

# Maximice el valor de los datos

Cómo preparar sus datos para la innovación  
de la IA

# Contenido

## 03 /

De la dispersión de datos a la preparación para la IA: una base de datos unificada

## 04 /

Aborde las complejidades y desafíos de los datos

## 07 /

Cree una plataforma unificada de datos e IA con un almacén de lago y OneLake

## 11 /

Desbloquee el poder de su entorno de datos con Azure Databricks

## 13 /

Comience a unificar sus datos para la IA

# De la dispersión de datos a la preparación para la IA: una base de datos unificada

**99,1 %** de las empresas encuestadas afirman que invertir en datos e IA es una de las principales prioridades organizacionales<sup>1</sup>

La IA está revolucionando a las industrias, automatizando los procesos y descubriendo información a un ritmo que antes era inimaginable. Impulsados por datos amplios y diversos, la IA y los agentes pueden identificar patrones de manera efectiva, tomar decisiones informadas y llevar a cabo acciones. Esta nueva frontera de sistemas de inteligencia impulsados por IA requiere datos confiables, gobernados y actuales para que los copilotos, agentes y flujos de trabajo de decisiones funcionen de forma fiable.

Sin embargo, muchas iniciativas de IA se estancan por razones familiares. Sus datos están repartidos entre dominios y herramientas. Los equipos mantienen múltiples copias. Los ciclos de actualización no siguen el ritmo de las necesidades empresariales. Las definiciones varían según la función. Cuando los datos están aislados o desactualizados, los análisis y los resultados de IA se vuelven inconsistentes. El nuevo mandato es una forma sencilla y eficaz de limpiar, integrar y enriquecer los activos de datos de la organización.

Cumplir con ese mandato requiere un entorno de datos unificado. Microsoft Fabric es la solución impulsada por IA que unifica experiencias en ingeniería de datos, integración de datos, ciencia de datos, análisis en tiempo real e inteligencia empresarial en un entorno de software como servicio (SaaS) optimizado. En el núcleo se encuentra OneLake, un lago de datos unificado para toda su organización. Unificar su entorno de datos a través de OneLake ayuda a reducir la duplicación y dispersión de datos, a la vez que permite que la información esté disponible en el flujo de trabajo.

Con datos unificados y gobernados, y un significado empresarial compartido, puede potenciar un sistema de inteligencia. Puede personalizar copilotos de IA que ayudan a los equipos a explorar, crear y actuar utilizando lenguaje natural. O puede construir agentes de IA que trabajen junto a las personas para supervisar condiciones, automatizar tareas, conectar herramientas e impulsar flujos de trabajo basados en eventos. La información en tiempo real y la automatización mantienen a los copilotos y agentes basados en los datos más recientes y confiables para que puedan recomendar o iniciar acciones en el momento. Las herramientas impulsadas por IA integradas en la plataforma ayudan a los equipos a hacer más en menos tiempo.

## Acerca de este eBook

Este eBook es para líderes de datos y análisis y para quienes toman decisiones técnicas que desean modernizar su entorno de datos y prepararlo para análisis escalables, aplicaciones de IA y agentes. Ayuda a maximizar el valor de sus datos al realizar lo siguiente:

- Unificar sus datos antes de escalar a IA
- Usar OneLake para simplificar la complejidad
- Construir y escalar una plataforma lista para IA

## Aborde las complejidades y desafíos de los datos

Es tentador impulsar la innovación de la IA lo más rápido posible. Pero logrará mejores resultados cuando comience con la base adecuada. Esto se debe a que muchos entornos de datos fueron optimizados para informes y transacciones, no para el uso continuo y multidominio de IA. Como resultado, es posible que esté gestionando una complejidad creciente a medida que crea y escala modelos y procesos analíticos en sistemas desconectados. Sin embargo, puede empezar a abordar los principales desafíos de datos con dos pasos: unificar los datos y reducir la duplicación.

**92,7 %** de las empresas informan que el interés en la IA ha llevado a un mayor enfoque en los datos<sup>2</sup>



## La necesidad de datos unificados

Para obtener resultados confiables en BI e IA, necesita datos limpios, precisos y consolidados. Cuando los datos están distribuidos en distintos sistemas, los equipos dedican demasiado tiempo a extraerlos, moverlos y transformarlos solo para que puedan usarse. El resultado es información más lenta, mayores costos, definiciones inconsistentes y datos que están desactualizados para cuando llegan a las personas y procesos que los necesitan.

Los entornos integrados conectan herramientas y multiplican copias, mientras que un entorno de datos unificado establece un significado compartido, una gobernanza consistente y la reutilización en todas las cargas de trabajo de análisis y de IA. Una base unificada permite reducir el movimiento innecesario de datos y la fricción asociada, lo que incluye:

### → **Coherencia de datos**

Cuando las definiciones divergen entre equipos y herramientas, sus métricas ya no se concilian y las salidas de IA se vuelven poco confiables. Una estrategia unificada de datos alinea el significado empresarial y la gobernanza, para que los equipos usen definiciones consistentes en todas las cargas de trabajo.

### → **Seguridad y cumplimiento normativo**

Cada transferencia añade una sobrecarga de seguridad y cumplimiento. Reducir el movimiento de datos permite aplicar controles, auditoría y políticas de manera más uniforme en todo su entorno de datos.

### → **Latencia**

Mover grandes volúmenes de datos importantes introduce latencia, lo que retrasa las aplicaciones en tiempo real y las acciones que provocan. Una plataforma unificada reduce las transferencias y el procesamiento previo, de modo que el análisis y la IA se ejecutan sobre los datos más actuales con menos demora.

### → **Rendimiento del modelo**

Los silos de datos limitan el entrenamiento del modelo a datos parciales o inconsistentes, lo que ralentiza el desarrollo y sesga los resultados. Una base unificada brinda a los modelos un acceso más amplio y gobernado a los datos de su organización, lo que mejora la precisión y la coherencia entre los casos de uso.

### → **Transformación de datos**

En entornos fragmentados, los equipos suelen repetir la misma lógica de limpieza y transformación en distintas herramientas. Una plataforma unificada le permite aplicar las transformaciones una sola vez sobre una base compartida y reutilizar tablas y definiciones seleccionadas en distintas cargas de trabajo, lo que reduce el retrabajo y mejora la consistencia.

### → **Costos y uso de los recursos**

El movimiento de datos genera costos recurrentes y recursos de cómputo para las transferencias, almacenamiento para la preparación y esfuerzo operativo para ejecutar y supervisar trabajos. Una plataforma unificada ayuda a reducir estos costos al minimizar el movimiento innecesario y simplificar las operaciones diarias.

### → **Restricciones de ancho de banda de red**

Las transferencias frecuentes de grandes conjuntos de datos agotan la capacidad de la red y pueden generar costos de salida, especialmente con datos de sensores de alto volumen y contenido multimedia. Una plataforma unificada reduce el movimiento entre sistemas para que los equipos puedan trabajar más cerca de donde ya residen los datos.

## Reducir la duplicación

Cuando los mismos datos se almacenan en varios lugares, los duplicados se acumulan a medida que los equipos integran información entre sistemas. Esa duplicación genera confusión sobre qué datos están actualizados y cuáles son correctos, y ralentiza su capacidad de ofrecer análisis y soluciones de IA confiables, ya que se dedica más tiempo a reconciliar registros que a generar información.

Una base unificada reduce la duplicación de datos y sus desafíos, incluidos los siguientes:

### → **Precisión y confiabilidad**

Con una vista única y confiable de los datos, el análisis y los modelos producen resultados más fiables.

### → **Confianza y rendición de cuentas**

Los datos duplicados generan una carga constante de conciliación de impuestos. Los equipos dedican tiempo a investigar discrepancias, trazar el origen de los datos y defender resultados, en lugar de avanzar en su labor.

### → **Productividad y tiempo de amortización**

Con una base confiable y gobernada, los equipos se incorporan más rápido e iteran con mayor agilidad, y dedican menos tiempo a validar datos y más a entregar información, modelos y automatización.

### → **Eficiencia de almacenamiento y recuperación de datos**

Menos copias duplicadas reducen el costo de almacenamiento y mantienen la rapidez en las consultas al minimizar los datos innecesarios que deben escanearse y gestionarse.

### → **Sobrecarga operativa**

Una base unificada simplifica la gestión continua de datos al reducir el manejo de excepciones, los ciclos de corrección y la coordinación requerida para conciliar los conjuntos de datos en competencia.

“Microsoft Fabric fue un punto de inflexión gracias a su capacidad para crear accesos directos sin tener que mover físicamente los datos de un lugar a otro. Antes, si tenía que incorporar tres orígenes, tenía que crear canalizaciones para traer los datos. Esa canalización tenía un costo. El movimiento de datos tenía un costo. Con Fabric, basta con dos clics y listo”.

### **José Viegas**

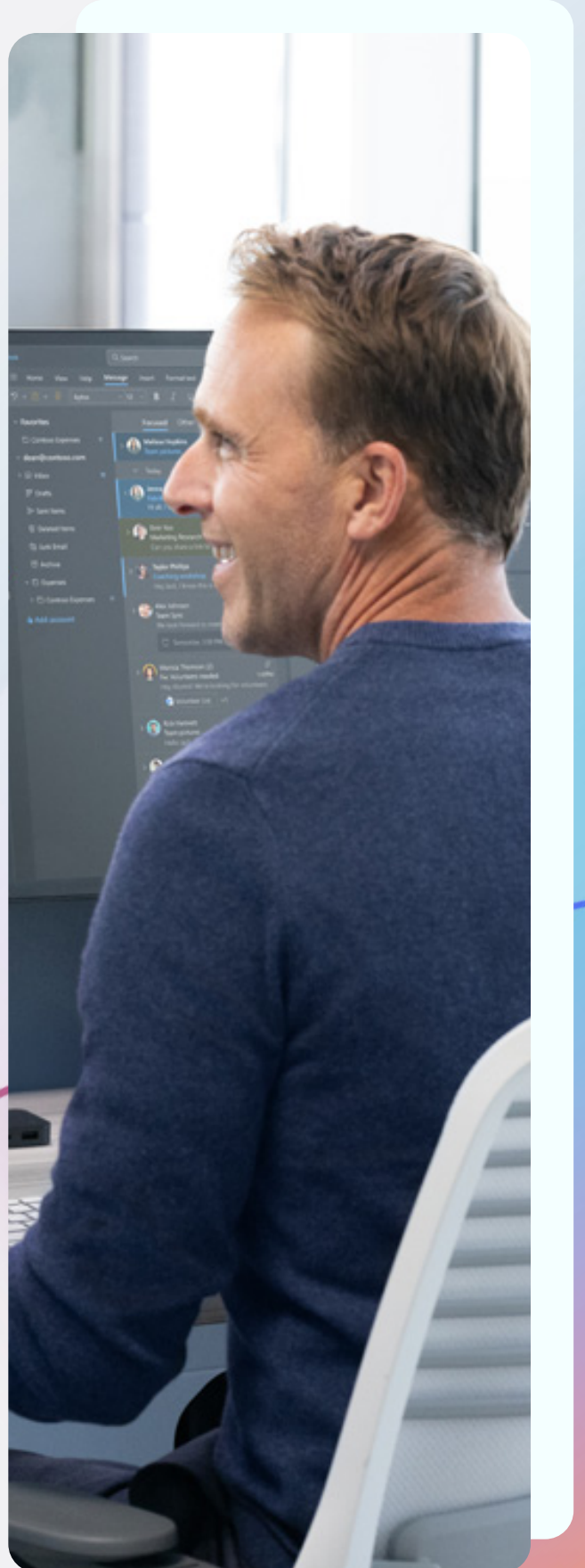
Arquitecto de datos sénior, IWG

[Lea el caso completo](#) 

# Cree una plataforma unificada de datos e IA con un almacén de lago y OneLake

Para reducir el movimiento y la duplicación de datos a gran escala, necesita una base unificada. Por ejemplo, considere a un equipo de comercio electrónico que agrega personalización básica y luego la deja atrás rápidamente a medida que aumenta el tráfico y cambian los comportamientos de los clientes. Si los datos detrás de las recomendaciones se dividen entre sistemas o se actualizan de forma demasiado lenta, los resultados se retrasan y la experiencia se ve afectada.

OneLake es el único lago de datos lógico y unificado para su organización. Almacena datos y los hace accesibles en diferentes cargas de trabajo. Un lago de datos es el patrón arquitectónico que se implementa sobre esa base para combinar la flexibilidad de un lago de datos con la gestión tipo almacén de datos para análisis e IA.



## OneLake: la base para la arquitectura de almacén de lago

Un almacén de lago es una arquitectura moderna de datos que permite almacenar datos estructurados y no estructurados juntos y analizarlos tanto con Spark como con SQL, sin mover datos entre sistemas. Implementado en OneLake, reduce la fricción de dividir los datos en plataformas separadas, lo que facilita la aplicación de una gobernanza consistente y el uso de los mismos datos para análisis e IA.

Sin duplicación ni movimiento de datos, puede unificar los orígenes de datos en múltiples nubes y locales. Los *accesos directos* de OneLake proporcionan acceso sin duplicación a datos en almacenamientos externos, como Amazon S3 y Almacenamiento en la nube de Google. La función de *replicación* de Fabric replica continuamente los datos operativos en OneLake como tablas Delta abiertas para análisis. También puede conectar visualmente fuentes locales y en la nube con más de 150 conectores. Esos datos gobernados se vuelven accesibles en cargas de trabajo, como Data Factory, Ingeniería de datos, Ciencia de datos, Almacén de datos, Inteligencia en tiempo real y Power BI.

Con OneLake como base compartida, obtiene varias ventajas que ayudan a los equipos a escalar análisis e IA, que incluyen:

### → Almacenamiento y procesamiento escalables

Los almacenes de lago escalan para admitir mayores volúmenes de datos y más usuarios con el tiempo. En Fabric, los datos de los almacenes de lago se almacenan en OneLake y pueden procesarse con múltiples motores, incluido Spark y SQL, a medida que aumentan sus necesidades.

### → Una única fuente de verdad

Un almacén de lago ayuda a estandarizar datos consistentes y gobernados a través de equipos y herramientas. OneLake organiza los datos de almacén de lago y almacén en una única plataforma de datos para toda la organización.

### → Acceso de copia cero con accesos directos

Los accesos directos y la replicación le permiten conectarse a los datos donde ya residen y hacerlos disponibles en su almacén de lago sin crear otra copia. Puede reducir la duplicación y las demoras en la preparación, mientras mantiene el acceso gobernado mediante los permisos y credenciales existentes de la plataforma.

### → Optimizados para análisis e IA

La arquitectura de almacén de lago estandariza los datos en tablas Delta abiertas que varios motores pueden utilizar desde el mismo origen, lo que mejora el rendimiento y la coherencia para las cargas de trabajo de análisis e IA.

### → Baja latencia de consultas

Los equipos pueden recuperar información rápidamente para informes interactivos y cargas de trabajo de IA. OneLake reduce la latencia causada por copias duplicadas y transferencias, lo que permite rutas de acceso rápidas, como Direct Lake sobre tablas Delta en el almacén de lago.

### → Actualización de los datos

Los almacenes de lagos mantienen los datos actualizados mediante transmisión y actualizaciones frecuentes desde sistemas transaccionales. OneLake permite que esos datos estén disponibles entre equipos y cargas de trabajo, mediante replicación para una copia casi en tiempo real y atajos para acceder a los datos en su ubicación original.

### → Gobernanza y seguridad de los datos

OneLake centraliza los permisos y controla el acceso tanto a nivel de administración como de datos. El catálogo de OneLake centraliza la gestión de seguridad y la integración con Microsoft Purview añade etiquetas de sensibilidad, prevención de pérdida de datos y auditoría para cumplimiento. Y ahora puede aplicar el acceso basado en roles hasta el nivel de carpetas o tablas, incluidas las restricciones de nivel de fila y columna.

## Desde la base hasta capacidades listas para IA

OneLake proporciona el fundamento de datos compartido, pero preparar los datos para la IA también requiere un conjunto de capacidades esenciales para reducir el movimiento, limitar la duplicación y preparar los datos para la IA. Basadas en OneLake, estas capacidades trabajan juntas sobre los mismos datos gobernados, en vez de crear nuevos silos.

### Integración de datos

Al optimizar la integración de datos, puede reducir los costos de almacenamiento, simplificar la ingeniería y mejorar la confianza en los datos compartidos. Puede trabajar con réplicas casi en tiempo real de fuentes como Azure SQL, Azure Database for PostgreSQL y Azure Cosmos DB para acceder fácilmente a todos sus datos en OneLake.

Además, Fabric puede trabajar con las tablas de Unity Catalog de Azure Databricks para reflejar el metadato del catálogo y acceder a los datos subyacentes mediante accesos directos en OneLake a fin de evitar el movimiento y la duplicación de datos.

### Transformación de datos

Puede transformar datos con flujos de datos de código bajo y automatizar el proceso completo con canalizaciones. Los flujos de datos en Fabric le permiten aplicar limpieza y modelado paso a paso, y las canalizaciones programan y coordinan flujos de trabajo de extremo a extremo. Los almacenes de lagos de Fabric también admiten Spark y SQL sobre los mismos datos, por lo que puede publicar tablas seleccionadas una sola vez y reutilizarlas en diferentes cargas de trabajo.

### Transmisión de datos

Para actuar sobre lo que está sucediendo ahora, necesita convertir las señales entrantes en información oportuna y respuestas automatizadas. Esto es esencial para flujos de trabajo de agentes, decisiones basadas en eventos y escenarios de la Internet de las Cosas, donde los sistemas detectan patrones como umbrales, anomalías o cambios de estado y responden de inmediato.

Fabric Real-Time Intelligence ingiere, transforma de manera opcional y enruta eventos en tiempo real sin necesidad de código. Fabric Activator supervisa continuamente los flujos con latencia inferior a un segundo para desencadenar acciones posteriores, como notificaciones o flujos de Power Automate, cuando se cumplen las condiciones.

### Consulta de datos

Las consultas permiten obtener la porción adecuada de datos —características, registros y ventanas de tiempo— de modo que los modelos de análisis e IA se entrenen y validen con entradas relevantes y de alta calidad. Puede realizar consultas a almacenes de datos y puntos de conexión SQL con T-SQL, almacenes de lago con Spark y SQL o a datos en tiempo real y de registro con KQL. Luego puede aplicar esos resultados a través de modelos semánticos y Direct Lake para un análisis rápido y gobernado.

### Visualización y colaboración de datos

Cuando los datos gobernados se convierten en conocimiento compartido, sus equipos pueden monitorear resultados, identificar cambios y tomar decisiones más rápido. Fabric incluye Power BI como una carga de trabajo principal para la generación de informes interactivos sobre los datos en la plataforma. Direct Lake facilita el análisis interactivo sobre grandes tablas Delta almacenadas en el lago, lo que disminuye la sobrecarga de actualización en comparación con los patrones tradicionales de importación.

Además, Microsoft Copilot en Fabric está disponible en todas las cargas de trabajo para ayudarle a interactuar con sus datos mediante lenguaje natural, transformarlos y analizarlos, generar información y crear visualizaciones e informes en Fabric y Power BI.

“Ayudamos a los usuarios a desbloquear el valor de sus datos haciendo preguntas específicas de su negocio para tomar mejores decisiones basadas en datos”.

### Srihari Kumar

Director de Productos, Información y Datos, JLL

[Lea el caso completo](#) 

#### **Bases de datos de Azure: entregue aplicaciones y agentes de IA de manera más rápida, con capacidades integradas para IA generativa y cargas de trabajo de LLM**

Azure ofrece servicios totalmente gestionados para SQL, NoSQL, PostgreSQL y MySQL, lo que proporciona la base para cargas de trabajo críticas de cualquier tamaño que su organización utiliza para lograr la transformación impulsada por IA y la innovación empresarial. Con Azure Databases, las empresas obtienen hasta cinco veces más rendimiento, detección inteligente de amenazas y hasta un 58 % menor costo total de propiedad.<sup>3</sup> Las nuevas aplicaciones, agentes de IA y soluciones impulsadas por datos se benefician del desarrollo asistido por GitHub Copilot, recuperación sin contacto, latencia de un solo dígito en milisegundos, capacidades avanzadas de búsqueda vectorial e integraciones con las herramientas de desarrollo, plataformas de análisis y tecnologías de IA más populares. Con Azure Databases, sus datos se convierten en un activo estratégico, lo que ayuda a convertir sus ambiciones de IA en resultados comerciales reales.

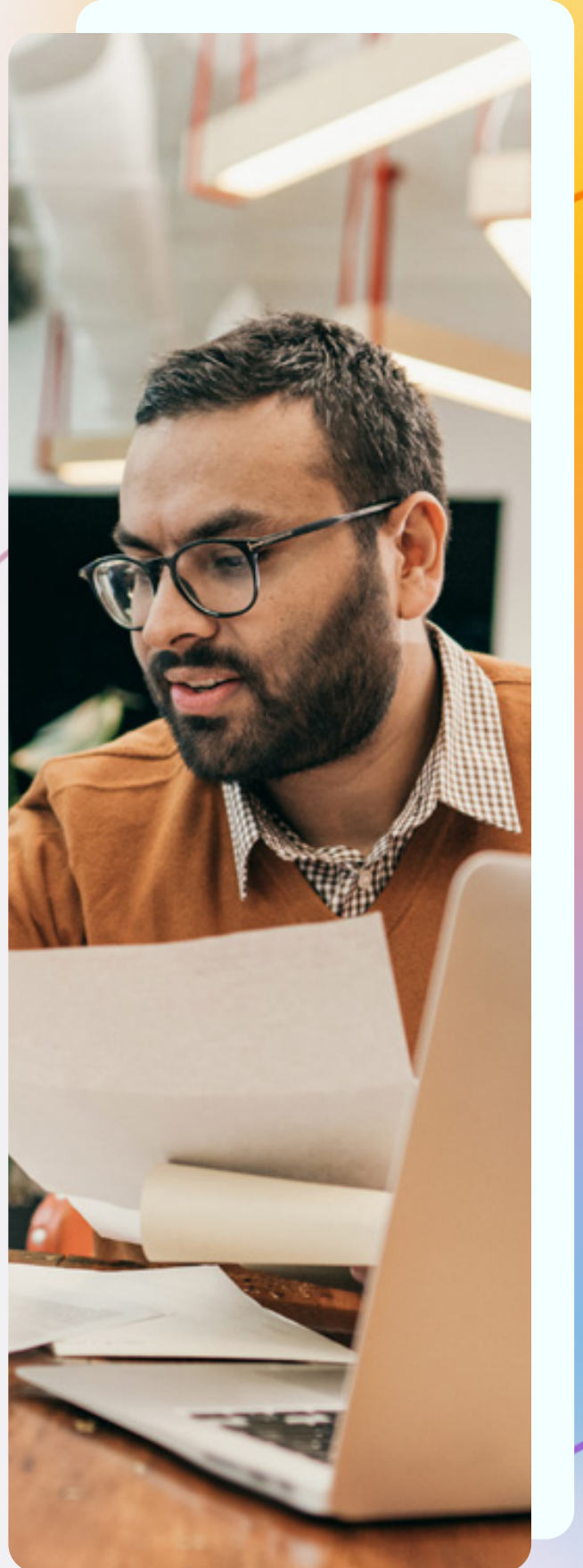
# Desbloquee el poder de su entorno de datos con Azure Databricks

En el panorama en evolución del análisis de datos, es fundamental seleccionar la plataforma de nube adecuada. Databricks, ampliamente adoptado, proporciona una plataforma unificada para datos, análisis e IA basada en estándares abiertos y confiables. Azure Databricks representa una asociación única entre Databricks y Microsoft. Una colaboración de ingeniería conjunta garantiza una integración sólida y un rendimiento óptimo, lo que le permite aprovechar un entorno de datos verdaderamente unificado. Sus equipos de ingeniería de datos pueden trabajar con la plataforma de inteligencia de datos de Databricks con la interoperabilidad de Unity Catalog con OneLake para acelerar el análisis y la IA abiertas y gobernadas en Azure.

Los beneficios clave incluyen:

## → Integración perfecta

Azure Databricks está diseñado para trabajar con Entra ID y otros servicios de Azure para identidad y autenticación, almacenamiento y servicios de observabilidad, lo que facilita la configuración y gestión de sus cargas de trabajo de datos.



### → Escalabilidad y rendimiento

Los clústeres pueden escalar automáticamente hacia arriba o hacia abajo según las demandas de la carga de trabajo, para un uso óptimo de los recursos y eficiencia en costos. Azure ofrece una variedad de opciones de cómputo, incluidas instancias habilitadas para GPU, que aceleran las cargas de trabajo de "machine learning" y el aprendizaje profundo.

### → Seguridad y cumplimiento normativo

Azure cuenta con certificaciones de cumplimiento líderes de la industria, lo que garantiza que sus cargas de trabajo de Azure Databricks cumplan con los estándares reglamentarios.

### → Optimización de costos

Azure proporciona una experiencia de facturación unificada, lo que le permite administrar los costos de Azure Databricks sin problemas junto con otros servicios de Azure. Puede aprovechar las instancias de acceso puntual de Azure para reducir costos en cargas de trabajo no críticas o elegir Instancias Reservadas para ahorrar en el uso a largo plazo de Azure Databricks. También tiene opciones para supervisar y administrar sus costos al usar Databricks IA en Azure Databricks.

### → Servicios integrados

Dentro del ecosistema de Azure, puede conectar Azure Databricks con Microsoft Foundry para modelado avanzado de IA, y trabajar con Power BI para visualizaciones y reportes en tiempo real, entre muchas otras integraciones, como Microsoft Fabric, Copilot Studio y Power Platform.

"Elegimos Azure Databricks porque funciona bien con nuestro ecosistema existente de Azure, lo que nos permite aprovechar y ampliar nuestras inversiones actuales. Esto significaba que nuestros equipos podían continuar usando sistemas conocidos mientras accedían a nuevas funciones de análisis."

### Benedict Ho

Director adjunto principal, Departamento de DSA,  
Central Provident Fund Board

[Lea el caso completo](#) 

## Comience a unificar sus datos para la IA

Para que la IA genere un valor real para el negocio, comience con una base unificada de datos que cada equipo pueda confiar y reutilizar. El camino más rápido hacia el impacto es reducir la fricción: menos silos, menos copias y menos tiempo dedicado a reconciliar cuál versión es la correcta.

Una plataforma integral brinda a cada equipo lo que necesita para pasar de datos sin procesar a inteligencia impulsada por IA. En todos los roles y departamentos, sus usuarios pueden explorar los datos de manera más eficiente, obtener información valiosa y crear aplicaciones y agentes de IA, impulsando la transformación de IA de su organización con una plataforma de datos completa.

## Dé los siguientes pasos

- ➔ **Comience con una plataforma de datos completa a través de Microsoft Fabric**
- ➔ **Obtenga más información sobre el valor de sus datos con Azure Databricks.**
- ➔ **Inicie el desarrollo de la IA con Microsoft Foundry.**

<sup>1</sup>Bean, Randy and Davenport, Thomas H. "Survey: How Executives Are Thinking About AI in 2026". Harvard Business Review. 12 de enero de 2026.

<sup>2</sup> *Ibid.*

<sup>3</sup> Las cifras relacionadas con el rendimiento y el ahorro de costos se extraen de estudios encargados por Microsoft y de casos prácticos de clientes. Los resultados reales pueden variar en función de la carga de trabajo, la configuración y el contexto de implementación. Consulte, por ejemplo, McAfee, Nathan. "The Economic Benefits of Migrating On-premises Databases to Azure Database for PostgreSQL". Enterprise Strategy Group. Marzo de 2025.