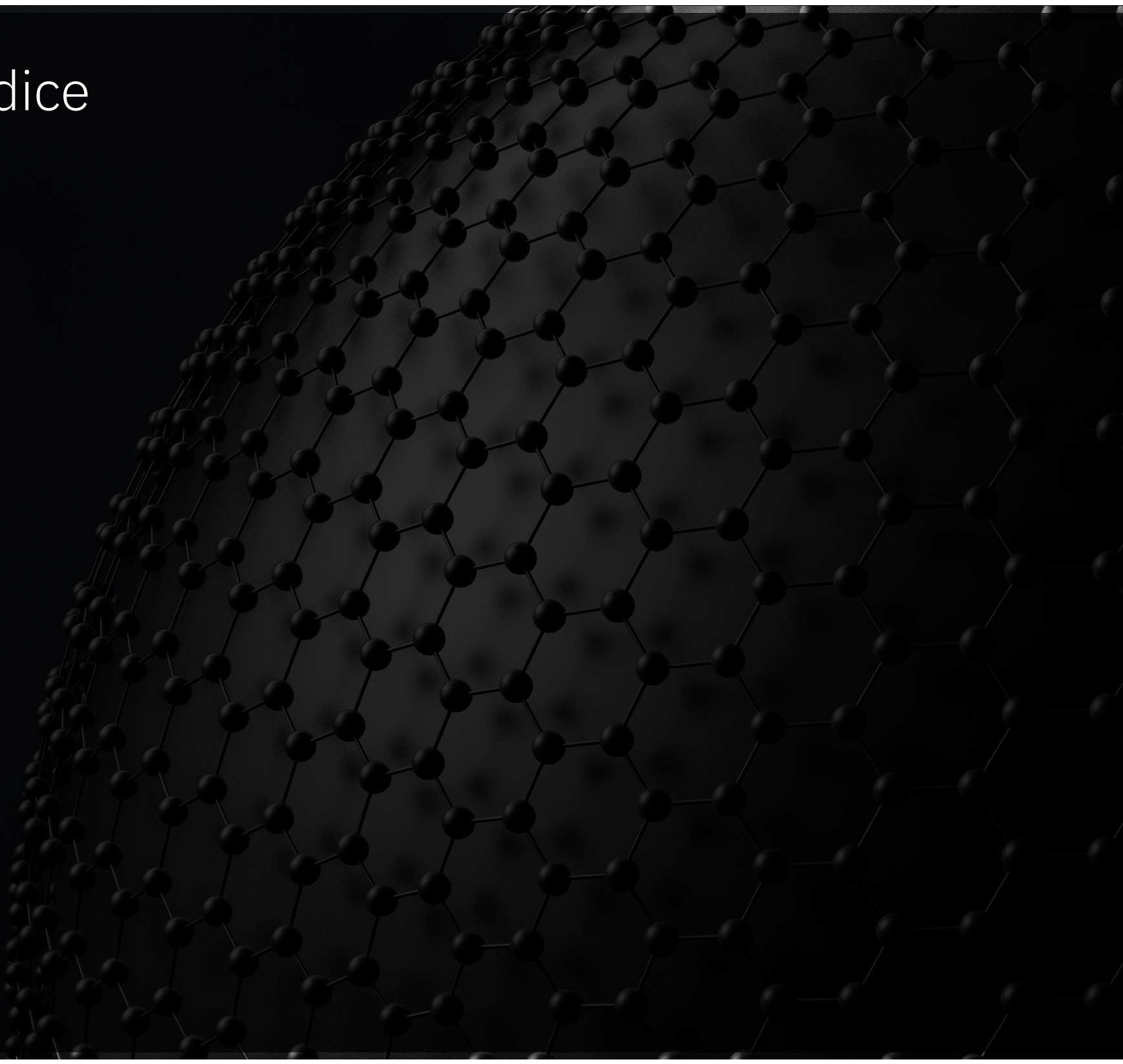


Db2

La base
de datos
de la IA



Índice



- 3 La base de datos de la IA:
basada en y creada para la IA
- 4 [Basada en IA](#)
- 5 Virtualización de datos
- 6 Gestión adaptativa del volumen de trabajo
y la optimización de los recursos
- 9 Optimización de consultas mediante
machine learning
- 10 Consultas basadas en la confianza
- 11 [Creada para la IA](#)
- 12 Exploración de datos más rápida con
consultas en lenguaje natural
- 14 Utilización de lenguajes de programación
populares y ejemplos de código para
desarrollar aplicaciones basadas en IA
- 17 Modelado de relaciones complejas con
grafos y SQL
- 18 Análisis de datos de blockchain de forma
nativa
- 19 [La importancia de Db2 en una
organización impulsada por la IA](#)

La base de datos de la IA: basada en y creada para la IA

El aumento del volumen y la variedad de los datos exige nuevos conocimientos, procesos e infraestructura para impulsar la transformación digital en todas las industrias. El machine learning (ML) y la inteligencia artificial (AI) son partes fundamentales de este esfuerzo, ya que los ordenadores procesan los datos y aprenden de su experiencia sin la continua intervención humana.

La IA y el ML contribuyen a la realización de tareas informáticas tradicionales y no tradicionales, desde el análisis y la logística hasta la interacción con el lenguaje natural, la composición musical, el diseño químico y la detección de fraudes. Acercan la inteligencia basada en datos a las empresas, de modo que estas puedan comprender los hechos acontecidos, los que están sucediendo y los que es probable que ocurran en el futuro.

El camino hacia la IA es como una escalera donde cada componente es un peldaño. Quienes emprenden este camino fortalecen la escalera y la hacen más útil.

Incorporar

Ponga la IA en funcionamiento en todo su negocio.

Analizar

Cree y escale la IA con confianza y transparencia.

Recopilar

Haga que los datos sean simples y accesibles.

Modernizar

Prepare sus datos para un mundo de IA y nubes híbridas.

Organizar

Diseñe una base analítica adecuada para su negocio.

El camino hacia la IA se basa en una sólida infraestructura de datos que permite a las organizaciones acceder, recopilar, organizar y hacer accesibles todos los datos para su análisis, independientemente de su naturaleza, fuente o despliegue. Gracias a la gestión de datos basada en ML, la IA representa tanto el objetivo como el medio para lograrlo.

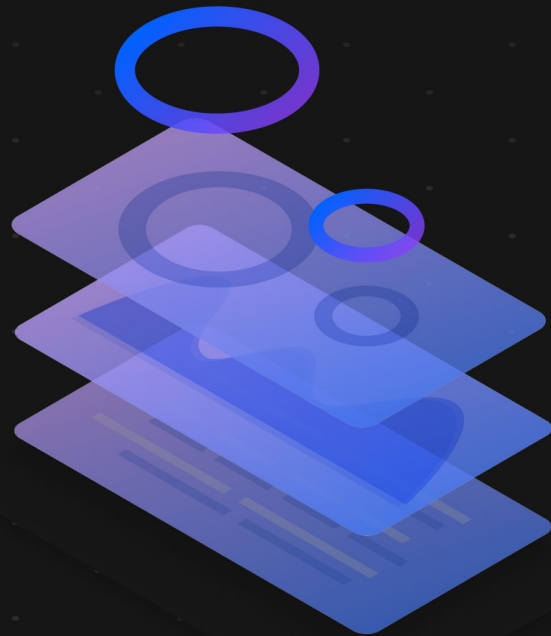
Las soluciones integradas provistas de IA y ML hacen que la gestión de datos sea más sencilla, rápida e inteligente, lo que permite a los desarrolladores y científicos de datos utilizar los datos de nuevas formas e incorporar las capacidades de la IA en toda la empresa, desde las operaciones hasta la creación de modelos y el desarrollo de aplicaciones.

Basada en IA

Las funcionalidades de machine learning e IA integradas en IBM® Db2® impulsan diversas funciones de gestión de datos innovadoras, entre las que se incluyen:

- Virtualización de datos de pila completa
- Gestión automática de la carga de trabajo y la optimización de los recursos
- Rendimiento de consultas 10 veces mayor
- Resultados de consultas basadas en la confianza

Virtualización de datos



Qué es

Una combinación de federación de datos junto con una capa de abstracción. Permite a todos los usuarios y aplicaciones interactuar con múltiples fuentes de datos a la vez desde un único punto de acceso, independientemente de la naturaleza, el formato, el tamaño o la ubicación de los datos subyacentes. Esta mayor accesibilidad contribuye a eliminar los silos de datos y permite que los datos se utilicen de forma más rápida y eficiente para escalar la adopción de la IA.

Por qué es importante

Disponer de un único punto de entrada a los datos proporciona las siguientes ventajas:

– Acceso simple a los datos para sus profesionales de datos

El hecho de tener una visión única de todos los datos permite realizar búsquedas y acceder a ellos a través de diversas fuentes de datos: estructuradas o no estructuradas, relacionales o NoSQL sin tener que invertir tiempo y recursos en mover los datos a otro lugar. Los desarrolladores e ingenieros de datos tienen acceso a todos los tipos de datos posibles, tanto históricos como recientes, y pueden extraer información de las combinaciones inesperadas de esos datos.

– Un único punto de control para el gobierno y la seguridad

Las disposiciones relativas a la seguridad y la gobernanza son mucho más sencillas, más sólidas y menos propensas a los fallos cuando sus administradores pueden simplemente dirigir a todos los usuarios y aplicaciones a un único punto de acceso para todos sus depósitos de datos.

– Reducción de las transferencias de datos

El procesamiento de los datos puede llevarse a cabo en los propios repositorios, sin necesidad de transferir previamente los datos consultados a un repositorio local. Esto puede reducir drásticamente los costes de latencia y de ancho de banda.

Cómo funciona

Los entornos de datos de las empresas modernas dependen de múltiples almacenes de datos. Entre ellos se incluyen los datos transaccionales almacenados en una base de datos relacional, los datos de sentimientos almacenados en un clúster Hadoop, los datos históricos de clientes almacenados en un almacén de datos y los datos de comercio electrónico de clickstream almacenados en un clúster de datos rápidos.

Sin la virtualización, todas las personas que quieran interactuar con estos datos, desde los administradores de bases de datos (DBA) hasta los desarrolladores de aplicaciones y científicos de datos, se verían obligados a implementar un método de acceso personalizado para cada almacén de datos. En el mejor de los casos, supondría un enorme dolor de cabeza para todo aquel que estuviese involucrado. Limitaría la información que los usuarios podrían obtener de la comparación de diferentes tipos de datos entre sí (como, por ejemplo, el historial de los clientes con el comportamiento de navegación en tiempo real). Y, en el peor de los casos, se correría el riesgo de que se produjera una corrupción de los datos y nuevas brechas de seguridad.

Proporcionar un único punto de acceso a diferentes datos en toda la empresa es el motivo por el que la federación de datos se ha convertido en algo tan popular. Sin embargo, la virtualización de datos de IBM va más allá de la federación, ya que proporciona una capa de abstracción y permite a los usuarios definir su propia terminología y metodología para el acceso a los datos (véase la figura 1).

En acción

Virtualización de datos

En la actualidad, el banco europeo ING colabora con IBM para implementar un punto único de acceso a los datos para todos sus usuarios a nivel mundial. Esto le permite al banco administrar una infraestructura de datos global al mismo tiempo que ofrece el rendimiento, la escalabilidad, la seguridad y las características de gobierno que necesitan. Asimismo, esto reduce las desconexiones entre las diferentes áreas de la empresa.

Gracias a la solución de IBM, se pueden añadir nuevos datos a la plataforma ING desde cualquier lugar del mundo y a los que cualquiera de los usuarios internacionales del banco puede acceder sin necesidad de modificar los esquemas de acceso a los datos individuales o los permisos de seguridad.



Figura 1. Representación del modo en que la virtualización permite el acceso y la gestión unificados de múltiples fuentes de datos.

Gestión adaptativa del volumen de trabajo y la optimización de los recursos

Qué es

Tecnología inteligente que asigna automáticamente recursos de datos para gestionar una variedad de cargas de trabajo. Esta tecnología está automatizada y es adaptativa, ya que responde a las condiciones anteriores y también predice las demandas futuras.

Por qué es importante

Con la gestión adaptativa de la carga de trabajo, se obtiene un sistema estable y de alto rendimiento, sin necesidad de realizar complicados ajustes y puestas a punto. Esto es esencial para las empresas modernas debido a la complejidad de sus plataformas de datos y a la alta tasa de fallos de las estrategias de gestión de la carga de trabajo manual.

Cuando hay demasiadas cargas de trabajo a la vez, una base de datos debe determinar cómo gestionar la demanda adicional. Los llamados esquemas “abiertos” o manuales no tienen ningún mecanismo de respuesta; simplemente

requieren que el usuario establezca límites predeterminados para determinar el número o el tamaño de las cargas de trabajo. Pero esos límites pueden desaparecer cuando se enfrentan a un conjunto complejo de cargas de trabajo, e incluso en las mejores circunstancias requieren una monitorización y un ajuste manual constantes. Esto da lugar a problemas de rendimiento de la base de datos, a una infrautilización o incluso a fallos de esta.

Por el contrario, Db2 detecta y predice las tendencias de utilización, de modo que puede notificar a los usuarios o incluso corregir automáticamente el problema antes de que la tendencia se convierta en algo serio. Esto simplifica el trabajo de los DBA y reduce el coste de la propiedad de cada base de datos al reducir el tiempo de intervención necesario para la configuración y la puesta a punto. Las pruebas de IBM demostraron que esta mejor utilización produjo mejoras en el rendimiento general de la base de datos de hasta un 30 %.





Cómo funciona

La gestión adaptativa de la carga de trabajo incorpora el machine learning como mecanismo de respuesta, de modo que la base de datos supervisa constantemente el tiempo de ejecución previsto y real de las diferentes cargas de trabajo y, a continuación, ajusta los recursos disponibles para hacer frente a cualquier carencia. Las bases de datos Db2 pueden funcionar sin necesidad de una configuración mínima para la mayoría de las cargas de trabajo.

Esta tecnología adaptativa facilita el proceso en caso de que se necesite una configuración manual para situaciones especialmente complejas. Los DBA pueden crear múltiples clases de cargas de trabajo y especificar objetivos de rendimiento para cada una de ellas. A continuación, el gestor adaptativo monitoriza las cargas de trabajo entrantes y asigna recursos para alcanzar esos objetivos.

En acción

Gestión automatizada de la carga de trabajo

La gestión automatizada y adaptativa de las cargas de trabajo aborda específicamente muchos de los puntos débiles más comunes que conlleva la gestión de una base de datos empresarial.

Por ejemplo, en los entornos de almacenamiento de datos, la base de datos puede recibir una gran variedad de trabajos, desde ingestas en flujo hasta informes de baja latencia con tiempos de respuesta inferiores a un segundo (como la carga de páginas web), así como informes de lotes más pesados que necesitan minutos para completarse. El administrador debe asegurarse de que el sistema no se bloquee en los trabajos más pesados e impida que se completen los trabajos urgentes.

El administrador de cargas de trabajo adaptativas facilita enormemente esta labor, de modo que los administradores pueden crear agrupaciones de recursos (también denominadas clases de cargas de trabajo o de servicios) y asignar una parte del sistema a cada agrupación. A continuación, la base de datos asigna los recursos de forma inteligente para alcanzar los objetivos de rendimiento. Todo esto ocurre sin necesidad de monitorizar y ajustar manualmente las cargas de trabajo y los recursos de forma individual.

La IA puede convertirse tanto en el objetivo como en el medio para alcanzar ese objetivo.



Optimización de consultas mediante machine learning

Qué es

El optimizador Db2 basado en machine learning proporciona un nivel adicional de optimización inteligente que utiliza machine learning no supervisado para ofrecer estrategias de ejecución de consultas que mejoran la optimización tradicional de consultas basadas en los costes.

Por qué es importante

Los optimizadores de costes básicos de cargas de trabajo pueden sugerir estrategias de ejecución para una consulta determinada, pero no son capaces de detectar cambios recientes en la base de datos y no pueden aprender de la experiencia. Aunque pueden recomendar la estrategia de ejecución más rápida, continuarán recomendando esa misma estrategia, aunque no funcione como se espera.

Por el contrario, el optimizador Db2 basado en machine learning incorpora la información del rendimiento real de la consulta para recomendar estrategias de ejecución que ofrezcan los mejores resultados, concretamente, en su infraestructura de datos y perfeccionar la ruta de la consulta con cada ejecución.

Con un optimizador basado en machine learning, los DBA no tienen que dedicar su tiempo a supervisar el rendimiento del sistema y a tratar de optimizar las consultas. En su lugar, pueden centrarse en actividades que aporten más valor a la organización como la implementación de aplicaciones de IA y el desarrollo de estrategias de uso de datos, así como contribuir a que los usuarios empresariales de toda la organización hagan un mejor uso de los datos disponibles.

Cómo funciona

El optimizador Db2 basado en machine learning imita los patrones de la red neural para optimizar las rutas de las consultas, lo que permite que algunas consultas se completen con una velocidad hasta 8-10 veces mayor (pruebas internas de IBM).

En acción

Optimización de consultas mediante ML

Un optimizador de costes tradicional utiliza modelos estadísticos y de recursos para evaluar las estrategias de ejecución de una consulta determinada. Devuelve la primera opción que se muestra en la figura 2 a continuación: unir dos pares de tablas y, a continuación, unir los resultados para devolver una respuesta. Sin embargo, un algoritmo de machine learning puede aprender de la experiencia y comprender que existe una estrategia de ejecución superior: unir un par de tablas, unir una tercera tabla a ese resultado, y luego unir una cuarta tabla a ese resultado para devolver la misma respuesta.

Sin machine learning



Con machine learning

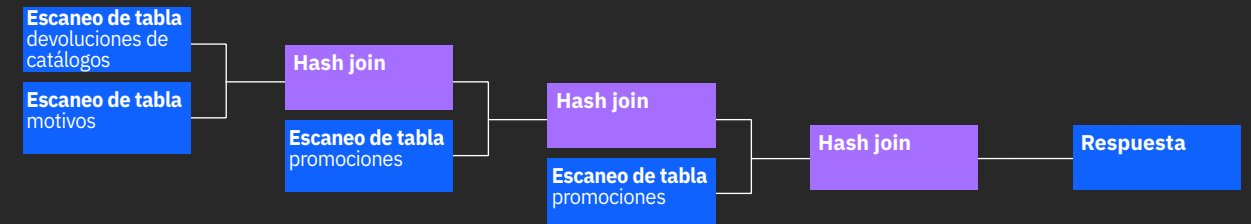


Figura 2. Ejemplo de estrategias de ejecución de consultas optimizadas en función de los costes (arriba) y optimizadas en función del machine learning (abajo).

Estos cambios en la estrategia de ejecución pueden aplicarse en tiempo real, en respuesta a las ejecuciones de consulta reales. En el caso de las aplicaciones dependientes del tiempo, como la detección de fraudes, esta mejora en tiempo real puede ser crítica.



Consultas basadas en la confianza

Qué es

Una función avanzada que ofrece resultados de consultas SQL en términos de probabilidades, o “mejores coincidencias”, en lugar de una simple respuesta afirmativa o negativa.

Por qué es importante

La correspondencia basada en la confianza está disponible en los ajustes de machine learning, pero esta característica de Db2 se extiende también a las expresiones SQL. Para un ingeniero de SQL, esta característica amplía drásticamente la cantidad de posibles tareas de datos que pueden realizarse sin la intervención de un científico de datos, lo que convierte a los ingenieros de SQL en personal más valioso para la empresa y disminuye la carga de los científicos de datos con exceso de trabajo.

Cómo funciona

Las consultas basadas en la confianza añaden extensiones de machine learning a SQL mediante la implementación de redes neuronales de alimentación profunda. Esta variedad de machine learning profundo sin supervisión encuentra áreas donde tanto la “semejanza” como la probabilidad de encontrar una coincidencia son altas. Las consultas basadas en la confianza se pueden aplicar a:

- consultas de similitud/disimilitud
- consultas de razonamiento inductivo (por ejemplo, agrupamiento semántico, analogías, odd-man-out)
- operaciones semánticas de grupo
- anomalías en los patrones (por ejemplo, detección de fraudes)
- imágenes, audio y vídeo

En acción

Consultas basadas en la confianza

La base de datos de la policía es un escenario con muchas variables (altura, peso, edad, color de la piel, color de los ojos, rasgos distintivos, etc.) y prácticamente todas ellas son imprecisas debido a la falta de fiabilidad de los testigos. Una búsqueda tradicional en una base de datos utiliza una consulta SQL muy compleja para responder a esta incertidumbre, ya que crea manualmente un “rango” de valores que se corresponderán con la declaración de un testigo determinado. Por ejemplo, una declaración de un peso de 88 kg puede relacionar a los sospechosos con pesos reales de entre 84 y 93 kg. Estos rangos son difíciles de utilizar y pueden eliminar por error a los sospechosos si el valor proporcionado por un testigo se encuentra demasiado lejos del rango. En cambio, un agente que intente establecer una correspondencia entre las declaraciones de los testigos y el perfil de un sospechoso puede utilizar una consulta basada en la confianza para generar una simple declaración SQL probabilística que encuentre al sujeto que mejor se ajuste al perfil general de los testigos, incluso cuando los puntos de datos individuales estén fuera del rango original de correspondencias aceptables.

Creada para la IA

Db2 también está desarrollada para la IA, lo que significa que sus profesionales de datos pueden habilitar con mayor facilidad potentes actualizaciones de la carga de trabajo impulsadas por la IA y mejoras en el flujo de trabajo. Entre estas se incluyen:

- la exploración de datos a mayor velocidad con consultas en lenguaje natural;
- la utilización de lenguajes de programación populares y ejemplos de código para desarrollar aplicaciones basadas en IA;
- el modelado de relaciones complejas con grafos y SQL;
- el análisis de datos de blockchain de forma nativa;

Exploración de datos a mayor velocidad con consultas en lenguaje natural

Qué es

La aplicación Augmented Data Explorer es una interfaz basada en la web “similar a la de Google” para explorar datos comerciales de gran complejidad. Ofrece búsqueda en lenguaje natural para los usuarios y REST API para los desarrolladores. Una pregunta formulada en lenguaje natural generará visualizaciones y resúmenes de datos extraídos de todas las fuentes de datos disponibles.

Por qué es importante

A medida que el panorama de los datos se hace más complejo, un mayor número de usuarios se enfrenta a conjuntos de datos cada vez más voluminosos. Las habilidades para desarrollar consultas SQL o aplicaciones analíticas son escasas, pero cada vez más personas quieren obtener respuestas de los datos, y en muy poco tiempo.

Augmented Data Explorer es una herramienta de exploración de datos en fase inicial. Permite realizar sofisticadas consultas de datos a usuarios no técnicos, que simplemente necesitan introducir una consulta en lenguaje natural como “ventas para 2019” o “¿con qué frecuencia llueve en octubre?”. Esto generará una declaración SQL para

consultar todos los datos relevantes y devolver el resultado deseado en forma de lenguaje natural, una visualización dinámica de datos, o ambos.

En el pasado, los usuarios comerciales generalmente necesitaban una combinación de procesos manuales para traducir las solicitudes en lenguaje natural a la consulta SQL necesaria y el asesoramiento de expertos para hacer frente a los inevitables inconvenientes que se presentaban. Augmented Data Explorer se ha diseñado para evitar ambos requisitos y ofrecer nuevas vías para encontrar nuevos valores en los datos ya existentes.

Los científicos de datos y los usuarios más sofisticados pueden utilizar Augmented Data Explorer para realizar consultas iniciales sobre conjuntos de datos de gran tamaño o desconocidos, para obtener así una idea de lo que contienen esos conjuntos de datos y determinar dónde pueden comenzar las investigaciones. Augmented Data Explorer también ofrece REST API, que permite a los desarrolladores integrar fácilmente estas capacidades de búsqueda y predicción en lenguaje natural en sus propias aplicaciones en lugar de tener que invertir tiempo en codificar su propia solución.



Cómo funciona

Augmented Data Explorer utiliza la tecnología de ML para rastrear e indexar los datos con el fin de descubrir información oculta. Sus capacidades de IA se enfocan en el apoyo al lenguaje natural (por ejemplo, para entender el contexto, los sinónimos y la sintaxis de una amplia variedad de disciplinas y campos), así como en la predicción de temas de interés relacionados basados en los términos de búsqueda. La herramienta se encuentra en un contenedor para facilitar su despliegue y gestión, y cuenta con una sencilla interfaz basada en la web.

También incluye funcionalidad de grafos, con la que se puede descubrir a simple vista cuáles son los conjuntos de datos a los que el motor del explorador ha accedido y rastreado. El usuario puede ver fácilmente qué datos (como esquemas y tablas) se utilizan para producir las respuestas que proporciona el sistema.



El valor de Augmented Data Explorer

Augmented Data Explorer resulta más valioso en situaciones en las que sus usuarios de datos no saben con certeza qué preguntas hacer. Facilita a estos usuarios la posibilidad de hacer preguntas, descubrir las relaciones dentro de los datos y desarrollar conocimientos procesables. Además, amplía el grupo de personas que pueden abordar estos retos empresariales. En los entornos empresariales en constante cambio, donde los usuarios tienen muchas responsabilidades, Augmented Data Explorer representa el primer paso para hacer frente a la abrumadora cantidad de datos que se encuentran disponibles en la actualidad y obtener nuevo valor de estos.

Piense en un gerente de ventas de una empresa minorista. Es posible que, cada día o cada mes, esta persona se plantee muchas preguntas diferentes, desde desglosar las ventas por región hasta comprender las tasas de retención de clientes, identificar las causas de abandono de los carritos de la compra, predecir por qué en ocasiones no vuelven los clientes habituales o explicar cómo afecta el clima a los plazos de entrega de los inventarios.

Augmented Data Explorer está diseñado para identificar y responder rápidamente a nuevas preguntas, lo que permite expandir las capacidades de la IA a un segmento mucho más amplio de operaciones comerciales.

Utilización de lenguajes de programación populares y ejemplos de código para desarrollar aplicaciones basadas en IA

Qué es

Db2 11.5 incluye soporte nativo para lenguajes y bibliotecas conocidas, incluidos Python, JSON, GO, Ruby, PHP, Java, Node.js, Sequelize y Jupyter Notebooks. Los profesionales de datos pueden desarrollar aplicaciones basadas en IA en un lenguaje de programación que dominan para, posteriormente, conectarlas a los datos en Db2.

Por qué es importante

Al disponer de soporte nativo para los lenguajes de programación para machine learning más populares, IBM acorta distancias entre la gestión de datos y el rápido desarrollo de aplicaciones con un ecosistema de datos centrado en el ML y la IA. Los desarrolladores y científicos de datos pueden mantener sus datos en una base de datos empresarial y seguir utilizando los lenguajes de programación y las bibliotecas más populares que existen en la actualidad.

El soporte de lenguajes de programación y de bibliotecas de forma nativa aporta múltiples ventajas que conducen a un desarrollo más rápido y fácil de las aplicaciones basadas en IA:

- Los desarrolladores pueden aprovechar esos lenguajes y bibliotecas populares, junto con ejemplos de códigos específicos de la industria en GitHub, para desarrollar rápidamente aplicaciones destinadas a sus propios usos.
- Los desarrolladores pueden desarrollar aplicaciones que utilicen datos Db2 de forma nativa con facilidad, lo que elimina la necesidad de un código back-end personalizado para acceder a los datos Db2, y elimina un punto común de fallo o administración para sus aplicaciones.
- Las nuevas bases de datos pueden desplegarse rápidamente y escalarse en la nube si es necesario y, posteriormente, pueden conectarse sin problemas a las aplicaciones.
- Los desarrolladores tienen acceso a entornos de desarrollo de machine learning (IBM Watson® Studio) dentro del mismo ecosistema de bases de datos.
- Los científicos de datos pueden invertir su tiempo en modelos de machine learning en lugar de resolver problemas de acceso a las bases de datos.
- Las empresas pueden contratar a personas con las habilidades necesarias para analizar los datos sin necesidad de aprender nuevas habilidades o lenguajes de programación, únicamente para acceder a los datos.

Cómo funciona

El proceso de desarrollo de aplicaciones basadas en IA plantea varias cuestiones importantes:

- ¿Dónde se almacenan los datos relevantes?
- ¿Cómo se puede acceder a ellos y explorarlos?
- ¿Cómo se aplican las prácticas de machine learning mientras se desarrolla la aplicación?
- ¿Qué lenguajes de programación se pueden utilizar?

Dentro del ecosistema de IBM, existen respuestas empresariales a todas estas preguntas. Los datos se pueden almacenar en Db2 o Db2 Warehouse, IBM Db2® Event Store o Hadoop, en la nube o de forma local, en prácticamente cualquier combinación de bases de datos. Todos ellos tienen en común la disponibilidad empresarial con total apoyo, fiabilidad y rendimiento. La base de código Db2 y el motor SQL común subyacente permiten que las aplicaciones puedan consultar datos de cualquier componente del ecosistema sin necesidad de reescribir código para su ejecución en diferentes partes del ecosistema. Asimismo, gracias a una serie de nuevos controladores disponibles para los principales lenguajes y marcos de programación de código abierto, ahora resulta más sencillo para los desarrolladores analizar e incorporar modelos de machine learning en las aplicaciones que utilizan Db2.

Los profesionales de datos pueden utilizar algunas de las otras características descritas en este libro electrónico, como el Augmented Data Explorer, para obtener una rápida comprensión de la información que contienen los datos. Una vez almacenados en Db2, los datos pueden conectarse directamente al entorno de desarrollo de machine learning en Watson Studio. Las bibliotecas específicas de la industria, los ejemplos de código y las plantillas proporcionan a los desarrolladores una ventaja en cuanto al desarrollo de una aplicación funcional. Por último, un conjunto de controladores nativos convierte las sentencias SQL en un modelo específico de lenguajes para ML (como Python) para su posterior desarrollo en el lenguaje de programación seleccionado.

El valor del soporte nativo de lenguajes de programación

El mercado potencial para las aplicaciones habilitadas para la IA es enorme. Algunos ejemplos son:

Comercio minorista: motores de recomendación, contenido dirigido y promociones

Seguros: Analíticas de fraudes y procesamiento de solicitudes

Logística: gestión de flotas y mantenimiento predictivo

Transporte: rutas en tiempo real y guía de recorridos basada en, entre otros, el tráfico o el clima

Existe una legión de desarrolladores y científicos de datos que trabajan en el mercado para desarrollar aplicaciones y algoritmos inteligentes, por lo general mediante el uso de Python, Java u otro lenguaje de programación popular. Los datos que alimentan estas aplicaciones se encuentran en una o más bases de datos. Hay bases de datos que se conectan de forma nativa a Python, pero, normalmente, no están equipadas para ofrecer la escala, el rendimiento o la fiabilidad que requiere una empresa. Sin embargo, sin una integración nativa entre los lenguajes de programación para ML seleccionados y la base de datos empresarial subyacente, es difícil y engorroso para los desarrolladores conectar la propia aplicación con los datos necesarios.

Gracias a la integración nativa entre Db2 y los lenguajes de programación más populares, los desarrolladores pueden acceder a los datos sin necesidad de recurrir a un especialista en DBA o SQL. Esto se traduce en una mayor productividad para sus desarrolladores y científicos de datos, además de aplicaciones más robustas y desarrolladas con mayor rapidez.

La IA permite a usuarios de cualquier nivel aprovechar los datos de manera más eficiente que antes.



Modelado de relaciones complejas con grafos y SQL

Qué es

Db2 integra profundamente la funcionalidad de los grafos con los datos relacionales y SQL de modo que las aplicaciones de grafos pueden ejecutarse directamente desde los datos relacionales, mientras que el motor SQL de Db2 puede consultar directamente los datos de los grafos.

Por qué es importante

IBM Graph es una potente herramienta que proporciona información importante, aunque anteriormente era incompatible con los sistemas transaccionales y OLTP. Para evitar esta incompatibilidad, las organizaciones mantenían su infraestructura relacional ordinaria, pero extraían determinados datos, los introducían en una base de datos de grafos y ejecutaban aplicaciones de grafos en ella. Este tipo de duplicación, retraso y sobrecarga ya no es aceptable. Las organizaciones necesitan una visión basada en grafos en sus datos relacionales, y los necesitan en segundos.

En muchas industrias, los usuarios desean añadir información en forma de grafos a su actual conjunto de herramientas de analítica, pero sin alterar su infraestructura de datos ni añadir una nueva base de datos únicamente para respaldar los grafos. Db2 hace esto posible.

Cómo funciona

Los datos de los grafos se almacenan en tablas dentro del entorno de trabajo relacional de Db2 o Db2 Event Store. Esto significa que el motor SQL puede consultar los datos de los grafos directamente, y las aplicaciones de grafos pueden consultar los datos directamente en las tablas relacionales. Esta arquitectura también soporta lenguajes de consulta de grafos de código abierto, como Gremlin y Tinkerpop.

La profunda integración entre los datos relacionales y las capacidades de grafos en Db2 ofrece varias ventajas importantes:

- Los grafos se pueden ejecutar directamente contra los datos relacionales existentes, lo que añade un nuevo nivel de información además de las analíticas basadas en SQL.
- Las analíticas SQL pueden ejecutarse directamente contra los datos de los grafos almacenados en las tablas relacionales.
- Las aplicaciones de grafos pueden utilizar una API personalizada o Spark para conectarse a Db2.
- Las transacciones ACID actualizan los grafos en tiempo real sin afectar a las aplicaciones relacionales existentes, lo que significa que los grafos pueden utilizarse para el procesamiento de transacciones en tiempo real, además de para realizar analíticas.
- Existen soluciones específicas para industrias que requieren de grafos con gran cantidad de datos, como el sector sanitario y el financiero.



El valor del soporte nativo de grafos

Existen muchos casos en los que la información basada en grafos puede permitir nuevas eficiencias y flujos de trabajo con datos relacionales. Por ejemplo, un comerciante minorista puede utilizar relaciones de grafos en tiempo real para registrar un patrón de devoluciones de un artículo determinado en múltiples ubicaciones. De este modo, esta información puede actualizar la base de datos de transacciones del minorista de forma inmediata y hacer que el artículo se retire de las estanterías el mismo día.

En un ámbito diferente, un investigador de fraudes al seguro puede utilizar un grafo para reunir múltiples conjuntos de datos tales como el número y el tamaño de las reclamaciones, las relaciones entre los reclamantes o las identidades de los proveedores de servicios, entre otros. Las relaciones que surgen en el grafo pueden revelar la probabilidad de fraude.

Análisis de datos de blockchain de forma nativa

Qué es

La API Blockchain Connector nativa en Db2 proporciona una transparencia sin precedentes y una visión de los datos altamente comprimidos contenidos en los registros contables de las cadenas de bloques.

Por qué es importante

La tecnología blockchain se ha adoptado rápidamente en muchas industrias, pero este hecho genera cuellos de botella analíticos. Existen enormes cantidades de información potencialmente valiosa en forma comprimida en cada registro de la cadena de bloques, pero aún no se ha encontrado una forma sencilla de ver y analizar esa información. Db2 Blockchain Connector cambia esa situación. Estos registros compactos, transparentes, controlables e inalterables de blockchain están ahora abiertos al análisis, al igual que cualquier otra fuente de datos de la empresa. Las empresas obtienen todos los beneficios del motor Db2 (rendimiento, flexibilidad, escalabilidad y seguridad) sin necesidad de desarrollar soluciones de informes ad-hoc únicamente para blockchain.

El conector blockchain es aún más valioso debido a la conectividad e interoperabilidad nativa de otros tipos de datos disponibles en Db2. Las empresas pueden integrar fácilmente datos adicionales de otros almacenes de datos para ganar contexto y abrir nuevas posibilidades analíticas para sus datos de blockchain. Por último, este conector abre la posibilidad de utilizar los datos de blockchain para aplicaciones de IA, un caso de uso que anteriormente habría sido muy engorroso debido a que los datos de blockchain eran, efectivamente, su propio silo. Los desarrolladores de IA pueden ahora incorporar fácilmente conjuntos de datos de blockchain, ya sea como fuente de datos primaria para sus aplicaciones o para proporcionar detalles granulares adicionales.

Cómo funciona

Blockchain Connector extrae los datos transaccionales que se comprimen y almacenan dentro del registro de blockchain y los presenta como una tabla relacional en Db2. Una robusta estrategia de almacenamiento en caché utiliza la funcionalidad Db2 existente para crear una tabla de caché para los datos de blockchain con el fin de mejorar el rendimiento de la consulta y, al mismo tiempo, permitir a los usuarios obtener los datos más recientes cuando sea necesario.



El valor del soporte nativo de blockchain

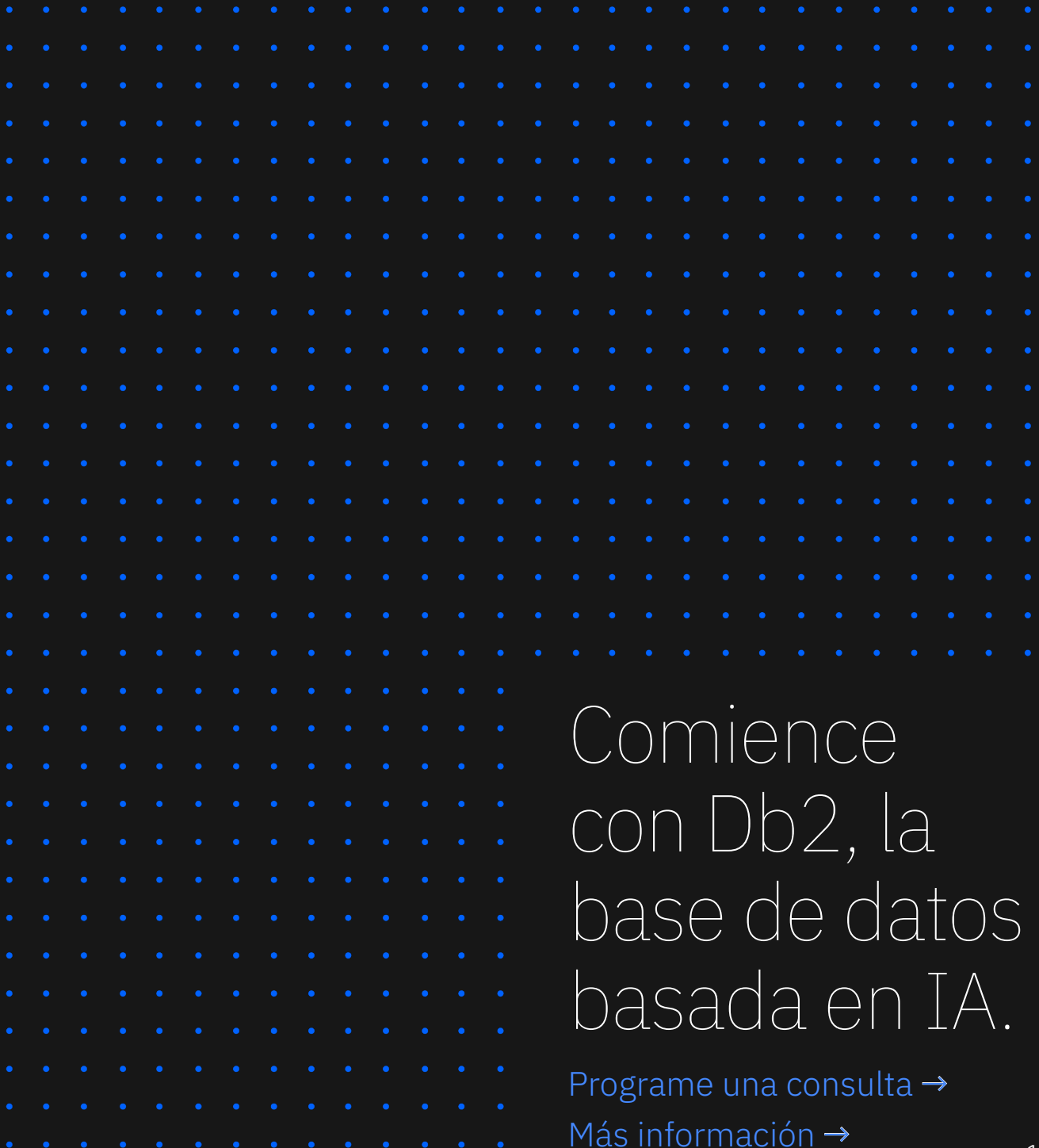
En la actualidad, industrias como la de seguros, finanzas, logística y sanitaria utilizan la tecnología blockchain para almacenar un registro inmutable de transacciones, pero estas industrias también quieren realizar analíticas para comprender mejor sus datos. Por ejemplo, las empresas de seguros quieren saber cuántas reclamaciones se han tramitado por cliente en una región geográfica determinada, mientras que las empresas de transporte quieren hacer un seguimiento de los contenedores y vehículos, y las empresas del sector sanitario quieren conocer los historiales de tratamiento de los pacientes.

En lugar de desarrollar un back-end personalizado para extraer estos datos, pueden simplemente conectar Db2 a su almacén de datos de blockchain y ejecutar sus analíticas, del mismo modo que lo harían con cualquier otro dato. Debido a la conectividad y flexibilidad inherentes de Db2, pueden correlacionar datos contextuales adicionales con los datos transaccionales de blockchain (por ejemplo, el clima durante un envío particular o las personas asociadas a un reclamante de seguros que podrían haber presentado reclamaciones similares) y obtener una información mucho más rica y completa. Ahora es posible ejecutar aplicaciones de IA directamente sobre datos de blockchain.

La importancia de Db2 en una organización impulsada por la IA

Tanto si despliega las capacidades de Db2 como un despliegue tradicional de forma local, en una base de datos relacional en la nube o en un almacén de datos, la plataforma de gestión de datos híbridos por suscripción (Hybrid Data Management Platform), el nuevo IBM Cloud Pak™ for Data u otro modelo de despliegue, la tecnología base es emocionante y los beneficios empresariales son espectaculares.

Db2 tiene la capacidad de contribuir a que la IA sea una realidad para los usuarios de organizaciones de cualquier tamaño. La propuesta de valor es clara. Desarrolle, ejecute y gestione la IA de su empresa en cualquier infraestructura, de forma local o en la nube. Aproveche todas sus fuentes de datos, sin importar su ubicación física. Aproveche las capacidades de la IA infundida en la base de datos para ofrecer una visión predictiva y proactiva de las decisiones basadas en los datos. Además, con la base de datos Db2, las empresas pueden desarrollar y conectar aplicaciones basadas en IA de forma más eficiente, lo que permite obtener rápidamente conocimientos útiles y hacer que la IA sea ahora accesible a un mayor número de personas.



Comience
con Db2, la
base de datos
basada en IA.

Programe una consulta →
Más información →



© Copyright IBM Corporation 2020. IBM España

Santa Hortensia, 26-28
28002 Madrid
España

Producido en Estados Unidos en enero de 2020.

IBM, el logotipo de IBM, **ibm.com**, Db2, IBM Cloud Pak e IBM Watson son marcas comerciales de International Business Machines Corp. registradas en diversas jurisdicciones de todo el mundo. Otros nombres de productos y servicios pueden ser marcas comerciales de IBM o de otras empresas. Encontrará una lista de las marcas actuales de IBM en el apartado "Copyright and trademark information" de la página web: ibm.com/legal/copytrade.shtml.