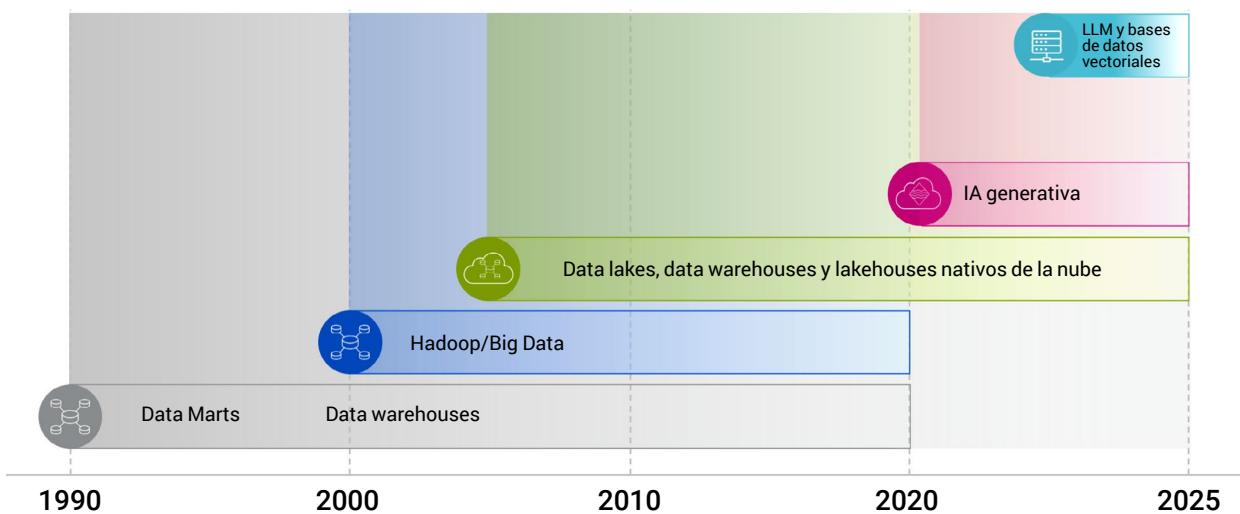


Figura 1: La evolución del almacenamiento de datos para casos de uso de análisis

Cree su ecosistema de datos listo para la IA

A través de estos cambios, un tema emergente ha sido la separación de los recursos informáticos y el almacenamiento. Esta evolución arquitectónica ha desbloqueado modelos de gestión de datos más flexibles y rentables al liberar a las empresas de pilas rígidas y monolíticas. (Figura 2)

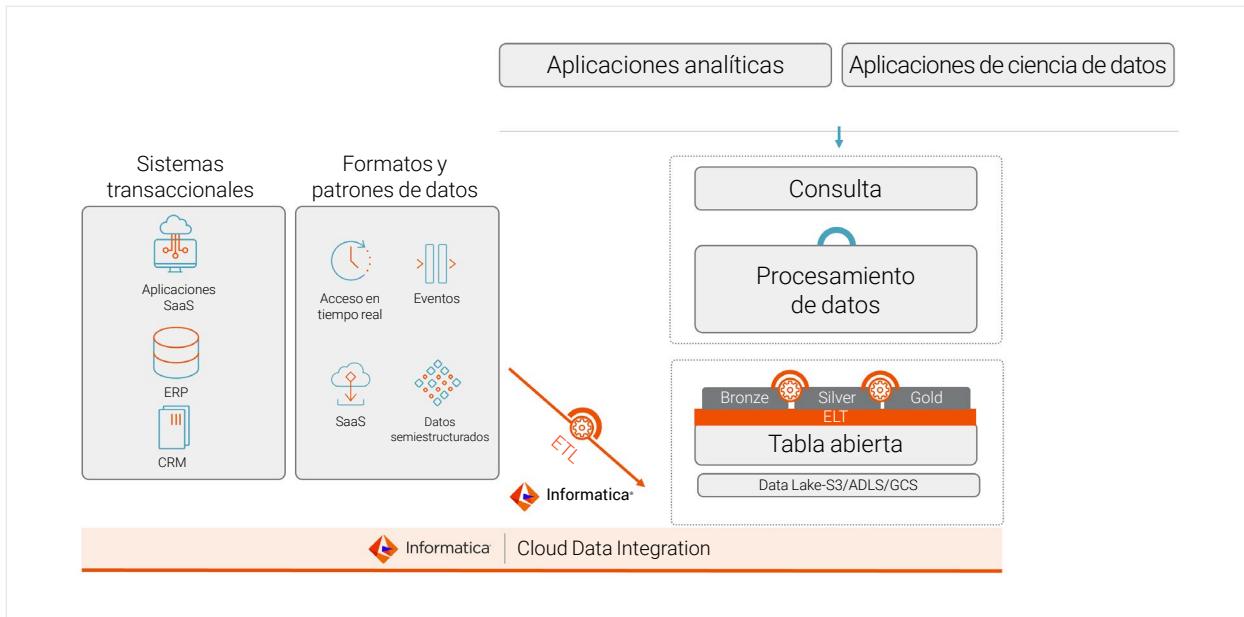


Figura 2: El desacoplamiento del almacenamiento y los recursos informáticos en la era de los OTF

Este es el motivo por el que los formatos de tablas abierta (OTF) como Apache Iceberg, Delta Lake y Apache Hudi han ganado relevancia y tracción. Los OTF han revolucionado el almacenamiento de datos transformando los almacenes de objetos del cloud en capas de tablas inteligentes, que permiten consultas y versionadas que pueden admitir tanto cargas de trabajo de análisis avanzado como de IA/ML a la velocidad que las empresas necesitan.

En el ecosistema de datos empresarial más amplio, los OTF demuestran ser fundamentales para las arquitecturas modernas como data lakehouse, data mesh y data fabric. La capa de almacenamiento unificado, estandarizada e interoperable habilitada por los OTF permite a los motores informáticos trabajar de manera fluida y eficiente en entornos diversos y distribuidos, haciendo que las arquitecturas de datos estén preparadas para la IA desde el diseño.

Como se muestra en esta arquitectura de referencia del lakehouse (Figura 3), los OTF se sitúan en el centro de la pila de la plataforma de datos, y tienden un puente entre el almacenamiento en la nube sin procesar con almacenes en la nube, catálogos y análisis avanzados.

Al aportar la agilidad, la interoperabilidad y el rendimiento necesarios para los complejos entornos multimotor de hoy en día, los OTF se han convertido en el tejido conectivo de la pila de datos e IA moderna.

Cree su ecosistema de datos listo para la IA

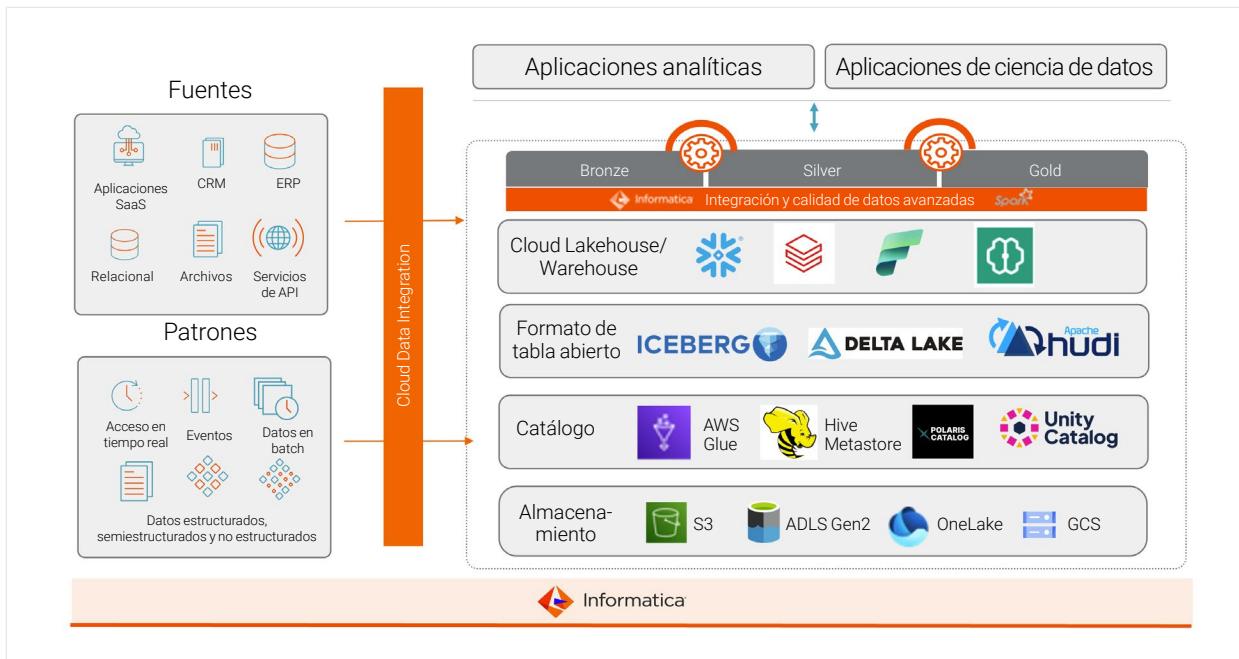


Figura 3: Los OTF son el tejido conectivo de su arquitectura de datos moderna

## Gestión del punto de inflexión de la IA

Hoy en día, usted se encuentra en un punto de inflexión, buscando la ruta óptima para escalar más allá de los pilotos de IA hacia los sistemas operativos basados en IA. Al igual que muchos otros profesionales de datos que están pensando en utilizar los OTF, es posible que le surjan preguntas clave a medida que realiza la transición.

Por ejemplo, ¿cómo puede garantizar la interoperabilidad de los datos en diversos motores? ¿Cómo evitar que la dependencia de un proveedor y ofrecer datos fiables y preparados para la IA lo suficientemente rápido como para mantenerse al día en cuanto a la innovación en análisis y casos de uso de la IA?

Informatica ha ido constantemente un paso por delante en cuanto a habilitar este recorrido, admitir los OTF, optimizar el almacenamiento y garantizar que las empresas puedan escalar con confianza hacia los casos de uso basados en la IA.

En todos los sectores y verticales, los equipos de datos empresariales pueden cumplir mejor con la promesa de los OTF con soluciones de gestión de datos inteligentes que automatizan procesos fiables y preparados para la IA, reducen la complejidad de la integración y garantizan arquitecturas híbridas preparadas para el futuro.

Este práctico white paper explora cómo puede aprovechar los formatos de tabla abiertos que estén preparados para ecosistemas de datos preparados para la IA y cómo Intelligent Data Management Cloud (IDMC) de Informatica le permite hacerlo con confianza y facilidad.

## Parte 2

### Creación de un ecosistema de datos preparado para la IA

#### ¿Por qué los formatos de tabla abiertos necesitan la base de gestión de datos adecuada para optimizar la entrega de datos preparados para la IA?

Mientras que los OTF resuelven el problema del almacenamiento y recuperación estandarizados, solo son una pieza del rompecabezas. Optimizar el almacenamiento es vital para garantizar que los ecosistemas de datos continúen ofreciendo un alto rendimiento en entornos cada vez más diversos e híbridos, que abarcan el cloud, el entorno local, los modelos Edge y los modelos público/privado.

En la era de la IA, el almacenamiento en sí está evolucionando rápidamente para gestionar la explosión de fuentes, tipos de datos, formatos, volúmenes y escala de datos empresariales. Los OTF están liderando la carga, pero sin una base de datos unificada, su capacidad para proporcionar datos preparados para la IA sigue siendo limitada.

Los datos preparados para la IA no solo se caracterizan por la velocidad, sino por un acceso rápido a datos fiables y utilizables. Desde una perspectiva de almacenamiento y arquitectura, hay una necesidad de agilidad, rendimiento, auditabilidad, linaje e integración, lo que garantiza que los datos no solo estén disponibles, sino que sean fiables y procesables para la IA.

Los OTF como Apache Iceberg, Delta Lake y Hudi proporcionan la capa de almacenamiento inteligente en arquitecturas de lakehouse; sin embargo, no gestionan cómo los datos entran en el lakehouse ni garantizan su calidad.

#### Para ofrecer correctamente datos preparados para la IA para los OTF, los lakehouses multicloud necesitan:

1. Movimiento de datos flexible a escala en dominios y el cloud.
2. Sincronización operacional en tiempo real (**ETL/ETL inverso**) para impulsar acciones empresariales oportunas.
3. **Catalogación de metadatos** y gobierno centralizados para garantizar el cumplimiento, la coherencia y la confianza en todos los procesos de datos e IA.

Aquí es donde su elección de una solución de integración de datos se vuelve crucial. Las herramientas de integración proporcionan el motor de incorporación y transformación, y extraen los datos de diversas fuentes (CRM, ERP, API, IoT, etc.), los transforman y los limpian para obtener coherencia y calidad, y los cargan en tablas respaldadas por los OTF en el lakehouse.

Sin una sólida **integración de datos**, los OTF corren el riesgo de convertirse en islas fragmentadas, lo que mina la preparación para la IA mediante procesos rotos, conjuntos de datos incoherentes e información que se proporciona demasiado tarde. Algunos ejemplos de esto en la práctica podrían ser los siguientes:

- Tablas que pueden contener datos incompletos, incoherentes o de baja calidad.
- Los desajustes de los esquemas y los errores de incorporación pueden interrumpir los análisis posteriores.
- Los modelos de IA entrenados con datos poco fiables o sesgados pueden dar lugar a resultados defectuosos.

El potencial combinado de la integración de datos y los OTF puede abordar estos desafíos. De hecho, las preocupaciones planteadas por los profesionales destacan por qué la gestión inteligente de datos es fundamental para el rendimiento de los OTF:

- **Preparación empresarial:** Si bien los proyectos desde cero tienden a adoptar los OTF de manera más fácil, los entornos legacy complejos requieren herramientas inteligentes para preparar y alinear los datos para una adopción exitosa de los OTF.
- **Formatos adaptados específicamente:** Cada OTF tiene puntos fuertes: Hudi para los procesos de alto nivel de incorporación, Iceberg para una amplia interoperabilidad, Delta Lake para los flujos de trabajo nativos de Databricks. La integración inteligente garantiza que las empresas puedan seleccionar y organizar el formato adecuado para la carga de trabajo correcta.
- **Gestión y mantenimiento continuos a escala:** La implementación de los OTF en sí no es suficiente: la orquestación inteligente y el gobierno, impulsados por soluciones como IDMC, son necesarios para mantener los procesos de datos escalables, fiables y preparados para la IA.
- **Integración basada en el contexto:** Las altas tasas de mutación o los archivos pequeños no gestionados pueden degradar el rendimiento de los OTF a menos que lo respalden la integración y optimización de datos basadas en el contexto.

Mientras que los OTF optimizan la forma en que se almacenan y se consultan los datos, su potencial completo solo se materializa cuando va aparejado de una capa de integración y gestión de datos completa. Juntas, garantizan que los datos que fluyen hacia el lakehouse sean coherentes, de alta calidad y preparados para la IA: arquitecturas modernas y preparadas para el futuro para las demandas de la era de la IA.

La siguiente tabla de comparación ilustra por qué los OTF y la **integración de datos** son más fuertes juntos, y por qué las empresas deberían verlos como componentes complementarios de una base de datos unificada preparada para la IA. (Tabla 1)

<b>Aspecto</b>	<b>Integración de datos</b> (Por ejemplo, Informatica IDMC)	<b>Formatos de tabla abiertos (OTF)</b> (Por ejemplo, Apache Iceberg, Delta Lake, Hudi)
<b>Finalidad</b>	<ul style="list-style-type: none"> <li>Gestiona la forma en que se incorporan, transforman y entregan los datos</li> <li>Responsable de obtener y transformar los datos que fluyen hacia el lakehouse</li> </ul>	<ul style="list-style-type: none"> <li>Gestiona la forma en que se almacenan los datos y se accede a ellos</li> <li>Permite tablas que permiten consultas, eficientes y versionadas en el lakehouse</li> </ul>
Informatica cuenta con el conjunto más amplio de conectores y automatización inteligente para crear y optimizar procesos de datos, lo que garantiza un flujo de datos fluido a escala en todo el ecosistema de datos. Entre las funcionalidades se incluyen las siguientes:		<ul style="list-style-type: none"> <li>Solución completa de incorporación y replicación para todos los patrones de incorporación (ETL/ELT/ETL inverso)</li> <li>Integración de datos multicloud</li> <li>Integración de datos en tiempo real y en batch</li> <li>SQL ELT compatible con ecosistemas multicloud</li> </ul>
<b>Objetivo</b>	Garantiza que los datos estén limpios, coherentes, conectados y preparados para su uso	Optimizado para operaciones de análisis, consultas y almacenamiento de datos
Informatica garantiza que los datos sean correctos, completos y coherentes, así como conectados, contextuales y continuos para todos los casos de uso de IA y análisis actuales y futuros. Entre las funcionalidades se incluyen las siguientes:		<ul style="list-style-type: none"> <li>Más de 50 000 conexiones basadas en metadatos</li> <li><b>Inteligencia de metadatos basada en CLAIRE</b></li> <li>Validación de datos integrada basada en asistente</li> <li>Orquestación automatizada de procesos de datos</li> </ul>
<b>Foco</b>	Lógica del proceso, reglas de negocio, calidad de datos.	Rendimiento en el nivel de tabla, evolución del esquema, time travel
La base de metadatos de Informatica garantiza la calidad y el gobierno de los datos para admitir todos los catálogos, arquitecturas y patrones. Entre las funcionalidades se incluyen las siguientes:		<ul style="list-style-type: none"> <li>Inteligencia de metadatos basada en CLAIRE</li> <li><b>Calidad y observabilidad de los datos</b> para la supervisión de los procesos</li> <li><b>Datos basados en IA y gobierno de IA</b></li> <li>Controles de acceso a datos detallados</li> </ul>
<b>Interdependencia</b>	Envía datos procesados a los destinos OTF	Depende de los procesos de integración para llenar las tablas con datos de alta calidad
	<p>IDMC proporciona datos limpios, particionados y alineados con esquemas para ayudar a los OTF a rendir de manera óptima, evitando al mismo tiempo errores comunes como desajustes de esquemas, consultas lentas o duplicación de datos que ralentizan los OTF.</p>	<p><b>Entre las funcionalidades se incluyen las siguientes:</b></p> <ul style="list-style-type: none"> <li>Conectores nativos para Iceberg, Delta, Hudi</li> <li>Acceso a OTF en Snowflake, Databricks y Microsoft Fabric</li> <li>Gestión de esquemas y metadatos</li> <li>Flexibilidad en la elección de almacenamiento y catálogo</li> </ul>

Tabla 1: La conexión estratégica entre la integración de datos y los OTF: ambos trabajan en conjunto para garantizar que los flujos de datos preparados para la IA lleguen a los usuarios de negocio.

### Resumen: Conclusiones ejecutivas

- Los OTF estandarizan el almacenamiento y la recuperación, pero por sí solos no pueden garantizar los datos preparados para la IA.
- Desde el punto de vista del almacenamiento y la arquitectura, la preparación para la IA significa mucho más que la rapidez del acceso: requiere procesos de datos fiables, gobernados y de alta calidad.
- La integración de datos es el eslabón que falta para garantizar que los datos limpios, coherentes y basados en el contexto fluyan hacia los lakehouses respaldados por los OTF.
- La combinación correcta de los OTF + la integración reduce el riesgo de rendimiento deficiente, desajustes de esquemas y modelos de IA poco fiables.
- Informatica permite esta sinergia, y ayuda a las empresas a escalar la adopción de los OTF con automatización inteligente, integración y gobierno.

## Parte 3

### La estructura de la solución

La solución de gestión de datos adecuada puede complementar y mejorar los resultados de sus implementaciones de los OTF, y garantizar así datos de entrada fiables y de alta calidad, y optimizar el rendimiento del almacenamiento de datos.

Tras haber establecido la importancia estratégica de los OTF y su conexión con la integración de datos, su siguiente paso es definir cómo su inversión en los OTF puede ofrecer el máximo valor con datos preparados para la IA.

Desde el punto de vista del almacenamiento, sabemos que el concepto "datos preparados para la IA" significa tener acceso a datos fiables y utilizables. Desde la perspectiva de la integración de datos, el concepto de los datos "preparados para la IA" implica más que las nociones tradicionales de "alta calidad". Deben estar más que limpios, ser más que correctos y estar más que completos. También deben ser coherentes, continuos y contextuales.

Lograr esto en arquitecturas híbridas y multicloud está más allá de la capacidad de las soluciones de uso puntual, las herramientas de integración de datos independientes, los procesos codificados a mano ad hoc y las soluciones alternativas, que carecen de la profundidad y la amplitud necesarias para mantenerse a la altura de la velocidad que caracteriza a las iniciativas modernas de IA.

Solo una plataforma de gestión de datos completa, escalable e inteligente puede garantizar que los datos de confianza preparados para la IA fluyan a escala hacia arquitecturas respaldadas por los OTF. Combina la compatibilidad nativa con los OTF con funcionalidades integrales de integración, gobierno y automatización a nivel y escala empresarial, y le permite:

- **Superar los silos de datos** mediante la integración de diversos sistemas, formatos y fuentes de forma eficiente.
- **Unificar el acceso a los datos y la disponibilidad de los datos** para acelerar las consultas, admitir casos de uso en tiempo real/transmisión y habilitar varios motores de procesamiento.
- **Gestionar las transformaciones** para que los datos más actualizados y preparados para el negocio estén siempre disponibles.
- **Eliminar los cuellos de botella de rendimiento** con un control detallado de las variables y las cargas de trabajo de la sesión.
- **Respaldar cada caso de uso moderno** con funcionalidades modulares, flexibles y listas para los entornos multicloud.

### Scorecard de ajuste estratégico (para la solución que se está evaluando)

Dimensión	S/N
<b>Alineación empresarial</b> ¿Está alineado para proporcionar datos preparados para la IA para todos los casos de análisis de negocios y de uso de IA?	
<b>Madurez técnica</b> ¿Nos permite escalar de forma eficiente el almacenamiento y los recursos informáticos para satisfacer las necesidades empresariales en constante evolución?	
<b>Preparación para el cumplimiento</b> ¿Respalda el cumplimiento y el riesgo críticos del negocio?	
<b>Interoperabilidad del ecosistema</b> ¿Ofrece interoperabilidad entre diversas herramientas de análisis, IA y BI?	
<b>Plazo de amortización</b> ¿Tiene funcionalidades impulsadas por IA que aceleran el plazo de amortización y optimizan los costes de funcionamiento sobre la marcha?	
<b>Notas:</b>	

Aquí es donde Informatica IDMC marca la diferencia. IDMC proporciona compatibilidad nativa con los OTF líderes, al tiempo que añade la automatización inteligente, la calidad de los datos y el gobierno que las empresas necesitan. Garantiza que sus lakehouses respaldados por los OTF ofrezcan datos reales preparados para la IA: fiables, de alta calidad y optimizados para los análisis modernos y los casos de uso basados en IA. (Tabla 2)

Función o consideración de los OTF	Por qué es importante para el negocio	Cómo Informatica admite los OTF para entregar datos preparados para la IA
<b>Evolución del esquema</b>	Modifica los esquemas de tablas (añadir/soltar/cambiar el nombre de las columnas) sin volver a escribir los datos históricos. Permite el desarrollo y la experimentación ágiles sin interrumpir los procesos posteriores.	Informatica es compatible con la gestión de la desviación del esquema y el mapping dinámico de esquemas durante la incorporación. Admite CDC para actualizaciones en tiempo real de las tablas de OTF sin recargas completas.
<b>Transacciones de ACID</b>	Evita el daño de los datos durante operaciones simultáneas. Mantiene la coherencia, la confianza y la fiabilidad en entornos multiusuario.	Mientras que ACID es nativo de los OTF, Informatica garantiza la integridad transaccional durante las escrituras a través de los conectores.
<b>Time travel/ Creación de versiones de datos</b>	Admite datos dinámicos y versionados, así como reproducibilidad en flujos de trabajo de análisis y aprendizaje automático mediante la consulta de instantáneas históricas. Permite acceder a estados anteriores de una tabla para auditoría, depuración o reversión.	Totalmente compatible con el conector nativo de la tabla abierta: los usuarios pueden consultar instantáneas históricas o marcas de tiempo.
<b>Evolución de la partición y aislamiento de instantáneas</b>	Permite cambios dinámicos en las estrategias de partición a lo largo del tiempo. Garantiza lecturas coherentes durante escrituras simultáneas.	Los mappings de Informatica pueden gestionar de forma dinámica datos particionados y desarrollar estrategias de partición.
<b>Compatibilidad multicatálogo</b>	Permite la interoperabilidad, evita la dependencia de un proveedor y optimiza los costes en función de las cargas de trabajo.	Informatica permite a los usuarios seleccionar el catálogo de su elección (AWS Glue/Hive Metastore/ catálogos basados en REST). Esta flexibilidad significa que Informatica puede: <ul style="list-style-type: none"> <li>Incorporar, transformar y gobernar los datos almacenados en los OTF independientemente del catálogo subyacente</li> <li>Preservar y enriquecer los metadatos en todas las plataformas</li> <li>Habilitar un gobierno coherente y un seguimiento del linaje, incluso en entornos híbridos o federados</li> </ul>
<b>Interoperabilidad, compatibilidad nativa para múltiples OTF</b>	Garantiza la compatibilidad y el rendimiento.	IDMC se integra con todas las principales plataformas en el cloud, OTF y almacenes de datos, lo que garantiza la compatibilidad entre plataformas.

<b>Gestión y enriquecimiento de metadatos</b>	<p>Detección de datos más rápida, mejor gobierno y cumplimiento.</p> <p>Permite la reducción de particiones, la omisión de datos y la aceleración de consultas, lo que reduce los costes informáticos.</p>	<p>IDMC actúa como un sistema de registro de metadatos que integra metadatos técnicos, empresariales y operacionales.</p> <p>Esto garantiza que los metadatos sean coherentes en las tablas de los OTF y los motores de consulta, lo que permite que las funciones de evolución del esquema y time travel funcionen sin problemas.</p> <p>Informatica actúa como un puente para los metadatos entre:</p> <ul style="list-style-type: none"> <li>• Sistemas legacy y lakehouses modernos</li> <li>• Entornos locales y nativos del cloud</li> <li>• Diversas implementaciones de los OTF</li> </ul> <p>Esto es especialmente poderoso para las empresas que adoptan arquitecturas de datos híbridas, donde la fragmentación del catálogo (metadatos como definiciones de tablas, versiones de esquemas, linaje y políticas de acceso se encuentran dispersos en diversos catálogos que no se comunican entre sí) es un verdadero desafío.</p>
<b>Compatibilidad con la incorporación en batch y por transmisión</b>	<p>La latencia y la actualización de los datos son clave para potenciar el análisis en tiempo real y la IA.</p>	<p>IDMC permite la incorporación de datos de alto rendimiento desde diversas fuentes y admite procesos de incorporación de datos tanto en batch como en tiempo real, incluidos Kafka y flujos nativos del cloud.</p> <p>La integración de datos de alto rendimiento en flujos paralelos permite realizar consultas y actualizaciones más rápidas, y garantiza que solo los datos limpios y relevantes lleguen a la capa de los OTF.</p> <p>La optimización y orquestación inteligente de fuentes y destinos reduce la latencia y mejora el rendimiento.</p>
<b>Compactación y optimización de los datos</b>	<p>La mejora del rendimiento de las consultas y la reducción de los costes de almacenamiento aceleran la toma de decisiones y permiten la escalabilidad.</p>	<p>Mientras que la compactación se gestiona mediante el motor de los OTF, Informatica puede activar u orquestar flujos de trabajo de compactación.</p>
<b>Gobierno y seguridad</b>	<p>Reduce el riesgo y la vulnerabilidad, crea una mayor confianza en los datos para una toma de decisiones fiable.</p>	<p>Un sólido gobierno a través de la catalogación de datos, el gobierno de datos e IA, la calidad y observabilidad de los datos, el mercado de datos y la gestión del acceso a los datos garantizan la transferencia segura de los datos hacia los OTF, alineándose con los modelos de gobierno del lakehouse.</p> <p>El seguimiento del linaje de datos ayuda a rastrear cómo fluyen los datos a las tablas de los OTF, lo que es fundamental para el cumplimiento y la depuración.</p>

Tabla 2: Cómo Informatica permite un rendimiento óptimo de los OTF

### Resumen: Conclusiones ejecutivas

- Los datos preparados para la IA deben ser limpios, correctos, completos, coherentes, continuos y contextuales.
- Las soluciones puntuales y los procesos codificados a mano no tienen la profundidad ni la amplitud necesarias para proporcionar datos preparados para la IA en las demandas de IA híbrida y multicloud.
- Solo los sistemas completos e inteligentes complementan los OTF de forma eficaz. La plataforma correcta integra, transforma, gobierna y optimiza los flujos de datos de principio a fin.
- Informatica IDMC mejora de forma única los resultados de los OTF, lo que garantiza que las empresas obtengan datos fiables y preparados para la IA a escala.

## Parte 4

### Preparación para el futuro de los datos preparados para la IA con los OTF e Informatica IDMC

A medida que acelera sus estrategias de IA, las demandas de sus ecosistemas de datos solo se intensifican. Los OTF proporcionan la base para la interoperabilidad, la flexibilidad y el rendimiento, pero su valor completo solo emerge cuando se combina con una gestión de datos completa e inteligente.

Informatica IDMC ofrece precisamente eso: una plataforma modular de principio a fin que garantiza que los datos preparados para la IA fluyan a través de las arquitecturas híbridas, multicloud y Edge más complejas.

Informatica IDMC no solo es compatible con los OTF, sino que los mejora, y garantiza que proporcionan datos fiables, de alta calidad y preparados para la IA para cada caso de uso de análisis e IA moderno.

El aprovechamiento de sus OTF con Informatica IDMC significa que puede comenzar a entregar datos a escala empresarial y preparados para la IA en cuestión de días. Conéctese con los expertos en soluciones de Informatica para realizar una evaluación personalizada de su arquitectura de datos moderna hoy mismo.

### Funcionalidades que hacen de Informatica IDMC la base ideal para sus OTF

1. Conectividad completa: IDMC garantiza la **compatibilidad del ecosistema**, ya sea en el cloud, en el entorno híbrido o el Edge, sin dejar de ser modular, independiente de los proveedores y rentable. Esto garantiza que su arquitectura esté preparada para el futuro para que pueda proteger las inversiones de hoy y permitir las innovaciones del mañana.
2. Automatización basada en IA: Con la **automatización e inteligencia impulsadas por CLAIRE**, IDMC acelera la adopción, reduce la fricción y fomenta una cultura de datos en toda la empresa. Su **base de metadatos** va más allá de la integración para cubrir todo el ciclo de vida de los datos: incorporación, transformación, gobierno, calidad y entrega, gestionando las cargas de trabajo empresariales de manera eficiente.
3. **Plataforma unificada para la integración, la calidad y el gobierno:** A diferencia de las soluciones de uso puntual, IDMC es una plataforma de gestión de datos completa con integración, calidad de datos, gobierno, catalogación y funcionalidades de gestión de datos maestros incorporados. Este enfoque de principio a fin simplifica la gestión de los OTF en flujos de trabajo de IA y análisis complejos.
4. **Flexibilidad para admitir arquitecturas de datos modernas:** IDMC está diseñado para admitir arquitecturas emergentes como data lakehouses, data mesh y data fabric aprovechando los OTF como tecnología fundamental, mientras que otros proveedores a menudo se centran en arquitecturas legacy o aisladas.
5. **Gobierno y calidad de datos sólidos:** La estrecha integración de la calidad de los datos, el perfilado de datos y el linaje de principio a fin con los conjuntos de datos de los OTF garantiza datos fiables preparados para la IA, una funcionalidad que algunos competidores ofrecen solo como módulos independientes o complementarios.
6. **Escalabilidad y rendimiento:** Los motores de ejecución elásticos y contenerizados de Informatica y el ajuste del proceso basado en IA ayudan a optimizar el rendimiento y la rentabilidad al procesar grandes conjuntos de datos de los OTF, lo que le da una ventaja para la preparación para la IA a escala empresarial.

Cree su ecosistema de datos listo para la IA

## Recursos adicionales

### Solución de gestión de datos compatible con los OTF: Lista de comprobación de evaluación

¿Su sistema de gestión de datos complementa su moderna arquitectura de data lakehouse con tecnología de OTF para entregar datos preparados para la IA para casos de uso actuales y emergentes?

Utilice esta lista de comprobación para averiguarlo.

Módulo 1: Compatibilidad básica y adecuación arquitectónica	S/N
Conectores nativos para el OTF elegido (Iceberg, Delta Lake, Hudi)	
Se integra con el almacenamiento de objetos en el cloud (por ejemplo, S3, ADLS, GCS)	
Compatible con el motor informático	
Permite desacoplar flujos de trabajo de los recursos informáticos y el almacenamiento	
Facilidad de uso: poco código o ninguno, documentación, soporte de la comunidad	
Módulo 2: Inteligencia de los metadatos y del catálogo	
Admite múltiples catálogos (Glue, Hive, Unity, Nessie, Polaris)	
Proporciona gestión activa de metadatos, creación de versiones y pistas de auditoría	
Enriquecimiento semántico basado en IA y descubrimiento de dominios	
Armoniza los metadatos técnicos, empresariales y operacionales	
Módulo 3: Integración de datos y orquestación de procesos	
Latencia: Admite la incorporación en batch y por transmisión (Kafka, CDC, etc.)	
Limpieza de datos, transformación y evolución del esquema incorporadas para todos los formatos	
Permite las herramientas ETL/ELT y las plataformas de ETL inverso	
Permite la optimización y supervisión inteligentes del proceso (basadas en IA)	
Módulo 4: Calidad de datos	
Perfilado de datos, puntuación de calidad y detección de anomalías automatizados	
Controla la desviación, la actualización y la integridad del esquema	
Realiza un seguimiento del linaje de datos en los procesos y formatos	

Cree su ecosistema de datos listo para la IA

Módulo 5: Gobierno, seguridad y cumplimiento	S/N
Control de acceso basado en funciones y aplicación de políticas en todas las plataformas	
Etiquetado de privacidad de los datos y detección automática de datos sensibles	
Respaldo del cumplimiento normativo (RGPD, HIPAA, ISO, etc.)	
Linaje de principio a fin y observabilidad de todos los activos de datos	
Módulo 6: Habilitación de la IA y los análisis	
Entrenamiento del modelo de IA basado en metadatos	
Integración perfecta con BI, ML y plataformas de análisis	
Time travel y consultas versionadas para obtener información reproducible	
Permite procesos de datos fiables y escalables preparados para la IA	
Módulo 7: Escalabilidad operacional y preparación para el futuro	
Escalado elástico y optimización inteligente basada en la carga de trabajo	
Independiente del proveedor: motores multicloud/entornos híbridos/lakehouse	
Plan estratégico alineado con la evolución de data mesh, data fabric y data lakehouse	

Cree su ecosistema de datos listo para la IA

# Acerca de nosotros

Informatica (NYSE: INFA), una empresa líder en la gestión de datos empresariales en el cloud basada en IA, da vida a los datos y la IA al permitir a las empresas materializar el poder transformador de sus activos más importantes. Informatica Intelligent Data Management Cloud (IDMC)<sup>TM</sup>, una solución basada en tecnología de IA que lidera una nueva categoría de software, proporciona una plataforma integral de gestión de datos que conecta, gestiona y unifica los datos prácticamente en cualquier sistema híbrido y multicloud, democratiza los datos y permite a las empresas modernizar sus estrategias empresariales. Clientes de aproximadamente 100 países y más de 80 empresas de la lista Fortune 100 confían en Informatica para impulsar la transformación digital basada en datos.

**Informatica. Donde los datos y la IA cobran vida.**

## INFORMATICA

Paseo de Castellana 91, 2a Planta,  
oficina D17. Madrid 28046

Tel: +34 91 787 61 40

Fax: +34 91 754 29 50

Número gratuito en Estados Unidos:  
1.800.653.3871

[informatica.com/es](http://informatica.com/es)

[linkedin.com/company/informatica  
x.com/Informatica](https://www.linkedin.com/company/informatica-x.com/Informatica)

**PÓNGASE EN CONTACTO  
CON NOSOTROS**

## Where data & AI come to



IN09-5233-0925

© Copyright Informatica LLC 2025. Informatica y el logotipo de Informatica son marcas comerciales o marcas comerciales registradas de Informatica LLC en los Estados Unidos de América y en otros países. La lista actualizada de marcas comerciales de Informatica se encuentra disponible en esta web: <https://www.informatica.com/es/trademarks.html>. Otros nombres de empresas y productos pueden ser nombres comerciales o marcas comerciales de sus respectivos propietarios. La información de este documento está sujeta a cambios sin previo aviso y se proporciona "TAL CUAL", sin garantía de ningún tipo, expresa ni implícita.