

A short vertical green line is positioned to the left of the text.

CALIDAD Y GOBERNANZA DE DATOS

Cómo maximizar el valor de los datos en la era de la IA

Los productos de datos aceleran los resultados al tiempo que mejoran la gobernanza

En el acelerado panorama empresarial actual, los datos son un activo estratégico vital, una pieza clave para optimizar las operaciones, así como impulsar la innovación y el crecimiento. Los datos no solo reflejan cifras y estadísticas, también incluyen conocimientos, tendencias y patrones que pueden iluminar el camino hacia el éxito. Ahora más que nunca, con la atención centrada en la IA generativa y el aprendizaje automático, las empresas abordan una pregunta fundamental:

¿Cómo acceder rápidamente al valor integral que ofrecen los datos para acelerar los resultados empresariales, mejorando al mismo tiempo la gobernanza de los datos y la confianza en ellos?

En esta ecuación la velocidad es un elemento clave. Según un estudio reciente de IDC, el 75 % de los responsables de la toma de decisiones afirma que los datos pierden su valor en cuestión de días.¹

El reto va más allá de gestionar volúmenes abrumadores de datos. Implica depurar la información para convertirla en activos de buena calidad, seleccionados y de fácil acceso, adecuados para dominios empresariales específicos. Pero, como cualquiera que trabaje con datos ya sabe, es más fácil decirlo que hacerlo. Los responsables de los datos y análisis se enfrentan a varios retos que dificultan la plena realización de sus iniciativas de datos:

- **La gestión de datos orientada a proyectos carece de escalabilidad.** Los equipos de cada proyecto desarrollan su trabajo con su propio conjunto de parámetros, incluidos el alcance, los objetivos y las partes interesadas. Este enfoque fragmentado dificulta la colaboración interfuncional y limita la capacidad de extraer valor de los datos. Además, existen numerosos riesgos: desalineación de las partes interesadas entre los proyectos, esfuerzos duplicados, brechas de gobernanza, uso ineficiente de los datos, mayores costes y, en última instancia, pérdidas o retrasos de los resultados empresariales.

- En el conjunto de la empresa, **hay una falta de coherencia en la calidad de los datos entre los sistemas**, lo que socava los procesos de toma de decisiones y menoscaba la confianza en los conocimientos basados en datos.
- Toda la cadena de suministro de datos y valor **carece de una clara propiedad y rendición de cuentas** entre los productores de datos y los consumidores, lo que frustra el intento de extraer valor al ritmo del negocio.
- **Emplear soluciones fragmentadas y enfoques propios basados en necesidades puntuales** para la gestión de datos amplía el tiempo de valorización, y resulta costoso de escalar y mantener sin personal cualificado.

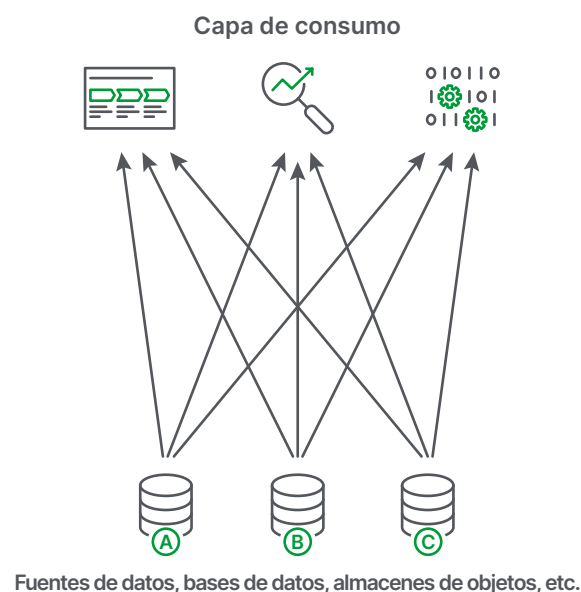
Estos retos amplían la brecha entre los productores y los consumidores de datos.

En un extremo, los productores de datos, que se centran en las TI, dan más importancia a las plataformas de datos que a comprender el uso de los datos y las necesidades de los consumidores. En el otro extremo, los consumidores de datos, atrincherados en sus dominios empresariales, tienen dificultades para transmitir los requisitos y no confían en los datos proporcionados. **Para cerrar esta brecha cada vez mayor, las organizaciones necesitan un enfoque estratégico y simplificado para la gestión de datos.**

¹ Blog de IDC, «Navigating the Planes of Enterprise Intelligence Architecture» (Recorrido por los planos de la arquitectura de la inteligencia empresarial), junio de 2023.

Cómo modificar los enfoques de la gestión de datos

Tradicionalmente, las organizaciones han empleado dos enfoques diferenciados para extraer valor de sus datos: el enfoque **fragmentado** y el enfoque de **datos centralizados**.



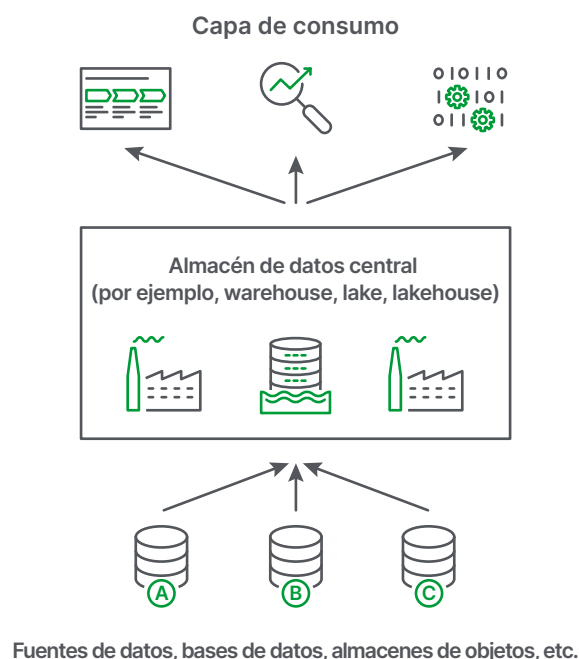
EL ENFOQUE FRAGMENTADO

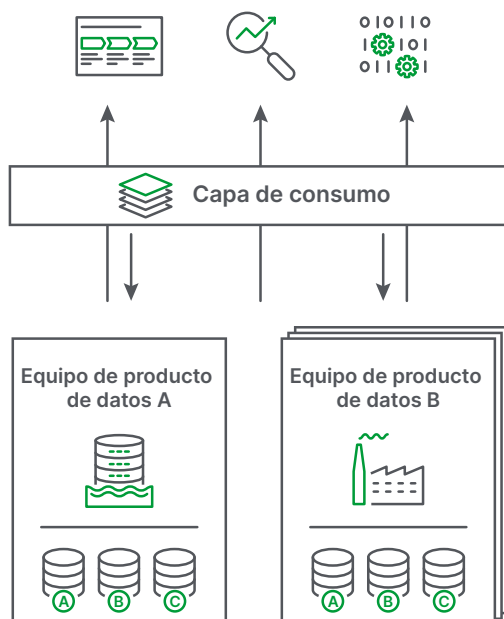
En el enfoque fragmentado, los equipos de «proyectos de datos» individuales recopilan datos de diversos sistemas de origen para abordar casos de uso empresariales específicos. Cada equipo consume los datos de manera independiente, utilizando sus propias herramientas y tecnologías para extraer, integrar y analizar datos para sus necesidades específicas. Aunque este método genera un valor inmediato a corto plazo, no se puede escalar para grandes empresas a largo plazo, lo que genera más silos de datos, esfuerzos duplicados, aumento de los costes y retrasos en los resultados empresariales.

EL ENFOQUE CENTRALIZADO

Por el contrario, un enfoque centralizado designa a un equipo de datos centralizado como responsable de extraer, limpiar y agregar los datos a gran escala. El objetivo del equipo central es abordar los requisitos analíticos de la organización de manera integral. Establecen canales de datos complejos para recuperar datos de diversos sistemas, limpiarlos y consolidarlos en data lakes, warehouses o lakehouses en la nube (como Delta Lake de Databricks, Snowflake, Big Query, S3, etc.).

Puede que la centralización parezca la solución a la fragmentación, y hasta cierto punto lo es, pero tiene un punto flaco fundamental: exige una amplia coordinación entre los departamentos para realizar cambios o actualizaciones, lo que impide la agilidad organizativa. Además, al centrarse en requisitos de datos comunes para toda la organización, el equipo de datos central a menudo pasa por alto las necesidades y preferencias específicas de las partes interesadas empresariales y los consumidores, lo que da lugar a una solución deficiente que no responde a los requisitos empresariales.





El enfoque federado ofrece varias ventajas clave:



Agilidad y flexibilidad

Al simplificar la gestión de datos de extremo a extremo, un enfoque federado permite a las organizaciones **acceder rápidamente al valor empresarial** de los datos. Promueve la agilidad al adoptar una mentalidad basada en los productos, por lo que la entrega y la creación de valor se distribuyen entre los equipos de productos interfuncionales, mientras que las funciones de ingeniería de plataformas se desempeñan de manera centralizada. Esto permite **reaccionar con mayor rapidez ante las necesidades empresariales cambiantes**, al tiempo que se garantiza la rendición de cuentas y la integridad de los datos durante todo el proceso.



Rentabilidad

El modelo federado adopta la **reutilización**, lo que permite a las organizaciones maximizar sus inversiones y reducir los costes operativos. Además, el intercambio centralizado de componentes básicos entre los diferentes equipos de productos **minimiza el trabajo redundante y mejora la capacidad de mantenimiento a largo plazo**.

EL ENFOQUE FEDERADO

Ambos enfoques tradicionales, el fragmentado y el centralizado, tienen sus limitaciones que incluyen carencias a la hora de satisfacer las necesidades de consumo de datos de las empresas. Sin embargo, se está produciendo un cambio de paradigma como respuesta a estos retos. El **enfoque federado**, que cada vez es más popular, distribuye estratégicamente las bases de datos aprovechando los productos de datos al tiempo que gestiona de manera centralizada funciones de ingeniería de plataformas.

Varios equipos de productos de datos pueden trabajar en paralelo, cada uno centrado en los productos de datos específicos de su dominio. Este enfoque acelera los resultados empresariales al aprovechar la colaboración interfuncional, que incluye a las partes interesadas tanto empresariales como técnicas.



Implementación acelerada de casos de uso

Al facilitarse la colaboración entre los equipos interfuncionales, se generan **resultados rápidos específicos del dominio**, personalizados para cumplir los requisitos empresariales cambiantes y las nuevas iniciativas basadas en datos. El enfoque federado ofrece **escalabilidad y permite implementar casos de uso personalizados específicos del dominio** mediante flujos de desarrollo paralelos en varios equipos de productos de datos.



Mejora de la confianza

El modelo federado **fomenta la confianza gracias a una gobernanza, transparencia y conformidad mejoradas**, lo que infunde confianza en la toma de decisiones basada en datos. La clara propiedad de los dominios de datos fomenta la alineación y la rendición de cuentas entre las partes interesadas de datos y análisis, para fortalecer aún más el marco de gobernanza de datos de la organización.

Introducción a los productos de datos

Un enfoque federado es el punto de partida más adecuado para los productos de datos porque, para muchas organizaciones, implica un cambio evolutivo que va del control a la colaboración, del proyecto al producto y de la rigidez a la agilidad, todo ello al tiempo que proporciona valor a escala. La [malla de datos](#) y la [ingeniería de datos](#), cada una con sus enfoques característicos de la gestión de datos, pueden combinarse para formar una potente sinergia y desbloquear todo el potencial de los productos de datos. En el paradigma de la malla de datos, la propiedad de datos está descentralizada, ya que quienes son propietarios de los datos y los gestionan son los equipos que los usan. El producto de datos es el aspecto más tangible e impactante del paradigma de la malla de datos, y la ingeniería de datos se convierte en la arquitectura de gestión de datos fundamental, que permite una entrega óptima de los productos de datos a los equipos de dominio.²



Según el informe de CDAO Agenda Survey de Gartner de 2024, el 50 % de las organizaciones participantes ya ha implementado productos de datos, y otro 29 % se ha comprometido a realizar pruebas piloto o a plantearse la implementación durante el próximo año.³

Entonces, ¿qué es un producto de datos? Un producto de datos ofrece bases de datos de gran calidad, seleccionadas, listas para usarse y preparadas para la IA, a las que las personas pueden acceder con facilidad y que se pueden aplicar a diferentes casos de uso empresariales en una organización. Se puede considerar que un producto de datos es un microservicio de datos. Al igual que los microservicios exponen capacidades de aplicación mediante unidades más pequeñas, más contextuales y componibles, un producto de datos proporciona acceso a datos pertinentes y orientados al dominio a través de interfaces de servicio. Un producto de datos cumple varios principios fundamentales:



Orientación al dominio y propiedad

Cada producto de datos debe identificarse con un dominio específico y debe tener un propietario de datos (persona o equipo) responsable de su éxito, es decir, de proporcionar valor y aumentar su adopción.



Pensamiento orientado al producto

Los productos de datos se crean aplicando un pensamiento orientado al producto en relación con los datos, que incluye la propiedad, las versiones y los clientes (en este caso, los consumidores de datos). Los productos se mejoran comprendiendo los patrones de uso y escuchando la opinión de los clientes. Al igual que los productos reales, tienen un ciclo de vida.



Estandarizado e interoperable

Los productos de datos cumplen formatos, esquemas y convenciones de metadatos estandarizados. Facilitan la interoperabilidad y la facilidad de integración con otros componentes del ecosistema de datos.



Fácil de encontrar y reconocer

Los productos de datos están registrados en un catálogo de productos de datos y los consumidores de datos pueden encontrarlos y reconocerlos con facilidad. Un producto de datos incluye su propia descripción: qué datos representa, cómo se puede acceder a él, su semántica empresarial y su linaje y calidad.



Reutilizable

Los productos de datos se crean una vez y se usan muchas veces en diversos casos de uso. Se crean otros productos de datos a partir de productos de datos existentes.

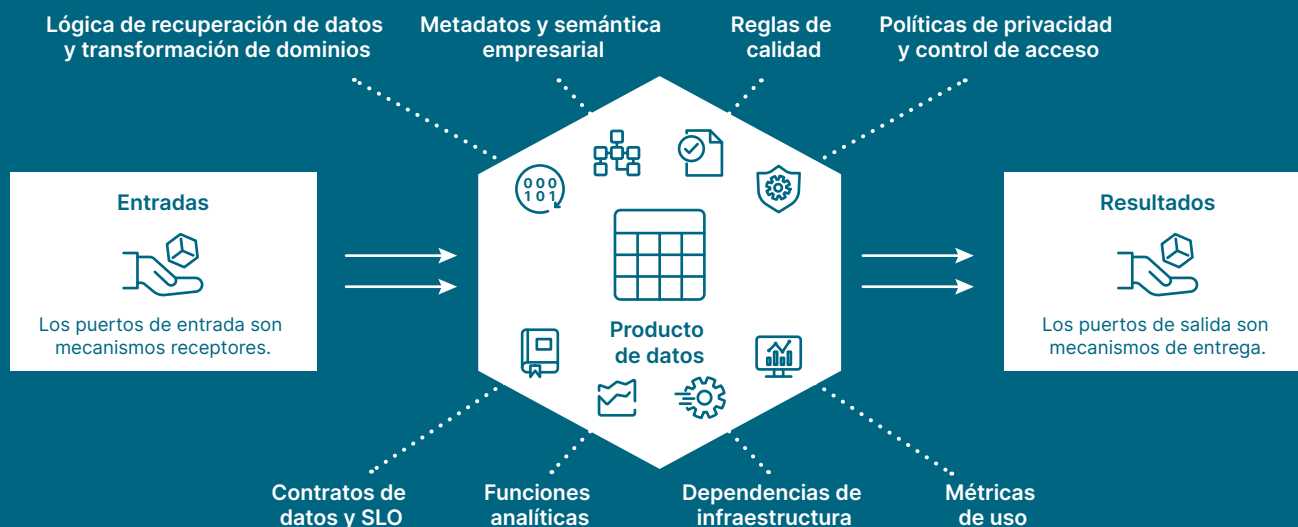


Accesible

Se accede a los productos de datos mediante un conjunto estándar de interfaces como SQL o una API REST que se definen y exponen cuando se crean los productos de datos. Las interfaces ocultan la complejidad del producto de datos.

² Gartner: 6 Lessons Data Leaders Can Learn From the Early Adopters of Data Mesh (Seis lecciones que los responsables de datos pueden aprender de los primeros usuarios de la malla de datos), por Robert Thanaraj, et al. 4 de marzo de 2024. GARTNER es una marca comercial registrada y una marca de servicio de Gartner, Inc. o de sus filiales en EE. UU. e internacionalmente. Se hace uso de ella en este documento con su permiso. Todos los derechos reservados.

³ Gartner: Gartner Chief Data and Analytics Officer Agenda Survey for 2024 (Encuesta de Gartner sobre la agenda de los directores de datos y análisis de 2024), 12 de marzo de 2024. Los resultados de este estudio no representan conclusiones globales ni al mercado en su conjunto, sino que reflejan la opinión de los encuestados y de las empresas encuestadas. GARTNER es una marca comercial registrada y una marca de servicio de Gartner, Inc. o de sus filiales en EE. UU. e internacionalmente. Se hace uso de ella en este documento con su permiso. Todos los derechos reservados.



Un producto de datos es una agrupación independiente de datos, transformaciones, código, políticas de gobernanza, acuerdos de nivel de servicio (SLA), patrones de acceso y dependencias de infraestructura. En el diagrama anterior se muestra una anatomía del producto de datos.

- 1 **Bases de datos:** existen en el núcleo central del producto. Pueden ser un conjunto de tablas, archivos, gráficos o representaciones vectoriales. Dentro de un producto de datos puede haber varios niveles de datos: un conjunto sin procesar que representa datos ingeridos desde un sistema de origen, un conjunto seleccionado que representa datos limpios e integrados, y una base de datos lista para consumo y modelada para facilitar el uso.
- 2 **Entradas:** representan los puertos de entrada o el origen de entrada de los datos. Podrían ser tablas en una aplicación SAP que deben ingerirse, tablas en Snowflake que deben usarse como origen, otro producto de datos del que se deriva uno nuevo o una cola de mensajes que recibe fuentes de datos en tiempo real.
- 3 **Salidas:** representan los mecanismos de entrega de datos orientados al consumo. Un producto de datos puede exponer una serie de interfaces estandarizadas como SQL, API REST, búsqueda vectorial, mensaje y archivo. Los puertos de salida también pueden variar en función del tipo de servicio que proporciona el producto de datos. Es posible que una salida proporcione acceso a datos históricos, y que la otra proporcione acceso a la instantánea de datos más reciente.
- 4 **Lógica de recuperación de datos y transformación de dominios:** representa el código y la lógica del canal de datos para adquirir los datos (por ejemplo, mover los datos de una base de datos de origen a una plataforma de datos en la nube o leer un conjunto de tablas), y luego aplicar la lógica de transformación para dar forma a los datos desde el formato de origen al formato de base de datos de destino, que puede ser en forma de tablas, vistas o integraciones vectoriales.
- 5 **Metadatos y semántica empresarial:** contienen los metadatos técnicos (nombres de tabla, nombres de archivo, definiciones de columna) y los metadatos empresariales (dominio, nombre de producto, tipo de producto, propósito, propiedad, etiqueta, etc.) para permitir una mejor comprensión del producto de datos.
- 6 **Reglas de calidad de datos:** definen las diversas comprobaciones de calidad de datos que deben ejecutarse en el producto de datos. Cada regla incluye información de configuración relacionada con cuándo se debe ejecutar, el alcance de la regla, si se debe ejecutar en todos los datos o es incremental, etc. Las reglas de calidad de datos se ejecutarán según el programa prescrito para garantizar la precisión de los datos y generar confianza.
- 7 **Políticas de privacidad y control de acceso:** son las políticas que controlan el acceso a las bases de datos dentro del producto de datos. Por ejemplo, si las bases de datos contienen información de carácter personal, una política de privacidad puede ocultar esta información para todos los usuarios o autorizar el acceso solo a determinados grupos de usuarios.
- 8 **Contratos de datos y objetivos de nivel de servicio (SLO):** definen los valores previstos del objetivo de servicio que se espera que cumpla el producto de datos. Por ejemplo, es posible que un SLO exija que una métrica de precisión de calidad de datos sea mayor del 95 % o que los datos se actualicen cada cinco minutos.
- 9 **Funciones analíticas:** incluyen una lógica empresarial personalizada adicional para que la función se pueda reutilizar. Una función en un producto de datos de pedidos de ventas podría ser ver los pedidos retrasados, que incluirá la lógica de dominio complejo para determinar qué pedidos están retrasados. Las funciones analíticas pueden exponerse mediante un puerto de salida, como una API.
- 10 **Dependencias de infraestructura:** información sobre los requisitos de infraestructura para las bases de datos asociadas con el producto de datos. Por ejemplo, las bases de datos se almacenan en una plataforma de datos (como Snowflake o Databricks) y el código asociado (por ejemplo, la lógica de transformación) requiere un entorno de computación (por ejemplo, SQL) para ejecutarse.

Ciclo de vida del producto de datos

Los productos de datos no son creaciones puntuales. Como los productos materiales, exigen una gestión y supervisión continuas. Cada producto de datos es propiedad de un equipo de datos alineado con el dominio, que lo gestiona y es responsable del éxito en la entrega de valor, el aumento de su adopción y el mantenimiento de su ciclo de vida.



A continuación se incluye un desglose de las etapas clave del ciclo de vida del producto de datos:

- 1 Conceptualización:** el proceso comienza con la identificación de las necesidades empresariales específicas del dominio, lo que incluye comprender a los usuarios, sus casos de uso y cómo suelen consumir los datos.
- 2 Diseño:** el siguiente paso consiste en dar forma a la solución, lo que incluye identificar las bases de datos necesarias y sus orígenes, así como definir las características del producto de datos por funcionalidad, interactividad, enfoque de presentación, etc.
- 3 Creación:** después del diseño, hay que crear la infraestructura técnica y las funcionalidades del producto de datos. Incluye recopilar e integrar las bases de datos sin procesar a partir de sus orígenes, garantizar la calidad de los datos y transformar los datos.
- 4 Implementación:** se trata de la etapa de «puesta a disposición», que implica completar la documentación del producto, asignar la propiedad, integrarlo en sistemas empresariales o flujos de trabajo de consumo existentes para facilitar el acceso, y activar el producto de datos o bien dejarlo preparado para el consumo.
- 5 Comercialización:** después de la activación, es importante dar a conocer el producto y promover su uso. Implica formar a los consumidores y establecer canales de comunicación para recabar su opinión.
- 6 Uso:** una vez que se lance con éxito, los consumidores comenzarán a interactuar con el producto de datos para buscar, encontrar, comprender y obtener acceso. Esta etapa se centra en supervisar el uso y recabar la opinión de los usuarios. Normalmente, se recopilan datos sobre cómo los usuarios interactúan con el producto de datos, su rendimiento y si satisface las necesidades de la empresa centradas en el dominio.
- 7 Mantenimiento:** los productos de datos no son entidades estáticas; deben mejorarse para garantizar un valor continuado. Según las opiniones recabadas de los usuarios y las nuevas necesidades de la empresa, se propone la incorporación de nuevas dimensiones de datos, características y funcionalidades a la siguiente versión del producto de datos.

Al seguir estas etapas del ciclo de vida, las organizaciones pueden crear productos de datos que proporcionan información importante, impulsar la toma de decisiones basada en datos y lograr sus objetivos empresariales.

Qué NO es un producto de datos

El mundo de los datos está abarrotado de términos, y su uso no siempre es coherente. Un buen ejemplo: el término «producto de datos» se suele vincular con términos como «bases de datos» o «catálogo de datos», que no son lo mismo. En esta sección se destacan algunas diferencias fundamentales entre estos conceptos.



Aunque las bases de datos son un componente fundamental de muchos productos de datos, es crucial comprender que solo son un componente. Un producto de datos va mucho más allá de las bases de datos sin procesar.

BASES DE DATOS: EL MATERIAL SIN PROCESAR

Las bases de datos suelen ser una serie de puntos de datos organizados, estén o no estructurados. Pueden incluir información sin procesar como detalles de clientes, registros contables o lecturas de sensores. Considere las bases de datos como materias primas, por ejemplo, montones de madera, que son potencialmente valiosas, pero que primero hay que refinar. Una de las cosas más importantes que diferencia las bases de datos de un producto de datos es la ausencia de contexto empresarial específico del dominio.

PRODUCTOS DE DATOS: LOS PRODUCTOS ACABADOS

Los productos de datos emplean bases de datos seleccionadas y procesadas meticulosamente, que a menudo se combinan con otros artefactos comentados en la sección anterior, para proporcionar valor. Siguiendo con la metáfora de la madera anterior, los productos de datos son muebles perfectamente fabricados a partir de madera (es decir, las bases de datos).

A continuación se describe lo que diferencia los productos de datos de las bases de datos:

- **Orientación al contexto y al dominio:** los productos de datos están diseñados para abordar necesidades específicas de las empresas o los usuarios. Van más allá de los datos sin procesar para incluir información contextual como definiciones, linaje y comprobaciones de calidad.
- **Facilidad de uso y accesibilidad:** los productos de datos están diseñados para aumentar el valor y facilitar el consumo. Se seleccionan y empaquetan para que los puedan consumir con facilidad consumidores de datos específicos del dominio con conocimientos técnicos limitados.
- **Conocimientos prácticos:** los productos de datos no solo presentan datos, también ofrecen conocimientos operativos, analíticos, predictivos y prescriptivos. Estos conocimientos se pueden entregar mediante cuadros de mando de visualización, modelos analíticos prediseñados o enlaces de integración con otras herramientas mediante SQL o API.
- **Propiedad y rendición de cuentas:** los productos de datos tienen líneas de propiedad claras, lo que garantiza la rendición de cuentas en la organización sobre diversos aspectos de su desarrollo, su consumo y la opinión sobre ellos. La propiedad de los productos de datos puede depender del contexto empresarial; el propietario es el productor que genera el producto de datos, el consumidor dentro del equipo de dominio responsable de los casos de uso pertinentes o un puesto de reciente creación denominado «gestor de productos de datos».



Los productos de datos transforman las bases de datos en activos valiosos y consumibles que impulsan la toma de decisiones informadas y el crecimiento empresarial. La publicación Harvard Business Review calcula que «las empresas que tratan los datos como un producto pueden reducir el tiempo que tardan en implementarlos en nuevos casos de uso hasta en un 90 %, disminuir sus costes totales de propiedad (tecnología, desarrollo y mantenimiento) hasta en un 30 %, y reducir su carga de riesgo y gobernanza de datos».⁴

⁴ «A Better Way to Put Your Data to Work» (Una mejor forma de sacar partido a los datos), Harvard Business Review, julio-agosto de 2022.

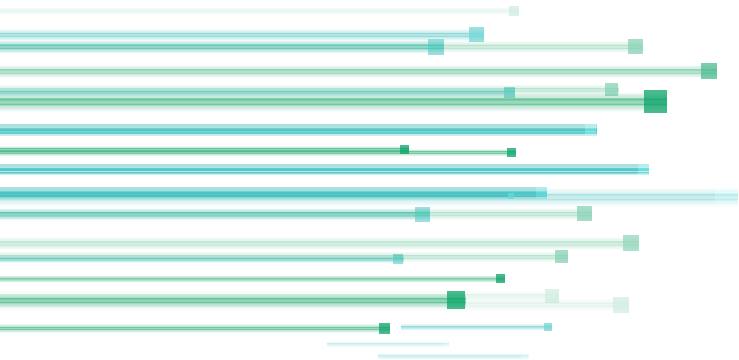
CATÁLOGO DE PRODUCTOS DE DATOS: EL CENTRO DE CONTROL DE LOS PRODUCTOS ACABADOS

Un catálogo de productos de datos actúa como un repositorio integral que aloja productos de datos y metadatos asociados (descripciones, etiquetas, versión, etiquetas de propiedad, etc.) para contribuir al reconocimiento y la asimilación de los productos de datos. Presta servicio tanto a los productores como a los consumidores de datos, ofreciendo soluciones personalizadas para satisfacer sus necesidades. Para los gestores de productos de datos, el catálogo proporciona herramientas para configurar, activar y gestionar el ciclo de vida completo de un producto de datos, desde la creación a la activación y el final de la vida útil. Para los consumidores, el catálogo actúa como un mercado, lo que permite una experiencia similar a la compra que impulsa el uso.

A continuación se describe lo que diferencia un catálogo de productos de datos de un catálogo de datos:

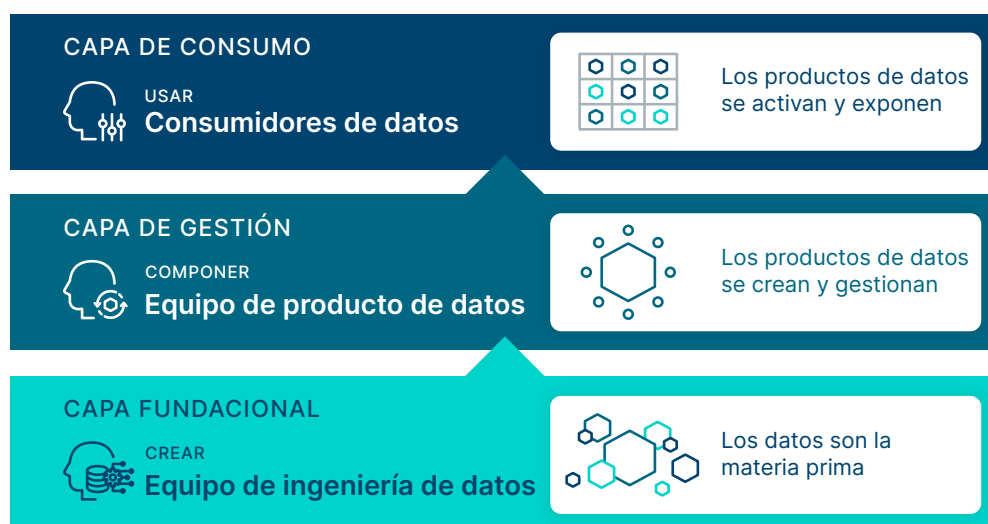
CARACTERÍSTICA	CATÁLOGO DE DATOS	CATÁLOGO DE PRODUCTOS DE DATOS
Misión	Reforzar la gobernanza de datos	Aumentar el valor de los datos
Ámbito	Centrado en la empresa	Centrado en el dominio
Modelo de gestión	Centralizado	Federado
Propósito	Reconocimiento, conformidad, transparencia	Facilidad de uso, confianza, capacidad de procesamiento
Contenido	Metadatos de diversos tipos, incluidos técnicos, operativos y empresariales, relacionados con cada activo de datos	Productos de datos de confianza, metadatos del ciclo de vida de los productos de datos, documentación de los productos de datos y propiedad
Entidades	Orden de cientos de miles	Orden de cientos
Audiencia	Principalmente gobernanza de datos	Productores y consumidores de datos

En el ámbito de la gestión de datos, el reto no solo consiste en crear plataformas de datos, sino en garantizar su adopción efectiva. Con demasiada frecuencia, las organizaciones invierten recursos considerables, a menudo millones de dólares, en crear infraestructuras de datos, pero se topan con un problema grave: **barreras de intercambio de datos entre productores y consumidores**. ¿Por qué? Porque los resultados generados por muchas plataformas tecnológicas carecen de fiabilidad y no se corresponden con los objetivos empresariales específicos de los consumidores. **La conclusión es que su empresa necesita tanto un catálogo de datos como un catálogo de productos de datos**. Pueden considerarse diferentes pasillos de un supermercado, uno con materias primas, como harina, mantequilla y azúcar, y el otro con productos acabados, como pastas y pasteles.



Catálogo de productos de datos: capas y actores

El catálogo de productos de datos es el centro de control principal para organizar los productos de datos en una organización. Está estructurado en tres capas, cada una de ellas adaptada a un conjunto diferente de partes interesadas. Ya se trate de ingenieros de datos, administradores, analistas o puestos de nueva creación como gestores de productos de datos, cada grupo interactúa de manera única con estas capas en función de sus objetivos y necesidades. Este enfoque garantiza la separación de intereses en función de las necesidades del usuario, con el nivel adecuado de abstracción para facilitar la adopción y el consumo.



LA CAPA BASE

La base del catálogo de productos de datos es una capa fundamental que establece la infraestructura técnica y proporciona las capacidades que sustentan las capas posteriores.

Las capacidades clave de esta capa son:

- **Entrega de datos:** traslado y procesamiento de datos a partir de diversos orígenes, como archivos, data warehouses en la nube, data lakes y streams de datos en tiempo real.
- **Gestión de metadatos:** recopilación y organización de información valiosa sobre los datos, como el linaje de datos, los diccionarios de datos, el esquema, etc.
- **Control de seguridad y acceso:** establecimiento de controles de seguridad de datos para proteger los datos y garantizar que solo los usuarios autorizados puedan acceder a ellos.
- **Seguimiento de la calidad y el linaje de los datos:** supervisión de la calidad de los datos para garantizar su precisión y seguimiento del linaje de los datos para asegurar su fiabilidad.

La capa base presta servicio principalmente al equipo de la plataforma de datos, incluidos puestos técnicos prácticos como ingenieros de datos, administradores y arquitectos. El objetivo es desarrollar las capacidades de esta capa en formato de autoservicio para que los usuarios sin habilidades técnicas puedan invocarlas con facilidad. Las capacidades necesarias en la capa base se pueden encontrar en diversas herramientas de gestión de metadatos, ingesta de datos, transformación de datos, calidad de datos y seguridad de datos.

LA CAPA DE GESTIÓN

La capa de gestión de productos de datos, un nuevo concepto, ofrece las capacidades necesarias para satisfacer las necesidades del puesto de nueva creación de gestor (o equipo) de productos de datos. Esta capa ofrece capacidades para desarrollar, gestionar y mantener productos de datos de confianza orientados al dominio. Se sirve de los servicios del plano básico para inventariar, trasladar, transformar y proteger los activos de datos. Actúa como punto de conexión entre los metadatos sin procesar que residen en la capa base y los conocimientos prácticos que los consumidores quieren extraer de sus datos.

Las capacidades clave de esta capa son:

- **Composición del producto de datos:** capacidades para diseñar un producto de datos, definir las entradas y salidas, y configurar sus diversos artefactos.
- **Implementación del producto de datos:** gestión de versiones y el flujo de trabajo para activar un producto de datos cuando está listo para el consumo por parte de los usuarios finales.
- **Mantenimiento del producto de datos:** seguimiento del uso, recabado de opiniones, cambios y actualizaciones, y registro de todas las etapas del ciclo de vida del producto de datos descritas en las secciones anteriores.

La capa de gestión de productos de datos presta servicio principalmente al equipo de gestión de productos, especialmente a los gestores de productos de datos, que dirigen el ciclo de vida de los productos de datos para cumplir los objetivos empresariales.

LA CAPA DE CONSUMO (ES DECIR, EL MERCADO)

La capa de mercado se sitúa en la parte superior del catálogo de productos de datos y representa la parte del sistema orientada al consumidor. Habilita el autoservicio en la exploración y el consumo de datos, con lo que promueve una cultura basada en los datos dentro de la organización.

Entre las capacidades clave de esta capa se incluyen:

- **Búsqueda y reconocimiento:** las interfaces intuitivas permiten a los usuarios explorar y reconocer con facilidad los productos de datos que mejor se adaptan a sus necesidades.
- **Aprendizaje y confianza:** los consumidores de datos pueden entender el producto de datos consultando los metadatos y los datos de muestra, aprender a utilizarlo consultando el código de muestra y la documentación de los datos, y afianzar su confianza en el producto revisando la calidad y el linaje de los datos.
- **Acceso y permisos:** esta capa controla quién puede acceder a productos de datos específicos, para garantizar la seguridad de los datos y el cumplimiento de las políticas de gobernanza de datos.
- **Interfaces de consumo de datos:** las API, las herramientas de visualización o incluso los puntos de acceso directo a las bases de datos, proporcionan a los usuarios los medios necesarios para consumir los productos de datos. En función del producto, puede haber disponibles diferentes interfaces adecuadas para distintas necesidades y habilidades de los usuarios.

La capa de consumo de datos presta servicio principalmente a los analistas de negocio, expertos en datos, usuarios de línea de negocio y otros responsables clave de la toma de decisiones empresariales.

VENTAJAS DE LOS PRODUCTOS DE DATOS PARA LOS DISTINTOS ACTORES DEL ECOSISTEMA DE DATOS

Los datos son el mayor activo de una organización, pero su verdadero potencial solo se materializa si se puede acceder a ellos con facilidad, se comprenden correctamente y se utilizan de manera efectiva. Los productos de datos sirven de vínculo, transformando los datos sin procesar en activos consumibles y seleccionados que capacitan a las partes interesadas de toda la organización.

Vamos a analizar cómo los productos de datos afectan a los actores clave del ecosistema de datos:



Para los **equipos de entrega de datos**, los productos de datos son un lenguaje común (un acuerdo bien definido) que aporta estructura y aclara la comunicación entre los productores y los consumidores de datos. Esto permite a los productores desarrollar productos de datos reutilizables de manera iterativa, lo que reduce el coste total de la propiedad. Además, el equipo de entrega de datos no necesita desarrollar grandes canales de datos centralizados y complejos; pueden centrarse en desarrollar canales más pequeños e independientes dedicados a la implementación de cada producto.



Para los **equipos de gobernanza de datos**, los productos de datos se convierten en un punto de auditoría central y, dado que estos productos son de confianza y se rigen por la gobernanza, es más sencillo minimizar el riesgo. El equipo de gobernanza define de manera centralizada las políticas y estándares; los distintos equipos de productos de datos aplican las políticas de manera federada.



Para los **gestores de productos de datos**, los productos de datos alinean mejor a los productores y consumidores de datos, con lo que impulsan una mayor colaboración y rendición de cuentas en todo el ecosistema de datos. Esto refuerza la propiedad y rendición de cuentas de los datos.



Para los **consumidores de datos**, los productos de datos son activos de datos seleccionados y centrados en el dominio, que proporcionan exactamente lo que se necesita para maximizar el valor empresarial. Gracias a ello se reducen las revisiones del trabajo, lo que da lugar a resultados más rápidos.

Unidas, estas partes interesadas pueden emplear un catálogo de productos de datos para aprovechar todo el valor de sus activos de datos de manera eficiente, impulsar la toma de decisiones basada en datos y lograr sus objetivos empresariales.

“ En el sector hipotecario, los productos de datos son mucho más que simples números y algoritmos: los consideramos agentes transformadores. Al definir productos de datos de autoservicio seleccionados, capacitamos a nuestros responsables de préstamos para que ofrezcan la solución hipotecaria adecuada al prestatario adecuado, y permitimos a los equipos de finanzas evaluar los costes y controlar el riesgo mejor. El enfoque de aprovechar los productos de datos en nuestra empresa está redefiniendo realmente el futuro de nuestras hipotecas. ”

Julia Fryk

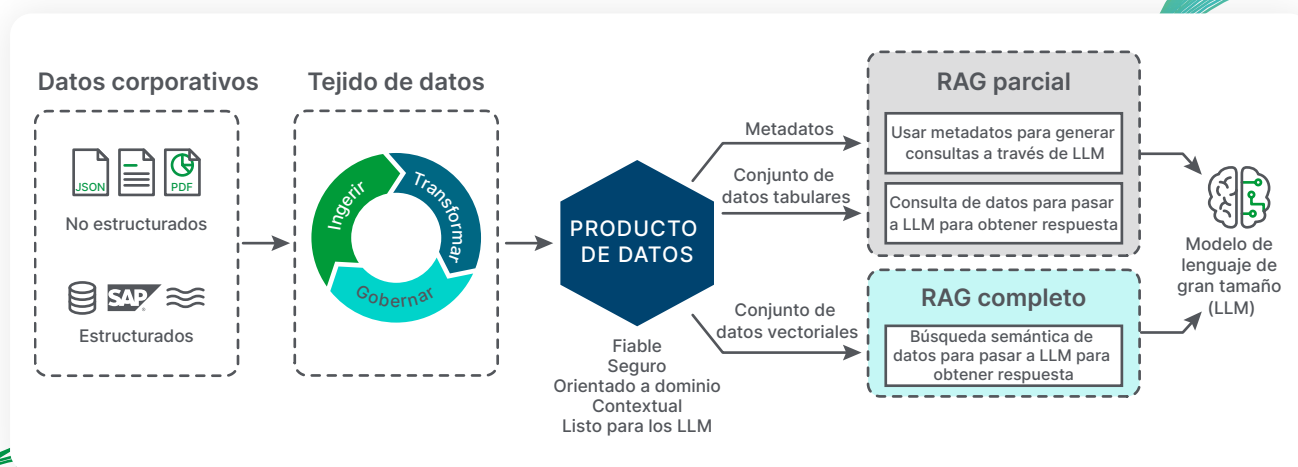
Arquitecta principal de datos, Waterstone Mortgage

Entrega de datos contextuales a los LLM con productos de datos

La IA generativa está creando enormes oportunidades para mejorar la productividad en todos los aspectos de la empresa. En el núcleo de la IA generativa están los modelos fundacionales (por ejemplo, los LLM como GPT-3, Anthropic, Llama y BERT) que se han entrenado previamente con un corpus masivo de datos de texto (páginas web, documentos, etc.). No obstante, los LLM fundacionales no comprenden los datos internos específicos de la empresa. Por eso primero tienen que cimentarse con datos empresariales.

Una de las técnicas más efectivas para proporcionar datos empresariales a los LLM es mediante un mecanismo denominado RAG o generación mejorada por recuperación. Tal y como su nombre indica, los datos contextuales relacionados con una indicación (pregunta) del usuario se **recuperan** y, a continuación, la indicación del usuario se **mejora** con estos datos y se traslada al LLM para generar una respuesta. Existen dos métodos de RAG: RAG parcial y RAG completo.

- En el RAG parcial (para datos estructurados), la aplicación de RAG traslada los metadatos al LLM para generar una consulta SQL. A continuación, la consulta se ejecuta en las bases de datos en formato de tabla y los resultados se trasladan al LLM para generar una respuesta. Este método se utiliza cuando se requiere una respuesta explícita de los datos.
- En el RAG completo (para datos estructurados o no estructurados), los datos primero se vectorizan. A continuación, la aplicación de RAG realiza una búsqueda semántica en estos datos vectorizados para recuperar los datos. Dichos datos se trasladan al LLM para generar una respuesta.



En ambos métodos, es importante contar con datos de gran calidad, limpios y contextuales disponibles para el RAG. **Los productos de datos proporcionan datos contextuales y orientados al dominio para distintas formas de consumo.** Tal y como se ha mencionado anteriormente, los productos de datos contienen metadatos enriquecidos (empresariales, técnicos y orientados al dominio) y bases de datos muy seleccionadas (archivos, tablas o datos vectorizados) para el consumo. Por tanto, los productos de datos se convierten en el vehículo perfecto para habilitar las aplicaciones basadas en RAG.

Para desarrollar una aplicación de RAG, primero hay que buscar en el catálogo de productos de datos o el mercado el producto de datos adecuado que se debe utilizar en función de la pregunta que se plantea. Por ejemplo, una pregunta relacionada con las ventas se puede responder mejor con un producto de datos «Oportunidad» procedente del CRM, a partir de los metadatos detallados asociados al producto de datos. O una pregunta relacionada con un pedido se puede responder mejor con un producto de datos «Pedido de ventas» procedente de SAP, a partir de los metadatos detallados asociados a ese producto de datos. Una vez que se ha identificado el producto de datos, la aplicación de RAG puede consultar directamente las bases de datos del producto de datos, ya sea en formato de tabla o como datos vectorizados, en función del caso de uso.

Conclusión

Mientras los datos sigan siendo la base de las empresas modernas, los productos de datos serán cada vez más importantes para **acceder al valor integral que ofrecen los datos** y acelerar los resultados empresariales, **mejorando al mismo tiempo la gobernanza de los datos y la confianza en ellos**. En Qlik, nos comprometemos a ayudarle a desarrollar una cimentación de datos para ofrecer capacidades de autoservicio. Para ello, capacitaremos a sus equipos de datos para que creen y gestionen productos de datos, y daremos a los consumidores de datos la capacidad de buscar, reconocer y utilizar rápidamente los productos de datos para impulsar resultados empresariales de éxito.

Un último apunte: aunque los productos de datos son herramientas potentes, no son una solución mágica. El verdadero poder de los productos de datos radica en la capacidad de identificar las necesidades empresariales, contar con una estrategia de datos federada bien definida, garantizar una gobernanza efectiva y colaborar entre los equipos.

“ Cada vez más, las organizaciones desean crear productos de datos centrados en el dominio que trasciendan la integración y calidad de los datos. El enfoque de Qlik en sus ofertas de calidad y gobernanza de datos en cuanto al concepto de productos de datos destaca por una gran madurez que identifica de inmediato las necesidades actuales de los clientes. ”

Stewart Bond

Vicepresidente, software de inteligencia e integración de datos, IDC



Acerca de Qlik

Qlik convierte entornos de datos complejos en información útil, generando resultados empresariales estratégicos. Con más de 40.000 clientes en todo el mundo, nuestra cartera de productos integra capacidades avanzadas de IA y ML empresarial, mejorando de manera constante la calidad de los datos. Destacamos en integración y gobernanza de datos, ofreciendo soluciones integrales que operan con fuentes de datos dispares. El análisis intuitivo en tiempo real de Qlik revela patrones ocultos, permitiendo a los equipos abordar desafíos complejos y aprovechar nuevas oportunidades. Nuestras herramientas de IA/ML, prácticas y escalables, permiten tomar mejores decisiones con mayor rapidez. Como socio estratégico, nuestra tecnología y experiencia hacen que nuestros clientes sean más competitivos, independientemente de la plataforma que utilicen.

qlik.com

**¿Desea acceder
al valor integral de
sus datos?**

[Más información](#)