PRESENTACIÓN DEL PRODUCTO · OCTUBRE DE 2026

BeryxTech

Calidad de datos, RGPD y coste de los datos en una sola capa: leída en la fuente, de solo lectura y con el cálculo de cada número a la vista.

Para el equipo de datos, para el DPD y para quien paga la factura de la nube.

CADA AFIRMACIÓN LLEVA UNA ETIQUETA

● EXISTE HOY
en producción y cubierto por pruebas

◐ PARCIAL
existe, con los límites indicados en la diapositiva

◐ POR ACTIVAR
hecho en el código, falta activarlo en producción

○ VISIÓN
todavía no existe

Datos verificados el 12/10/2026 frente al producto en producción y a las fuentes citadas.

Las decisiones y la IA ya funcionan sobre los datos. La confianza en ellos no ha seguido el mismo ritmo.

67%

no confían plenamente en los datos que usan para decidir (55% en 2023)

Precisely y Drexel LeBow, sept. 2024 · 565+ profesionales · estudio de proveedor

64%

señalan la calidad de los datos como su mayor reto de integridad (50% en 2023)

Precisely y Drexel LeBow, sept. 2024 · estudio de proveedor

71%

temen que resultados erróneos o alucinados lleguen a quien decide

dbt Labs, State of Analytics Engineering, abr. 2026 · 363 profesionales · estudio de proveedor

60%

de los proyectos de IA sin datos preparados para la IA se abandonarán hasta finales de 2026, según la previsión de Gartner

Gartner, comunicado del 26/02/2025

Cuando la IA responde sobre datos sin verificar, el error llega antes a quien decide.

Tres costes que no aparecen en ningún presupuesto

01 INGENIERÍA DE DATOS

El negocio descubre el fallo antes que el equipo de datos

Los datos parados, duplicados o inválidos no rompen ningún job. Aparecen en el cierre del mes, en un informe para la dirección o en una factura que no sale.

02 RGPD E IA

La exactitud de los datos personales es una obligación legal

El RGPD exige datos exactos y actualizados, responder a los interesados en un mes y notificar las violaciones de seguridad en 72 horas. Infringir los principios conlleva multas de hasta 20 M€ o el 4% del volumen de negocio mundial. El Reglamento de IA añade requisitos sobre los datos de entrenamiento a partir de diciembre de 2027.

RGPD, arts. 5, 12, 33 y 83; Reglamento de IA, arts. 10 y 113

03 FINOPS DE DATOS

La factura de los datos crece sin dueño

Las tablas que nadie lee se siguen cargando. Una consulta mal escrita multiplica el coste del día y solo se ve en la factura del mes. Se paga cómputo para procesar datos que no superan las comprobaciones.

Los tres costes tienen la misma raíz: nadie mide los datos de forma continua, con dueño y con evidencia.

Los datos fallan en silencio: los jobs siguen en verde y las cifras salen mal

ANATOMÍA DE UN INCIDENTE DE DATOS

EJEMPLO ILUSTRATIVO · DATOS SINTÉTICOS

VIERNES, 18 H

El conector del marketplace deja de sincronizar. Ningún job falla: simplemente dejan de llegar filas.

FIN DE SEMANA

Los pedidos del marketplace no entran en el data warehouse. El panel de ventas muestra una caída que nunca existió.

LUNES, 10 H

El cierre semanal no cuadra. Empieza la investigación: ¿qué tabla, desde cuándo y quién más usa estos datos?

MARTES

Encontrado el origen, falta saber qué informes y modelos ya han consumido los datos erróneos.

SEIS FORMAS DE FALLAR SIN ERROR

Frescura
la tabla dejó de recibir datos

Volumen
llegó la mitad de las filas

Esquema
una columna cambió de tipo o de nombre

Duplicados
un reintento grabó el mismo pedido dos veces

Validez
un NIF de 8 dígitos, un IBAN con el dígito de control erróneo

Integridad
pedidos sin cliente tras un borrado sin cascada

La historia y los ejemplos vienen de los packs de demostración de BeryX, con datos sintéticos. Cada fallo corresponde a un monitor o a una regla del producto.

El coste de no saber se calcula con tus cifras

PILAR	QUÉ CONTAR	CÓMO ESTIMARLO	QUÉ MIDE BERYX
Ingeniería de datos	Incidentes de datos al mes, horas de investigación y días hasta la detección	Incidentes × horas × coste por hora, más días hasta la detección × valor en riesgo por día (facturas por emitir, por ejemplo)	Histórico de incidencias e incidentes con fechas, e impacto estimado en euros por incidencia, etiquetado como heurística
RGPD	Solicitudes de interesados al mes y datasets con datos personales sin dueño o sin reglas	Solicitudes × horas por solicitud; exposición máxima de 20 M€ o el 4% del volumen de negocio mundial	Mapa de datos personales con la cobertura de reglas de cada columna
FinOps de datos	Tablas sin lecturas en 90 días, cómputo gastado en datos que no superan las comprobaciones y picos diarios	Coste mensual de esas tablas, más el cómputo atribuido a los datasets con incidencias abiertas	Coste leído de las vistas de sistema de la plataforma, por día (por dataset en BigQuery, por cuenta en Snowflake y Databricks), picos de coste y uso real por tabla

Fuentes: RGPD, art. 83.5

En una primera conversación rellenamos este cuadro contigo, con los datos de tus fuentes.

Lo que la ley pide a los datos, y que solo se demuestra con mediciones

OBLIGACIÓN	QUÉ EXIGE	PLAZO O LÍMITE	QUÉ PIDE A LOS DATOS
Exactitud · RGPD, art. 5.1.d)	Datos personales exactos y actualizados; los inexactos, suprimidos o rectificados sin dilación	Multa de hasta 20 M€ o el 4% del volumen de negocio mundial (art. 83.5)	Saber qué datos personales están mal, dónde y desde cuándo
Derechos de los interesados · art. 12.3	Responder a solicitudes de acceso, rectificación o supresión	Un mes, prorrogable dos más	Saber en qué sistemas están los datos de cada persona
Violación de la seguridad · art. 33.1	Notificar a la autoridad de control (en Portugal, la CNPD)	72 horas, siempre que sea posible	Saber qué datos y qué interesados se vieron afectados
Evaluación de impacto · art. 35	Evaluar antes de un tratamiento de alto riesgo	Antes de empezar el tratamiento	Inventario de categorías de datos y de flujos
Datos para IA de alto riesgo · Reglamento de IA, art. 10.3	Datos de entrenamiento, validación y prueba pertinentes, representativos y, en la mayor medida posible, sin errores y completos	A partir del 2 de diciembre de 2027 (anexo III)	Medir errores y completitud de los conjuntos de datos

Resumen de contexto; no es asesoramiento jurídico.

Fuentes: RGPD, Reglamento (UE) 2016/679 · Reglamento de IA, arts. 10 y 113 (AI Act Service Desk)

La autoridad portuguesa de protección de datos ya ha impuesto multas millonarias por fallos en la gestión de datos personales

4,3 M€

Instituto Nacional de Estadística de Portugal (INE)

DICIEMBRE DE 2022

Censo de 2021. Cinco infracciones: datos de salud y de religión tratados sin una base de licitud clara, información insuficiente a los encuestados, un encargado del tratamiento en EE. UU. elegido sin la diligencia debida, transferencias internacionales sin medidas complementarias y ninguna evaluación de impacto.

ESTADO El INE anunció que impugnaría la decisión en los tribunales.

CNPD, Deliberación/2022/1072 · Observador, 13/12/2022

1,25 M€

Ayuntamiento de Lisboa

ENERO DE 2022

Datos de promotores de manifestaciones enviados a embajadas y al Ministerio de Asuntos Exteriores: 225 infracciones.

ESTADO Los tribunales confirmaron la infracción. Con parte de las infracciones prescritas, el importe bajó a 738 000 € en el Tribunal Central Administrativo Sur (agosto de 2025), en una decisión todavía recurrible.

Renascença, 14/01/2022 · Observador, 06/08/2025

400 000 €

Centro Hospitalario Barreiro-Montijo

2018

La primera multa del RGPD en Portugal: 985 cuentas con perfil de médico para 296 médicos y acceso indebido a historias clínicas.

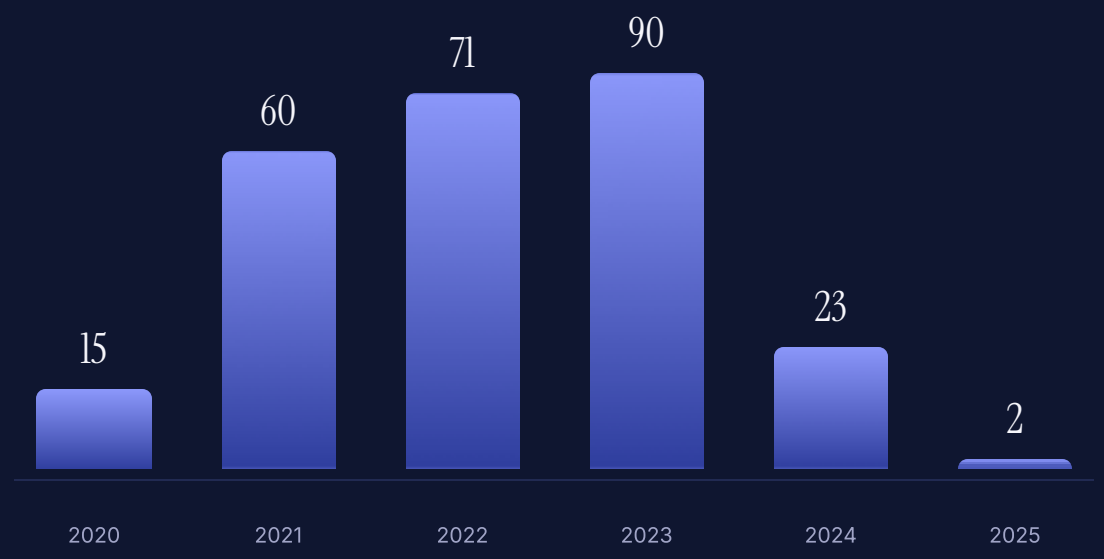
ESTADO Pasó a ser una multa única de 380 000 €. En julio de 2020, la CNPD aceptó dispensarla, alegando la pandemia y las cuentas del hospital.

Garrigues, 2018 · Renascença, 27/05/2021

No decimos que BeryX habría evitado estas multas. Los tres casos muestran lo que investiga la CNPD: dónde están los datos sensibles, quién accede a ellos y adónde van.

Las multas suben y bajan con los medios de la CNPD. Las violaciones notificadas siguen subiendo.

MULTAS IMPUESTAS POR LA CNPD, POR AÑO



472

expedientes por violación de datos personales en 2025, un 42% más que en 2024 (357 en entidades privadas, 115 en públicas)

CNPD, Informe de Actividades 2025, pp. 15 y 34

88

procedimientos sancionadores abiertos en 2025 que, según la CNPD, podrían haber dado lugar a 90 multas en lugar de 2

CNPD, Informe de Actividades 2025, p. 42

2037

investigaciones abiertas en 2025

CNPD, Informe de Actividades 2025, p. 42

> La propia CNPD atribuye la caída de 2025 a la falta de personal para tramitar expedientes y, en septiembre de 2025, propuso al Parlamento un régimen sancionador más rápido. Las obligaciones siguen ahí: notificar en 72 horas, responder en un mes, demostrar la exactitud.

Fuentes: CNPD, Informe de Actividades 2025, aprobado el 24/03/2026 (fig. 10, pp. 12, 42 y 43)

La factura de los datos crece sin dueño y sin relación con la calidad

84%

dicen que gestionar el gasto en la nube es hoy el mayor reto de la nube para las organizaciones

Flexera, State of the Cloud 2025 · 750+ profesionales · estudio de proveedor

- × Las tablas que nadie lee se siguen cargando, almacenando y monitorizando.
- × Una consulta mal escrita multiplica el coste del día, y solo se ve en la factura del mes.
- × Se paga cómputo para procesar datos que no superan las comprobaciones.
- × El coste de la IA crece sin tope y sin reparto por equipo.

La factura llega a finanzas y los problemas de calidad se quedan en el equipo de datos. BeryX pone los dos uno al lado del otro, por dataset siempre que la plataforma lo permite.

¿Dónde vive hoy la verdad sobre tus datos?

Hoy

- × En tests y scripts repartidos por equipos y repositorios
- × En la memoria de quien conoce las tablas
- × En un registro de actividades de tratamiento en una hoja de cálculo, revisado una vez al año
- × En la factura de la nube, a final de mes
- × En paneles que nadie relaciona con el origen de los datos

Con BeryX

- ✓ En reglas y contratos versionados, que se ejecutan en cada escaneo ● EXISTE HOY
- ✓ En un catálogo con dueño, dominio, sensibilidad y uso por dataset ● EXISTE HOY
- ✓ En un mapa de datos personales actualizado en cada escaneo ● EXISTE HOY
- ✓ En el coste diario leído de las vistas de sistema de BigQuery, Snowflake y Databricks ● EXISTE HOY
- ✓ En incidentes con linaje, dueño y plazo ● EXISTE HOY

> Un registro de confianza por dataset, compartido por ingeniería, el DPD y finanzas.

Por qué ahora

2025

Las violaciones de datos notificadas a la CNPD suben un 42%, hasta 472 expedientes.

CNPD, Informe de Actividades 2025

SEPT. 2025

La CNPD propone al Parlamento portugués un régimen sancionador más rápido, para evitar prescripciones.

CNPD, Informe de Actividades 2025, p. 12

2026

Previsión de Gartner: las organizaciones abandonan el 60% de los proyectos de IA sin datos preparados para la IA.

Gartner, 26/02/2025

2 AGO. 2026

El Reglamento de IA pasa a aplicarse con carácter general (art. 113).

AI Act Service Desk, art. 113

2 DIC. 2027

Obligaciones de alto riesgo del anexo III, incluida la calidad de los datos de entrenamiento, validación y prueba (art. 10).

AI Act Service Desk, arts. 10 y 113

2 AGO. 2028

Obligaciones de alto riesgo del anexo I, los productos regulados.

AI Act Service Desk, art. 113

> Poner los datos en orden lleva meses: perfilado, reglas, dueños y un histórico de mediciones que muestre la evolución.

Una capa de confianza sobre los datos que ya tienes, sin copiarlos

TUS FUENTES SE QUEDAN DONDE ESTÁN

- Bases de datos
- Data warehouses
- Ficheros y almacenamiento de objetos
- Aplicaciones de negocio: SAP, Salesforce, Dynamics 365
- Documentos
- Streaming

→

BERYX · SOLO LECTURA

Descubrir **Validar** **Monitorizar** **Explicar** **Demostrar**

- ✓ Agregados calculados en la fuente
- ✓ Muestras limitadas y con retención
- ✓ Propone correcciones, nunca escribe en las fuentes
- ✓ Cada número dice cómo se calculó

→

Ingeniería de datos
incidentes con linaje, SLO, reglas como código

DPD y jurídico
mapa de datos personales, sensibilidad, evidencia

FinOps y dirección
coste y calidad juntos, uso real, informes por perfil

DONDE EL EQUIPO YA TRABAJA: Slack · Teams · Jira · Power BI · Email · Webhooks · API · MCP

30 tipos de fuente en el registro de conectores; 1 todavía por instalar en el SaaS, Microsoft Fabric (ver Ecosistema).

Del dato conectado al dato demostrado, y de vuelta al principio

01
Conectar
credenciales de solo lectura

→

02
Descubrir
perfil, datos personales, uso

→

03
Reglas
propuestas por el perfil, aceptadas por personas

→

04
Monitorizar
frescura, volumen, esquema, anomalías, coste

→

05
Incidente
linaje, impacto, dueño, plazo

→

06
Corregir
SQL o pull request, ensayado antes

→

07
Demostrar
informes, evidencia, auditoría

↺ Lo que se aprende vuelve a las reglas: umbrales, falsos positivos marcados y reglas sugeridas a partir de incidencias resueltas.

Cada paso queda registrado: quién decidió, cuándo y con qué evidencia.

Cada dataset lleva consigo su registro de confianza

Dueño y dominio
quién responde y a qué área pertenece
EXISTE HOY

Sensibilidad por columna
pública, interna, confidencial o datos personales, con autor y fecha
EXISTE HOY

Reglas y contrato
reglas activas y contrato ODCS versionado
EXISTE HOY

Puntuación y SLO
cinco dimensiones, tendencia y objetivo de servicio
EXISTE HOY

Uso real
consultas al día: caliente, activa, fría o sin uso
EXISTE HOY

Coste
cómputo atribuido, cuando la plataforma lo reporta
EXISTE HOY

Incidentes y linaje
qué falló, aguas arriba y aguas abajo
EXISTE HOY

Auditoría
cada cambio, grabado en la misma transacción
EXISTE HOY

El ingeniero, el DPD y el director financiero leen el mismo registro. Nadie vuelve a explicar el contexto.

Conectar

ENTRADA

Una credencial de solo lectura de una base de datos, un warehouse, ficheros o una aplicación.

LO QUE HACE BERYX

Prueba la conexión, descubre esquemas y tablas e importa los metadatos. La credencial se cifra antes de guardarse y la dirección se valida contra SSRF antes de la primera petición.

● EXISTE HOY

QUIÉN DECIDE

El administrador del workspace elige la fuente, el alcance y la frecuencia de los escaneos.

GOBIERNO Y EVIDENCIA

Solo lectura impuesta en el código: el cliente HTTP rechaza las escrituras antes de salir y las sesiones de Postgres se abren en modo de solo lectura. En las pruebas, servidores reales de SQL Server, Oracle y MongoDB rechazaron escrituras.

● EXISTE HOY

SALIDA

Catálogo de datasets con esquema, programación y credencial cifrada.

● EXISTE HOY

SE CONECTA CON

PostgreSQL · SQL Server · Oracle · BigQuery · Snowflake · Databricks · S3 · SAP · Salesforce · Dynamics 365

PASA AL SIGUIENTE PASO

La lista de datasets y columnas pasa al perfilado.

VALOR

Sin credenciales de escritura, sin software instalado en las fuentes y sin copiar tablas.

● EXISTE HOY

Descubrir

ENTRADA

Los datasets conectados.

LO QUE HACE BERYX

Perfil por columna con agregados calculados en la fuente: nulos, distintos, patrones y distribuciones. Reconoce NIF, IBAN, email y código postal, construye el mapa de datos personales, clasifica la sensibilidad en cuatro niveles y estima el riesgo de reidentificación (k-anonimato). En las plataformas con telemetría, mide el uso real de cada tabla.

● EXISTE HOY

QUIÉN DECIDE

El DPD confirma o corrige la sensibilidad de cada columna. El escaneo nunca sobrescribe una decisión manual.

GOBIERNO Y EVIDENCIA

Cada cifra indica si es exacta, estimada o por muestreo, con un intervalo de confianza del 95%. Cada cambio de clasificación queda registrado con autor y fecha.

● EXISTE HOY

Métodos de BeryX

SALIDA

Perfil, puntuación por dimensión, mapa de datos personales, sensibilidad por columna y uso.

● EXISTE HOY

SE CONECTA CON

Tipos semánticos · Mapa de PII · k-anonimato · BigQuery · Snowflake

PASA AL SIGUIENTE PASO

El perfil alimenta las reglas sugeridas.

VALOR

El DPD ve dónde están los datos personales a partir de las propias fuentes, sin depender de un inventario manual.

● EXISTE HOY

Reglas y contratos

ENTRADA

El perfil, las sugerencias y los tests que el equipo ya tiene.

LO QUE HACE BERYX

Dieciséis tipos de regla: formatos con dígito de control (NIF, IBAN, tarjetas), unicidad, rangos, listas de referencia, integridad referencial, frescura, volumen y SQL propio de solo lectura. Exactitud frente a fuentes externas, opcional: VIES, GeoAPI.pt, planes de numeración y registros MX. Importa tests de dbt, Great Expectations, SodaCL y contratos ODCS.

● EXISTE HOY

QUIÉN DECIDE

El perfil y la IA proponen; una persona acepta cada regla.

GOBIERNO Y EVIDENCIA

El SQL del cliente pasa por una guarda: un único SELECT, en una transacción que siempre se revierte. La configuración también vive como código, en un datavault.yaml versionado.

● EXISTE HOY

SALIDA

Reglas activas por dataset y contrato versionado; exportación a dbt, con SodaCL y Great Expectations por activar.

◐ PARCIAL

SE CONECTA CON

dbt · Great Expectations · SodaCL · ODCS · VIES · GeoAPI.pt

PASA AL SIGUIENTE PASO

Las reglas aceptadas se ejecutan en cada escaneo y alimentan los monitores.

VALOR

Las reglas dejan de vivir en los scripts de una sola persona.

● EXISTE HOY

Monitorizar

ENTRADA

Reglas, histórico de perfiles y telemetría de coste y de uso.

LO QUE HACE BERYX

Monitores automáticos de frescura, volumen y esquema, hasta 500 por conexión en cada escaneo. Anomalías con z-score modificado y control de falsos positivos (Benjamini-Hochberg), deriva de distribución, SLO por dataset y picos de coste diario, con las consultas más caras del día en BigQuery.

● EXISTE HOY

QUIÉN DECIDE

El dueño del dataset define el SLO y la severidad. Un falso positivo marcado como «ignorado» silencia la serie.

GOBIERNO Y EVIDENCIA

Métodos publicados y probados contra el código. Ninguna cifra de calidad, de anomalía o de coste sale de un LLM.

● EXISTE HOY

SALIDA

Alertas en la app, por email, en Slack, en Teams y por webhook, con escalado de guardia.

● EXISTE HOY

SE CONECTA CON

Slack · Teams · Email · Webhooks firmados

PASA AL SIGUIENTE PASO

Una alerta abre o actualiza una incidencia, que puede unirse a un incidente.

VALOR

El fallo llega primero al equipo de datos, y no al cierre del mes.

● EXISTE HOY

Incidentes y causa raíz

ENTRADA

Una alerta o una incidencia abierta.

LO QUE HACE BERYX

Agrupa incidencias en incidentes a través del linaje (claves foráneas, SQL de las vistas, manifest de dbt, OpenLineage) y muestra el impacto aguas abajo. La IA propone la causa probable a partir de metadatos y de muestras limitadas. Plazos de respuesta por severidad, con escalado.

● EXISTE HOY

QUIÉN DECIDE

El responsable del dominio acepta, asigna y cierra el incidente.

GOBIERNO Y EVIDENCIA

Registro de auditoría grabado en la misma transacción que el cambio y exportable al SIEM: Splunk, Sentinel o Google SecOps.

● EXISTE HOY

SALIDA

Un incidente con linaje, impacto, dueño, plazo e historial, y un ticket en Jira, Linear, ServiceNow, Azure DevOps, GitHub o GitLab.

● EXISTE HOY

SE CONECTA CON

Jira · Linear · ServiceNow · Azure DevOps · GitHub · GitLab · OpenLineage

PASA AL SIGUIENTE PASO

La causa probable y el impacto pasan a la propuesta de corrección.

VALOR

Saber qué informes y modelos se vieron afectados antes de que alguien pregunte.

● EXISTE HOY

Corregir

ENTRADA

Un incidente con causa probable.

LO QUE HACE BERYX

Propone la corrección en SQL, en el dialecto de la fuente, o en PySpark, y la ensaya sobre una muestra antes de aceptarla. A partir del plan Business, la corrección aprobada abre un pull request en el GitHub o el GitLab del cliente. La normalización y la deduplicación también salen como propuestas.

● EXISTE HOY

QUIÉN DECIDE

Una persona aprueba y el cliente la aplica con su propio CI. BeryX nunca escribe en las fuentes.

GOBIERNO Y EVIDENCIA

Flujo propuesta, aceptada, aplicada y verificada: el siguiente escaneo confirma si la incidencia ha desaparecido.

● EXISTE HOY

SALIDA

Pull request o SQL aprobado, vinculado al dataset y al ticket.

● EXISTE HOY

SE CONECTA CON

GitHub · GitLab · SQL · PySpark · dbt

PASA AL SIGUIENTE PASO

La verificación del siguiente escaneo cierra el ciclo y queda en el historial.

VALOR

La corrección se revisa y se versiona como cualquier otro código.

● EXISTE HOY

Demostrar

ENTRADA

El historial de escaneos, incidentes y decisiones.

LO QUE HACE BERYX

Informes por perfil (ingeniería, gestión, dirección) con impacto estimado en euros, en PDF y programados; panel del CDO con objetivos y semáforos; dosier del art. 10 del Reglamento de IA (Business); feeds para Power BI; exportación de la auditoría; insignia pública por producto de datos.

● EXISTE HOY

QUIÉN DECIDE

La dirección lee; el DPD y la auditoría interna usan la evidencia.

GOBIERNO Y EVIDENCIA

El impacto en euros es una heurística declarada, que tiende a subestimar. Cada cifra dice cómo se calculó.

● EXISTE HOY

SALIDA

Informes, dosieres y exportaciones con fecha y fuente.

● EXISTE HOY

SE CONECTA CON

Power BI · Tableau · Looker · PDF · CSV

PASA AL SIGUIENTE PASO

Lo que se aprende vuelve a las reglas y a los umbrales del paso 3.

VALOR

Responder a un auditor o al regulador con historial, y no con una reconstrucción a toda prisa.

● EXISTE HOY

Lunes en la tienda online: del conector parado al informe semanal

Demostración con el pack «Loja online» de BeryX: datos sintéticos de una tienda online portuguesa, con ocho problemas plantados. EJEMPLO ILUSTRATIVO · DATOS SINTÉTICOS

01

Conectar

La base de pedidos se conecta en solo lectura. El primer escaneo perfila clientes, productos, pedidos y líneas de pedido en segundos.

02

Descubrir

El perfil marca siete columnas con datos personales (nombre, email, NIF y dirección) y la regla de formato detecta un 4% de emails inválidos.

03

Reglas

El pack de e-commerce instala reglas para el email, el país ISO, el código postal, la unicidad de los pedidos y la frescura. El perfil propone más reglas, que una persona acepta o rechaza.

04

Monitorizar

La regla de frescura falla: el pedido más reciente tiene 72 horas, con un límite de 24. El conector del marketplace se paró el viernes.

05

Incidente

El linaje agrupa las incidencias relacionadas en incidentes, cada uno con su origen, su severidad y los datasets afectados.

06

Corregir

Para los 14 order_id repetidos, BeryX propone el SQL de diagnóstico y el de corrección, también en PySpark. La corrección viene marcada para ejecutarse solo después de validar el diagnóstico.

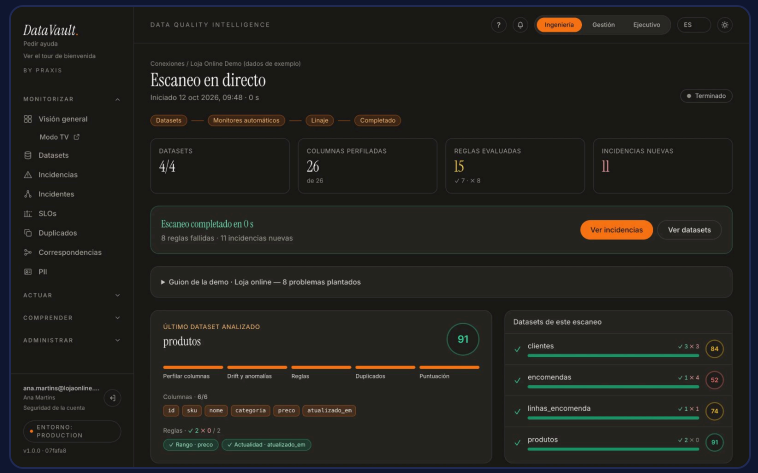
07

Demostrar

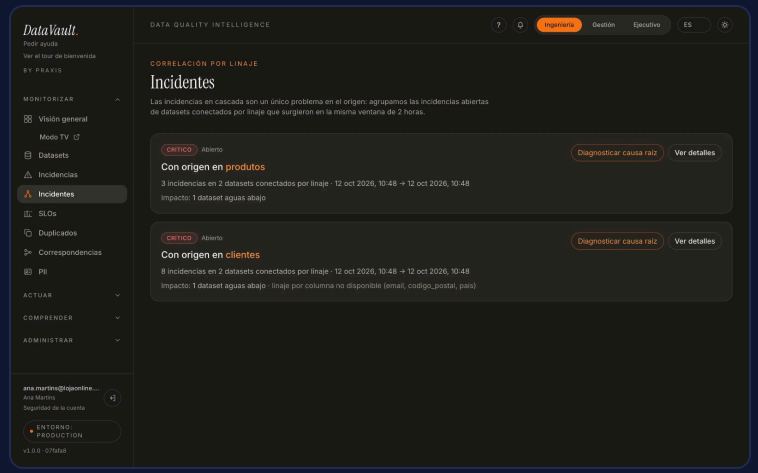
El informe ejecutivo resume el riesgo para el negocio, la exposición financiera estimada, la tendencia y los incidentes críticos, con la acción recomendada.

Reproducible: la cuenta gratuita incluye la demo de un clic con este pack y con tres más (ERP, Neobanco y Hospital).

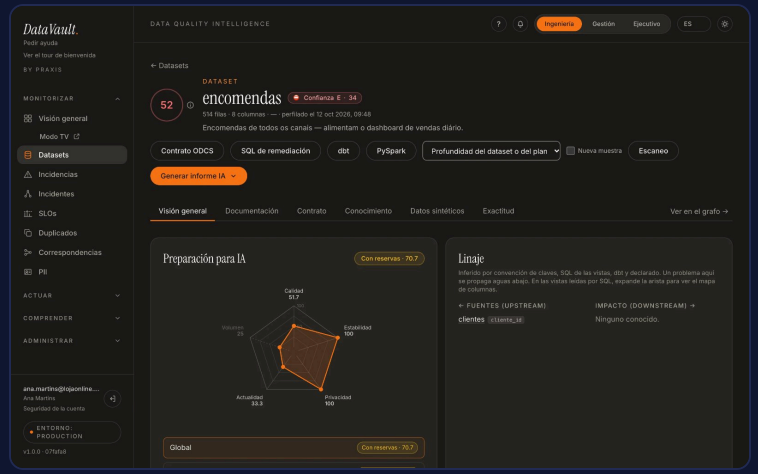
Lo que ve el equipo



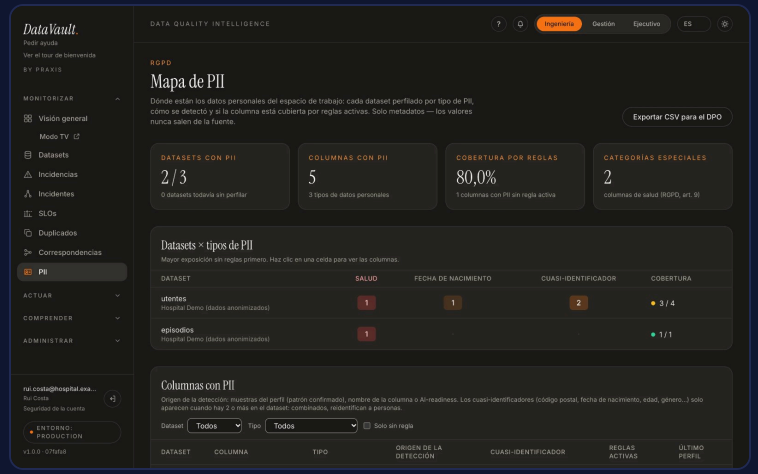
Escaneo en directo: reglas fallidas, incidencias nuevas y puntuación por dataset



Incidentes: incidencias agrupadas por el linaje, con el origen y el impacto



Detalle del dataset: puntuación por dimensión, preparación para la IA y linaje



Mapa de datos personales del workspace, con categorías especiales (salud)

Capturas de la app en funcionamiento con los packs de demostración (datos sintéticos).

Arquitectura de referencia

PERSONAS Y HERRAMIENTAS

- App web
- Slack y Teams
- Power BI, Tableau y Looker
- Agentes de IA por MCP
- API pública versionada

BERYX · GOOGLE CLOUD, EUROPE-WEST1

- API
- Workers asíncronos para escaneos e informes, con colas y DLQ
- Postgres con los metadatos, nunca las tablas
- Fachada de IA con un router por tarea
- Servidor MCP con 21 herramientas de solo lectura

TUS FUENTES

- Conexiones de solo lectura con los métodos oficiales de cada proveedor

> De las fuentes salen agregados, hasta 20 valores de ejemplo por columna y, para las reglas por fila, una muestra de hasta 5000 filas por tabla.

- SaaS compartido en la UE (europe-west1) — EXISTE HOY
- Instancia dedicada por cliente, con IP de salida propia y CMEK opcional — POR ACTIVAR
- Ejecución dentro del perímetro del cliente — VISIÓN

Encargados del tratamiento y regiones publicados en la página de seguridad, con 30 días de aviso antes de cualquier cambio.

IA con reglas: dónde entra, dónde no entra y quién decide

Dónde entra la IA

- Sugerencias de reglas, causa probable de los incidentes e informes por perfil — EXISTE HOY
- Asistente en lenguaje natural sobre el catálogo y las incidencias — EXISTE HOY
- Glosario y documentación de datasets, siempre para aprobación humana — EXISTE HOY

Dónde no entra

- Ninguna cifra de calidad, de anomalía o de coste sale de un LLM: los métodos son deterministas y están publicados — EXISTE HOY
- Proponer nunca ejecuta: cada acción pide aprobación humana y queda auditada — EXISTE HOY

Controles del cliente

- IA desactivada en cualquier plan, con informes en modo offline — EXISTE HOY
- Muestras en bruto o sintéticas, a elección del workspace — EXISTE HOY
- Clave propia de Anthropic (Business) o Claude en Amazon Bedrock o en Vertex AI, en regiones europeas (Enterprise) — EXISTE HOY

Guardas

- Tope diario de coste de IA y cuota mensual de tokens por plan — EXISTE HOY
- Router por tarea: por defecto, solo Anthropic recibe datos de clientes — EXISTE HOY
- Guarda de salida que enmascara secretos, emails, IBAN, tarjetas y NIF antes de guardar (todavía no cubre el asistente en streaming) — PARCIAL
- Defensa contra la inyección indirecta de instrucciones, con pruebas adversariales en el CI — EXISTE HOY
- Marcado del contenido generado por IA, como exige el art. 50 del Reglamento de IA — EXISTE HOY

POR ACTIVAR — Por defecto, muestras limitadas van a Anthropic, en EE. UU., al amparo de cláusulas contractuales tipo. El anonimizador (Presidio) está hecho, pero todavía no está en la imagen de producción.

RGPD: del inventario manual al mapa que se actualiza en cada escaneo

Saber dónde están los datos personales

- Mapa de datos personales: qué datasets guardan qué tipos de datos, cómo se detectaron y si tienen reglas activas — EXISTE HOY
- Sensibilidad por columna en cuatro niveles, con herencia por el linaje (sugerida) y curación con autor y fecha — EXISTE HOY
- Riesgo de reidentificación por cuasi-identificadores (k-anonimato); los ejemplos enviados a la IA salen generalizados — EXISTE HOY

Mantenerlos exactos

- Dígitos de control de NIF, IBAN y tarjetas, y listas de referencia — EXISTE HOY
- Exactitud frente a VIES, GeoAPI.pt, planes de numeración y registros MX, opcional — EXISTE HOY
- Auditoría de exactitud por muestreo humano, con intervalo de confianza — EXISTE HOY

Responder y demostrar

- Registro de solicitudes de los interesados con el plazo legal de cada jurisdicción: UE, Reino Unido, Suiza, Brasil, California y Angola — POR ACTIVAR
- Buscar y borrar el valor de un interesado en las muestras que guarda BeryX — POR ACTIVAR
- Retención de las muestras configurable por workspace y registro de auditoría exportable al SIEM — EXISTE HOY
- Búsqueda de credenciales olvidadas en código y notebooks — EXISTE HOY

Los casos de la CNPD, en preguntas que los datos responden

¿Dónde están los datos de salud, de religión o financieros?

El mapa de datos personales y la sensibilidad por columna. EXISTE HOY

¿Quién tiene acceso, y coincide con quien debería tenerlo?

Una regla de integridad referencial entre el directorio de cuentas y el registro de personal. EXISTE HOY

¿Adónde van los datos?

El linaje entre las fuentes conectadas. Los envíos por email o a terceros quedan fuera. PARCIAL

FinOps de datos: el coste leído de la propia plataforma y cruzado con la calidad

Coste y calidad juntos (Business)

- Telemetría nativa: BigQuery (INFORMATION_SCHEMA.JOBS), Snowflake (WAREHOUSE_METERING_HISTORY) y Databricks (system.billing.usage) — EXISTE HOY
- Cuánto del cómputo del mes se ejecutó sobre datos que no superan las comprobaciones — EXISTE HOY
- Coste por tabla en BigQuery; en Snowflake y Databricks el coste es de la cuenta, y el reparto aparece etiquetado como atribución. Sin telemetría, no aparece ninguna cifra — EXISTE HOY

Picos de coste

- Un pico diario abre una incidencia y avisa en los canales; en BigQuery, con las 5 consultas más caras del día y los literales enmascarados — EXISTE HOY
- Criterio publicado: $z \geq 3$, al menos un 50% y 1 USD por encima de la mediana, con 7 días de histórico — EXISTE HOY

Uso real

- Tablas calientes (20 o más consultas al día), activas, frías y sin uso (0 consultas en 90 días, solo con telemetría que cubra al menos 80 días) — EXISTE HOY
- Las tablas sin uso pasan a ser candidatas a darse de baja — EXISTE HOY
- Escaneos guiados por el uso: las calientes en profundidad en cada escaneo, las sin uso solo en el perfil mensual, de forma opcional y con aprobación del steward — EXISTE HOY

El coste del propio BeryX

- Tope de bytes antes de leer en BigQuery — EXISTE HOY
- Coste de IA por informe, por tarea y por proveedor; informes programados por la API de lotes a mitad de precio — EXISTE HOY
- Chargeback del uso por workspace, dominio y equipo, con CSV mensual y feed para Power BI — EXISTE HOY

PARCIAL — La telemetría de coste está hecha y probada, pero todavía no se ha demostrado en una cuenta real de BigQuery, Snowflake o Databricks. Mide el coste de cómputo, y no la eficiencia del formato de los datos.

Fuentes: Métodos de BeryX, anomalías de coste

Seguridad: solo lectura impuesta en el código, y las carencias puestas por escrito

Acceso a las fuentes

- Solo lectura en el código, probada contra servidores reales — EXISTE HOY
- Métodos oficiales de cada proveedor y credenciales cifradas (Fernet) con rotación de clave — EXISTE HOY
- Anti-SSRF con DNS fijado, también en los webhooks y los tickets — EXISTE HOY

Aislamiento e identidad

- Una prueba recorre todas las rutas de la API en busca de fugas entre clientes — EXISTE HOY
- SSO por OIDC y SAML 2.0 en todos los planes; MFA obligatoria para los administradores de la organización — EXISTE HOY
- Lista de IP permitidas por organización, también para los tokens de API — EXISTE HOY

Rastro y control

- Auditoría grabada en la misma transacción que el cambio, con exportación al SIEM — EXISTE HOY
- Acceso de soporte de BeryxTech solo con consentimiento, de solo lectura y durante 72 horas como máximo — EXISTE HOY
- Exportación de los datos del workspace y borrado en cascada con certificado firmado — EXISTE HOY

Lo que todavía no tenemos

- × Todavía no tenemos ISO 27001 ni SOC 2: están en curso, sin fechas.
- × Todavía no hemos hecho una prueba de penetración externa.
- × El WAF y Cloud Armor están escritos, pero todavía no están activos en producción.
- × Todavía sin un SLA contractual de disponibilidad ni de soporte.
- × Los modelos de contrato (DPA y MSA) todavía no los ha revisado un abogado.

Ecosistema: 30 tipos de fuente y las herramientas donde el equipo ya trabaja

BASES DE DATOS 7

PostgreSQL · MySQL y MariaDB · SQL Server, Azure SQL y Synapse · Oracle · Amazon Redshift · MongoDB · SQLite

WAREHOUSES 4

BigQuery · Snowflake · Databricks · Microsoft Fabric *

FICHEROS Y ALMACENAMIENTO DE OBJETOS 6

CSV · Google Sheets · Google Cloud Storage · Amazon S3 · Azure Blob y ADLS Gen2 · SFTP

APLICACIONES DE NEGOCIO 5

Salesforce · SAP S/4HANA y SAP Gateway · OData v2 y v4 · Dynamics 365 · REST y JSON

DOCUMENTOS 4

Confluence · Jira · SharePoint · Google Drive

STREAMING 2

Kafka · Pub/Sub

COLABORACIÓN 2

Slack · Teams

* Código hecho; el driver todavía no está en la imagen del SaaS.

- **Catálogos** · OpenMetadata y DataHub; Collibra y Purview validados solo con un transporte simulado — PARCIAL
- **BI** · Power BI, Tableau y Looker — EXISTE HOY
- **Tickets** · Jira, Linear, ServiceNow, Azure DevOps, GitHub y GitLab — EXISTE HOY
- **Orquestación** · dbt Cloud, dbt Core, Airflow, Fivetran y Airbyte — EXISTE HOY
- **Agentes de IA** · Servidor MCP publicado, con 21 herramientas de solo lectura — EXISTE HOY
- **Paquetes para desarrolladores** · SDK de Python, provider de Terraform, extensión de VS Code, Airflow y plugin de Claude Code, todavía sin publicar — POR ACTIVAR
- **SIEM** · Splunk, Microsoft Sentinel y Google SecOps — EXISTE HOY

Prueba por conector (matriz del 10/10/2026): 7 contra fuentes públicas reales, 2 contra servicios reales en Docker, 13 con transporte simulado, 6 sin driver en el entorno de prueba y 1 por verificar.

Cada cifra dice cómo se calculó

Tres tipos de cifra
exacta, estimada o por muestreo, marcadas en la app

EXISTE HOY

Intervalos de confianza
al 95%, de Wilson con corrección para población finita; z de Fisher en las correlaciones

EXISTE HOY

Puntuación con pesos fijos
completitud 25%, unicidad 25%, validez 20%, actualidad 15% y consistencia 15%

EXISTE HOY

Anomalías
z-score modificado (mediana y MAD) con control de falsos positivos de Benjamini-Hochberg

EXISTE HOY

Impacto en euros
heurística declarada, siempre estimada, que tiende a subestimar

EXISTE HOY

Métodos públicos
el documento está publicado y una prueba falla si una constante documentada cambia en el código

EXISTE HOY

> Ningún método de cálculo usa un LLM.

Fuentes: Métodos de BeryX

Seis perfiles, cada uno con su guía y su tour

Administrador del workspace

Conecta fuentes, gestiona el equipo, los canales de alerta y la política de IA.

● EXISTE HOY

Guía del perfil →

Miembro

Trabaja incidencias, reglas, sugerencias e incidentes.

● EXISTE HOY

Guía del perfil →

Lector

Consulta puntuaciones, informes y el catálogo sin cambiar nada. Todavía no se asigna en la página Equipo.

○ PARCIAL

Guía del perfil →

Responsable de dominio

Decide sobre los datasets, los incidentes y las reglas de su dominio.

● EXISTE HOY

Guía del perfil →

Administrador de la organización

Gobierno de todos los workspaces: SSO, IP permitidas, política de datos y chargeback.

● EXISTE HOY

Guía del perfil →

Facturación

Plan, facturas e informe de costes por equipo.

● EXISTE HOY

Guía del perfil →

En el primer inicio de sesión, un tour por perfil enseña la app. Las guías están publicadas en la documentación de BeryX.

Qué cambia para cada función

FUNCIÓN	HOY	CON BERYX
Ingeniería de datos	Se entera de los fallos por el negocio; tests repartidos por repositorios	Monitores automáticos, incidentes con linaje, reglas versionadas y un gate en el CI
DPD y jurídico	Inventario manual, revisado una vez al año	Mapa de datos personales en cada escaneo, sensibilidad con autor y evidencia exportable
FinOps y finanzas	La factura a final de mes, sin relación con la calidad	Coste diario por dataset o por cuenta, picos de coste con las consultas más caras y tablas sin uso
Dirección y CDO	Informes sin fuente ni tendencia	Panel con objetivos y semáforos, e informes por perfil
Equipos de IA	Datos de entrenamiento que nadie mide	Preparación por finalidad, dosier del art. 10 y datos sintéticos con evaluación de privacidad

Panorama: dónde es más fuerte cada enfoque

Suites enterprise · Collibra, Informatica

- Catálogo y glosario como sistema de registro de la empresa
- ISO 27001 y SOC 2, con un trust center público
- Limpieza y MDM que escriben en tus sistemas, con credenciales de escritura

COSTE Precio público en UK G-Cloud: de 74 999 £ a 122 919 £ al año, más la implantación

UK G-Cloud — Collibra Data Quality (servicio 136629392504870) · UK G-Cloud — Informatica IDMC (servicio 727452156234281) · Collibra — Trust Center

Open source · Great Expectations, Soda

- Gratis y ejecutándose dentro de tu perímetro
- SQL y Python sin límites en los tests
- Cada test lo escribe y lo mantiene tu equipo

COSTE Licencia 0; Soda Team a 750 $ al mes, más cómputo y horas de ingeniería

Great Expectations — planes y precios · Soda — planes y precios

Observabilidad · Monte Carlo, Sifflet

- Monitores por metadatos maduros en warehouses en la nube
- SOC 2 Type II e ISO 27001
- Detectan y alertan; la corrección queda en manos del equipo

COSTE Contratos anuales; Sifflet a unos 48 000 $ al año según su precio publicado

Monte Carlo — documentación: seguridad y cumplimiento · Monte Carlo — Trust Center · Sifflet — planes y precios

BeryX

- Calidad, datos personales y coste medidos en el mismo catálogo
- Propone la corrección en SQL o como pull request, sin credenciales de escritura
- Todavía sin certificaciones ni SLA contractual

COSTE Gratuito · desde 490 €/mes

Matriz completa, con una fuente fechada en cada celda →

Resumen de la matriz comparativa del sitio, con una fuente pública fechada en cada celda (verificadas el 28/09/2026).

Dónde está la diferencia, y lo que no afirmamos

Lo que se puede probar hoy

- Calidad, datos personales y coste medidos sobre el mismo catálogo y el mismo historial
- Solo lectura impuesta en el código, con correcciones como propuestas y pull requests
- Cada cifra con un método publicado y un intervalo de confianza
- IA con reglas de salida, modo offline y la opción de desactivarla
- Precio publicado y una cuenta gratuita sin llamada comercial

Lo que no afirmamos

- ✓ Que BeryX habría evitado una multa concreta
- ✓ Certificaciones: todavía no tenemos ISO 27001 ni SOC 2
- ✓ Un SLA contractual
- ✓ Que todos los conectores se hayan probado en una cuenta real
- ✓ Que sustituye a un MDM o a un catálogo enterprise
- ✓ Que corrige los datos por sí solo

La prueba que proponemos es un piloto, con métricas medidas antes y después.

Lo que pasa a ser medible

MÉTRICA	CÓMO LA MIDE BERYX	ESTADO
Puntuación de calidad por dataset y por dominio, con tendencia	Cinco dimensiones con pesos fijos, en cada escaneo	EXISTE HOY
Cumplimiento del SLO por dataset	Objetivo de servicio por dataset e historial de cumplimiento	EXISTE HOY
Incidencias abiertas y resueltas por severidad	Incidencias e incidentes con fechas, dueño y plazo de respuesta	EXISTE HOY
Columnas con datos personales sin reglas o sin dueño	El mapa de datos personales cruzado con las reglas y los dueños	EXISTE HOY
Plazo de cada solicitud de un interesado	Registro de solicitudes con el plazo de la jurisdicción	POR ACTIVAR
Cómputo gastado en datos que no superan las comprobaciones	Telemetría de coste de la plataforma cruzada con la puntuación	EXISTE HOY
Tablas sin uso en 90 días	Recuento de consultas por tabla en las vistas de sistema	EXISTE HOY
Coste de IA por informe y por equipo	Coste por llamada, por tarea y chargeback por equipo	EXISTE HOY

Los valores de partida salen del primer mes de medición; no usamos medias de mercado.

El libro honesto

● EXISTE HOY

- Perfilado, 16 tipos de regla, monitores automáticos, anomalías y SLO
- Incidentes con linaje, causa probable y tickets
- Corrección propuesta, ensayada y por pull request (Business)
- Mapa de datos personales, sensibilidad y riesgo de reidentificación
- Coste y calidad juntos, picos de coste, uso real y chargeback
- Informes por perfil, asistente y servidor MCP
- SSO por OIDC y SAML, MFA, IP permitidas, auditoría y SIEM
- Tour y guía escrita por perfil

◐ PARCIAL

- Conectores todavía sin prueba en una cuenta real, entre ellos BigQuery, Snowflake, Databricks y Oracle
- Collibra y Purview validados solo con un transporte simulado
- Páginas de metadatos y dosier de elementos críticos (BCBS 239) solo por API
- Correspondencia de entidades sobre muestras de hasta 50 000 filas
- La guarda de salida de la IA todavía no cubre el asistente en streaming
- El rol de Lector todavía no se asigna en la página Equipo

○ POR ACTIVAR

- Solicitudes de los interesados y exportación a SodaCL y Great Expectations: hechas, en el próximo despliegue
- Siete conectores fuera de la imagen del SaaS
- El anonimizador en la imagen de producción
- Instancia dedicada, alta disponibilidad, WAF e IP de salida fija
- SDK, Terraform, VS Code y Airflow sin publicar
- Pago en línea

○ VISIÓN

- Certificaciones ISO 27001, SOC 2, ISO 27701 e ISO 42001: todavía no tenemos ninguna
- Prueba de penetración externa
- SCIM y llaves de seguridad (WebAuthn)
- Conectividad privada y ejecución dentro del perímetro del cliente
- Sincronización con ServiceNow en los dos sentidos

Planes

Gratuito

0 € /mes

- 1 fuente
- 25 datasets
- 5 informes de IA al mes
- 30 días de histórico

Team

490 € /mes

- 5 fuentes
- 500 datasets
- 200 informes de IA al mes
- 365 días de histórico
- ✓ Tokens de API
- ✓ Webhooks de alertas
- ✓ Análisis de documentos para RAG
- ✓ Informes programados por email
- ✓ Datos sintéticos a partir del perfil

Business

1190 € /mes

- 15 fuentes
- 2000 datasets
- 1000 informes de IA al mes
- 730 días de histórico
- ✓ Dosier de conformidad del AI Act
- ✓ Telemetría de coste de las plataformas
- ✓ Remediación como Pull Request (GitHub/GitLab)
- ✓ Clave de Anthropic propia (BYOK)
- ✓ Elección de modelos de IA por tarea

Enterprise

Bajo consulta

- sin límite
- ✓ Claude en la cuenta cloud de la organización (Amazon Bedrock o Google Vertex AI)

En todos los planes: motor de calidad, sugerencias, modo offline, IA que se puede desactivar y SSO.

Precios sin IVA, con facturación anual; la facturación mensual cuesta un 20 % más. Mientras el pago en línea no esté activo, la suscripción se hace con BeryxTech.

Cómo empezar

01

Probarlo por tu cuenta

Cuenta gratuita con una fuente y la demo de un clic, con cuatro packs de datos sintéticos: Loja online, ERP, Neobanco y Hospital.

Crear una cuenta gratuita →

02

Una conversación con BeryxTech.

Treinta minutos para rellenar el cuadro de costes con tus cifras y ver la demo sobre tu caso.

Hablar con nosotros →

03

Un piloto en un dominio

Un dominio de datos, las métricas de la diapositiva anterior medidas al principio y al final, y una decisión tomada con cifras.

Proponer un piloto →

Datos en los que ingeniería, el DPD y finanzas confían, con la misma evidencia.

BeryX, un producto de BeryxTech.

beryxtech.com/es/apresentacao