APRESENTAÇÃO DO PRODUTO · OUTUBRO DE 2026

BeryxTech

Qualidade de dados, RGPD e custo dos dados numa só camada: lida na fonte, só de leitura e com o cálculo de cada número à vista.

Para a equipa de dados, para o DPO e para quem paga a fatura da cloud.

CADA AFIRMAÇÃO LEVA UMA ETIQUETA

● EXISTE HOJE
em produção e coberto por testes

◐ PARCIAL
existe, com os limites ditos no slide

◐ POR ATIVAR
feito no código, falta ligar em produção

○ VISÃO
ainda não existe

Factos verificados a 12/10/2026 contra o produto em produção e as fontes citadas.

As decisões e a IA já correm sobre os dados. A confiança neles não acompanhou.

67%

não confiam totalmente nos dados que usam para decidir (eram 55% em 2023)

Precisely e Drexel LeBow, set. 2024 · 565+ profissionais · estudo de fornecedor

64%

apontam a qualidade dos dados como o maior desafio de integridade (eram 50% em 2023)

Precisely e Drexel LeBow, set. 2024 · estudo de fornecedor

71%

temem que resultados errados ou alucinados cheguem a quem decide

dbt Labs, State of Analytics Engineering, abr. 2026 · 363 profissionais · estudo de fornecedor

60%

dos projetos de IA sem dados prontos para IA serão abandonados até ao fim de 2026, segundo a previsão da Gartner

Gartner, comunicado de 26/02/2025

Quando a IA responde sobre dados por verificar, o erro chega mais depressa a quem decide.

Três custos que não aparecem em nenhum orçamento

01 ENGENHARIA DE DADOS

O negócio descobre a falha antes da equipa de dados

Dados parados, duplicados ou inválidos não partem nenhum job. Aparecem no fecho do mês, num relatório à direção ou numa fatura que não sai.

02 RGPD E IA

A exatidão dos dados pessoais é uma obrigação legal

O RGPD pede dados exatos e atualizados, resposta aos titulares num mês e notificação de violações em 72 horas. Violar os princípios dá coimas até 20 M€ ou 4% do volume de negócios mundial. O Regulamento da IA junta exigências sobre os dados de treino a partir de dezembro de 2027.

RGPD, arts. 5.º, 12.º, 33.º e 83.º; Regulamento da IA, arts. 10.º e 113.º

03 FINOPS DE DADOS

A fatura dos dados cresce sem dono

Tabelas que ninguém lê continuam a ser carregadas. Uma query mal escrita multiplica o custo do dia e só se vê na fatura do mês. Paga-se compute para processar dados que falham a medição.

Os três custos têm a mesma raiz: ninguém mede os dados de forma contínua, com dono e com prova.

Os dados falham em silêncio: os jobs ficam verdes e os números ficam errados

ANATOMIA DE UM INCIDENTE DE DADOS

EXEMPLO ILUSTRATIVO · DADOS SINTÉTICOS

SEXTA, 18H

O conector do marketplace deixa de sincronizar. Nenhum job falha: só deixam de chegar linhas.

FIM DE SEMANA

As encomendas do marketplace não entram no data warehouse. O dashboard de vendas mostra uma quebra que não existiu.

SEGUNDA, 10H

O fecho semanal não bate. Começa a investigação: que tabela, desde quando e quem mais usa estes dados?

TERÇA

Encontrada a origem, falta saber que relatórios e modelos já consumiram os dados errados.

SEIS FORMAS DE FALHAR SEM ERRO

Frescura
a tabela deixou de receber dados

Volume
chegou metade das linhas

Schema
uma coluna mudou de tipo ou de nome

Duplicados
um retry gravou a mesma encomenda duas vezes

Validade
NIF com 8 dígitos, IBAN com o dígito de controlo errado

Integridade
pedidos sem cliente depois de um apagamento sem cascata

A história e os exemplos vêm dos packs de demonstração do BeryX, com dados sintéticos. Cada falha corresponde a um monitor ou a uma regra do produto.

O custo de não saber conta-se com os teus números

PILAR	O QUE CONTAR	COMO ESTIMAR	O QUE O BERYX MEDE
Engenharia de dados	Incidentes de dados por mês, horas de investigação e dias até à deteção	Incidentes × horas × custo por hora, mais dias até à deteção × valor em risco por dia (faturas por emitir, por exemplo)	Histórico de problemas e incidentes com datas, e impacto estimado em euros por problema, rotulado como heurística
RGPD	Pedidos de titulares por mês e datasets com dados pessoais sem dono ou sem regras	Pedidos × horas por pedido; exposição máxima de 20 M€ ou 4% do volume de negócios mundial	Mapa de dados pessoais com a cobertura de regras de cada coluna
FinOps de dados	Tabelas sem leitura há 90 dias, compute gasto em dados que falham a medição e picos diários	Custo mensal dessas tabelas, mais o compute atribuído aos datasets com problemas abertos	Custo lido das vistas de sistema da plataforma, por dia (por dataset no BigQuery, por conta na Snowflake e no Databricks), picos de custo e uso real por tabela

Fontes: RGPD, art. 83.º, n.º 5

Numa primeira conversa preenchemos este quadro contigo, com os dados das tuas fontes.

O que a lei pede aos dados, e que só se prova com medições

OBRIGAÇÃO	O QUE PEDE	PRAZO OU LIMITE	O QUE EXIGE DOS DADOS
Exatidão · RGPD, art. 5.º, n.º 1, al. d)	Dados pessoais exatos e atualizados; os inexatos apagados ou retificados sem demora	Coima até 20 M€ ou 4% do volume de negócios mundial (art. 83.º, n.º 5)	Saber que dados pessoais estão errados, onde e desde quando
Direitos dos titulares · art. 12.º, n.º 3	Responder a pedidos de acesso, retificação ou apagamento	Um mês, prorrogável por mais dois	Saber em que sistemas estão os dados de cada pessoa
Violação de dados · art. 33.º, n.º 1	Notificar a autoridade de controlo, em Portugal a CNPD	72 horas, quando possível	Saber que dados e que titulares foram afetados
Avaliação de impacto · art. 35.º	Avaliar antes de um tratamento de risco elevado	Antes de começar o tratamento	Inventário das categorias de dados e dos fluxos
Dados para IA de alto risco · Regulamento da IA, art. 10.º, n.º 3	Dados de treino, validação e teste relevantes, representativos e, tanto quanto possível, sem erros e completos	A partir de 2 de dezembro de 2027 (anexo III)	Medir erros e completude dos conjuntos de dados

Resumo para enquadramento; não é parecer jurídico.

Fontes: RGPD, Regulamento (UE) 2016/679 · Regulamento da IA, arts. 10.º e 113.º (AI Act Service Desk)

A CNPD já aplicou coimas de milhões por falhas na gestão de dados pessoais

4,3 M€

Instituto Nacional de Estatística

DEZEMBRO DE 2022

Censos 2021. Cinco contraordenações: dados de saúde e de religião tratados sem base de licitude clara, informação insuficiente aos respondentes, subcontratante nos EUA escolhido sem a diligência devida, transferências internacionais sem medidas complementares e nenhuma avaliação de impacto.

ESTADO O INE anunciou que ia impugnar a decisão em tribunal.

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

1,25 M€

Câmara Municipal de Lisboa

JANEIRO DE 2022

Dados de promotores de manifestações enviados a embaixadas e ao Ministério dos Negócios Estrangeiros: 225 contraordenações.

ESTADO Os tribunais confirmaram a infração. Com parte das contraordenações prescritas, o montante desceu para 738 mil € no Tribunal Central Administrativo Sul (agosto de 2025), numa decisão ainda passível de recurso.

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

400 mil €

Centro Hospitalar Barreiro-Montijo

2018

A primeira coima do RGPD em Portugal: 985 contas com perfil de médico para 296 médicos e acesso indevido a processos clínicos.

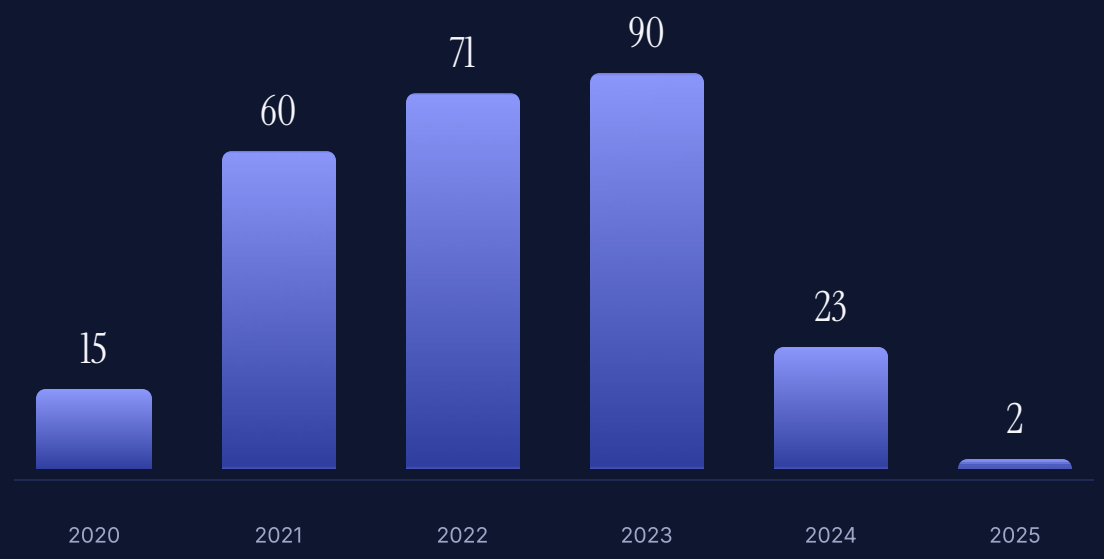
ESTADO Passou a coima única de 380 mil €. Em julho de 2020, a CNPD aceitou dispensá-la, invocando a pandemia e as contas do hospital.

Garrigues, 2018 · Renascença, 27/05/2021

Não dizemos que o BeryX teria evitado estas coimas. Os três casos mostram o que a CNPD investiga: onde estão os dados sensíveis, quem lhes acede e para onde vão.

As coimas sobem e descem com os meios da CNPD. As violações notificadas continuam a subir.

COIMAS APLICADAS PELA CNPD, POR ANO



472

processos de violação de dados em 2025, mais 42% do que em 2024 (357 em entidades privadas, 115 em públicas)

CNPD, Relatório de Atividades 2025, pp. 15 e 34

88

processos de contraordenação abertos em 2025, que «poderiam ter originado a aplicação de 90 coimas, em vez de 2»

CNPD, Relatório de Atividades 2025, p. 42

2 037

averiguações abertas em 2025

CNPD, Relatório de Atividades 2025, p. 42

A própria CNPD atribui a quebra de 2025 à falta de pessoas para instruir processos e, em setembro de 2025, propôs à Assembleia da República um regime de contraordenações mais rápido. As obrigações mantêm-se: notificar em 72 horas, responder num mês, provar a exatidão.

Fontes: CNPD, Relatório de Atividades 2025, aprovado a 24/03/2026 (fig. 10, pp. 12, 42 e 43)

A fatura dos dados cresce sem dono e sem ligação à qualidade

84%

dizem que gerir o gasto em cloud é hoje o maior desafio da cloud para as organizações

Flexera, State of the Cloud 2025 · 750+ profissionais · estudo de fornecedor

- Tabelas que ninguém lê continuam a ser carregadas, guardadas e monitorizadas.
- Uma query mal escrita multiplica o custo do dia e só se vê na fatura do mês.
- Paga-se compute para processar dados que falham a medição.
- O custo da IA cresce sem teto e sem repartição por equipa.

A fatura chega à direção financeira e os problemas de qualidade ficam na equipa de dados. O BeryX põe os dois lado a lado, por dataset sempre que a plataforma o permite.

Onde vive hoje a verdade sobre os teus dados?

Hoje

- × Em testes e scripts espalhados por equipas e repositórios
- × Na memória de quem conhece as tabelas
- × Num registo de tratamentos em folha de cálculo, revisto uma vez por ano
- × Na fatura da cloud, no fim do mês
- × Em dashboards que ninguém liga à origem dos dados

Com o BeryX

- ✓ Em regras e contratos versionados, que correm em cada scan ● EXISTE HOJE
- ✓ Num catálogo com dono, domínio, sensibilidade e uso por dataset ● EXISTE HOJE
- ✓ Num mapa de dados pessoais atualizado a cada scan ● EXISTE HOJE
- ✓ No custo diário lido das vistas de sistema do BigQuery, do Snowflake e do Databricks ● EXISTE HOJE
- ✓ Em incidentes com linhagem, dono e prazo ● EXISTE HOJE

Um registo de confiança por dataset, partilhado pela engenharia, pelo DPO e pela direção financeira.

Porquê agora

2025

As violações de dados notificadas à CNPD sobem 42%, para 472 processos.

CNPD, Relatório de Atividades 2025

SET. 2025

A CNPD propõe à Assembleia da República um regime de contraordenações mais rápido, para travar prescrições.

CNPD, Relatório de Atividades 2025, p. 12

2026

Previsão da Gartner: as organizações abandonam 60% dos projetos de IA sem dados prontos para IA.

Gartner, 26/02/2025

2 AGO. 2026

O Regulamento da IA passa a aplicar-se em geral (art. 113.º).

AI Act Service Desk, art. 113.º

2 DEZ. 2027

Obrigações de alto risco do anexo III, incluindo a qualidade dos dados de treino, validação e teste (art. 10.º).

AI Act Service Desk, arts. 10.º e 113.º

2 AGO. 2028

Obrigações de alto risco do anexo I, os produtos regulados.

AI Act Service Desk, art. 113.º

> Pôr os dados em condições leva meses: perfil, regras, donos e um histórico de medições que mostre a evolução.

Uma camada de confiança sobre os dados que já tens, sem os copiar

AS TUAS FONTES FICAM ONDE ESTÃO

- Bases de dados
- Data warehouses
- Ficheiros e object storage
- Aplicações de negócio: SAP, Salesforce, Dynamics 365
- Documentos
- Streaming

→

BERYX · SÓ DE LEITURA

Descobrir | Validar | Monitorizar | Explicar | Provar

- ✓ Agregados calculados na fonte
- ✓ Amostras limitadas e com retenção
- ✓ Propõe correções, nunca escreve nas fontes
- ✓ Cada número diz como foi calculado

→

Engenharia de dados
incidentes com linhagem, SLOs, regras como código

DPO e jurídico
mapa de dados pessoais, sensibilidade, evidência

FinOps e direção
custo e qualidade lado a lado, uso real, relatórios por perfil

ONDE A EQUIPA JÁ TRABALHA: Slack | Teams | Jira | Power BI | Email | Webhooks | API | MCP

30 tipos de fonte no registo de conectores; 1 ainda por instalar no SaaS, o Microsoft Fabric (ver Ecossistema).

Do dado ligado ao dado provado, e de volta ao início

01
Ligar
credenciais só de leitura

→

02
Descobrir
perfil, dados pessoais, uso

→

03
Regras
propostas pelo perfil, aceites por pessoas

→

04
Monitorizar
frescura, volume, schema, anomalias, custo

→

05
Incidente
linhagem, impacto, dono, prazo

→

06
Corrigir
SQL ou pull request, ensaiado antes

→

07
Provar
relatórios, evidência, auditoria

↺ O que se aprende volta às regras: limiares, falsos positivos marcados e regras sugeridas a partir dos problemas resolvidos.

Cada passo fica registado: quem decidiu, quando e com que evidência.

Cada dataset leva consigo o seu registo de confiança

Dono e domínio
quem responde e a que área pertence
EXISTE HOJE

Sensibilidade por coluna
pública, interna, confidencial ou dados pessoais, com autor e data
EXISTE HOJE

Regras e contrato
regras ativas e contrato ODCS versionado
EXISTE HOJE

Score e SLO
cinco dimensões, tendência e objetivo de serviço
EXISTE HOJE

Uso real
queries por dia: quente, ativa, fria ou sem uso
EXISTE HOJE

Custo
compute atribuído, quando a plataforma o reporta
EXISTE HOJE

Incidentes e linhagem
o que falhou, a montante e a jusante
EXISTE HOJE

Auditoria
cada mudança, gravada na mesma transação
EXISTE HOJE

O engenheiro, o DPO e o diretor financeiro leem o mesmo registo. Ninguém volta a explicar o contexto.

Ligar

ENTRADA

Uma credencial só de leitura de uma base de dados, de um warehouse, de ficheiros ou de uma aplicação.

O QUE O BERYX FAZ

Testa a ligação, descobre schemas e tabelas e importa os metadados. A credencial é cifrada antes de ser gravada e o endereço é validado contra SSRF antes do primeiro pedido.

● EXISTE HOJE

QUEM DECIDE

O administrador do workspace escolhe a fonte, o âmbito e a frequência dos scans.

GOVERNAÇÃO E PROVA

Só leitura imposto no código: o cliente HTTP recusa escritas antes de sair e as sessões Postgres abrem em modo só de leitura. Nos testes, servidores reais de SQL Server, Oracle e MongoDB recusaram escritas.

● EXISTE HOJE

SAÍDA

Catálogo de datasets com schema, agendamento e credencial cifrada.

● EXISTE HOJE

LIGA-SE A

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

PASSA AO PASSO SEGUINTE

A lista de datasets e colunas segue para o perfil.

VALOR

Sem credenciais de escrita, sem software instalado nas fontes e sem copiar tabelas.

● EXISTE HOJE

Descobrir

ENTRADA

Os datasets ligados.

O QUE O BERYX FAZ

Perfil por coluna com agregados calculados na fonte: nulos, distintos, padrões e distribuições. Reconhece NIF, IBAN, email e código postal, monta o mapa de dados pessoais, classifica a sensibilidade em quatro níveis e estima o risco de reidentificação (k-anonimato). Nas plataformas com telemetria, mede o uso real de cada tabela.

● EXISTE HOJE

QUEM DECIDE

O DPO confirma ou corrige a sensibilidade de cada coluna. O scan nunca sobrepõe uma decisão manual.

GOVERNAÇÃO E PROVA

Cada número diz se é exato, estimado ou por amostra, com intervalo de confiança a 95%. Cada mudança de classificação fica registada com autor e data.

● EXISTE HOJE

Métodos do BeryX

SAÍDA

Perfil, score por dimensão, mapa de dados pessoais, sensibilidade por coluna e uso.

● EXISTE HOJE

LIGA-SE A

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

PASSA AO PASSO SEGUINTE

O perfil alimenta as regras sugeridas.

VALOR

O DPO vê onde estão os dados pessoais a partir das próprias fontes, sem depender de um inventário manual.

● EXISTE HOJE

Regras e contratos

ENTRADA

O perfil, as sugestões e os testes que a equipa já tem.

O QUE O BERYX FAZ

Dezasseis tipos de regra: formatos com dígito de controlo (NIF, IBAN, cartões), unicidade, intervalos, listas de referência, integridade referencial, frescura, volume e SQL próprio só de leitura. Exatidão contra fontes externas, por opção: VIES, GeoAPI.pt, planos de numeração e registos MX. Importa testes dbt, Great Expectations, SodaCL e contratos ODCS.

● EXISTE HOJE

QUEM DECIDE

O perfil e a IA propõem; uma pessoa aceita cada regra.

GOVERNAÇÃO E PROVA

O SQL do cliente passa por uma guarda: um único SELECT, numa transação sempre revertida. A configuração também vive como código, num datavault.yaml versionado.

● EXISTE HOJE

SAÍDA

Regras ativas por dataset e contrato versionado; exportação para dbt, com SodaCL e Great Expectations por ativar.

◐ PARCIAL

LIGA-SE A

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

PASSA AO PASSO SEGUINTE

As regras aceites correm em cada scan e alimentam os monitores.

VALOR

As regras deixam de viver em scripts de uma só pessoa.

● EXISTE HOJE

Monitorizar

ENTRADA

Regras, histórico de perfis e telemetria de custo e de uso.

O QUE O BERYX FAZ

Monitores automáticos de frescura, volume e schema, até 500 por conexão em cada scan. Anomalias com z-score modificado e controlo de falsos positivos (Benjamini-Hochberg), drift de distribuição, SLO por dataset e picos de custo diário, com as queries mais caras do dia no BigQuery.

● EXISTE HOJE

QUEM DECIDE

O dono do dataset define o SLO e a severidade. Um falso positivo marcado como «ignorado» silencia a série.

GOVERNAÇÃO E PROVA

Métodos publicados e testados contra o código. Nenhum número de qualidade, de anomalia ou de custo vem de um LLM.

● EXISTE HOJE

SAÍDA

Alertas na app, por email, no Slack, no Teams e por webhook, com escalamento de piquete.

● EXISTE HOJE

LIGA-SE A

Slack · Teams · Email · Webhooks assinados

PASSA AO PASSO SEGUINTE

Um alerta abre ou atualiza um problema, que pode juntar-se a um incidente.

VALOR

A falha chega primeiro à equipa de dados, e não ao fecho do mês.

● EXISTE HOJE

Incidentes e causa raiz

ENTRADA

Um alerta ou um problema aberto.

O QUE O BERYX FAZ

Agrupa problemas em incidentes pela linhagem (chaves estrangeiras, SQL das vistas, manifest dbt, OpenLineage) e mostra o impacto a jusante. A IA propõe a causa provável a partir de metadados e de amostras limitadas. Prazos de resposta por severidade, com escalamento.

● EXISTE HOJE

QUEM DECIDE

O responsável do domínio aceita, atribui e fecha o incidente.

GOVERNAÇÃO E PROVA

Trilho de auditoria gravado na mesma transação da mudança e exportável para o SIEM: Splunk, Sentinel ou Google SecOps.

● EXISTE HOJE

SAÍDA

Incidente com linhagem, impacto, dono, prazo e histórico, e ticket no Jira, Linear, ServiceNow, Azure DevOps, GitHub ou GitLab.

● EXISTE HOJE

LIGA-SE A

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

PASSA AO PASSO SEGUINTE

A causa provável e o impacto seguem para a proposta de correção.

VALOR

Saber que relatórios e modelos foram afetados antes de alguém perguntar.

● EXISTE HOJE

Corrigir

ENTRADA

Um incidente com causa provável.

O QUE O BERYX FAZ

Propõe a correção em SQL, no dialeto da fonte, ou em PySpark, e ensaia-a numa amostra antes de a aceitar. A partir do plano Business, a correção aprovada abre um pull request no GitHub ou no GitLab do cliente. Normalização e deduplicação também saem como propostas.

● EXISTE HOJE

QUEM DECIDE

Uma pessoa aprova e o cliente aplica pelo seu próprio CI. O BeryX nunca escreve nas fontes.

GOVERNAÇÃO E PROVA

Fluxo proposta, aceite, aplicada e verificada: o scan seguinte confirma se o problema desapareceu.

● EXISTE HOJE

SAÍDA

Pull request ou SQL aprovado, ligado ao dataset e ao ticket.

● EXISTE HOJE

LIGA-SE A

GitHub · GitLab · SQL · PySpark · dbt

PASSA AO PASSO SEGUINTE

A verificação do scan seguinte fecha o ciclo e fica no histórico.

VALOR

A correção é revista e versionada como qualquer outro código.

● EXISTE HOJE

Provar

ENTRADA

O histórico de scans, incidentes e decisões.

O QUE O BERYX FAZ

Relatórios por perfil (engenharia, gestão, direção) com impacto estimado em euros, em PDF e agendados; painel do CDO com metas e semáforos; dossiê do art. 10.º do Regulamento da IA (Business); feeds para o Power BI; exportação da auditoria; badge público por produto de dados.

● EXISTE HOJE

QUEM DECIDE

A direção lê; o DPO e a auditoria interna usam a evidência.

GOVERNAÇÃO E PROVA

O impacto em euros é uma heurística declarada, que tende a subestimar. Cada número diz como foi calculado.

● EXISTE HOJE

SAÍDA

Relatórios, dossiês e exportações com data e fonte.

● EXISTE HOJE

LIGA-SE A

Power BI · Tableau · Looker · PDF · CSV

PASSA AO PASSO SEGUINTE

O que se aprende volta às regras e aos limiares do passo 3.

VALOR

Responder a um auditor ou à CNPD com histórico, e não com uma reconstrução à pressa.

● EXISTE HOJE

Segunda-feira na loja online: do conector parado ao relatório da semana

Demonstração com o pack «Loja online» do BeryX: dados sintéticos de uma loja portuguesa, com oito problemas plantados.

EXEMPLO ILUSTRATIVO · DADOS SINTÉTICOS

01

Ligar

A base de encomendas liga-se só de leitura. O primeiro scan perfila clientes, produtos, encomendas e linhas de encomenda em segundos.

02

Descobrir

O perfil marca sete colunas com dados pessoais (nome, email, NIF e morada) e a regra de formato acusa 4% de emails inválidos.

03

Regras

O pack de e-commerce instala regras para o email, o país ISO, o código postal, a unicidade das encomendas e a frescura. O perfil propõe mais regras, que uma pessoa aceita ou recusa.

04

Monitorizar

A regra de frescura falha: a encomenda mais recente tem 72 horas, para um limite de 24. O conector do marketplace parou na sexta.

05

Incidente

A linhagem junta os problemas relacionados em incidentes, cada um com a origem, a severidade e os datasets afetados.

06

Corrigir

Para os 14 order_id repetidos, o BeryX propõe o SQL de diagnóstico e o de correção, também em PySpark. A correção vem marcada para correr só depois de validado o diagnóstico.

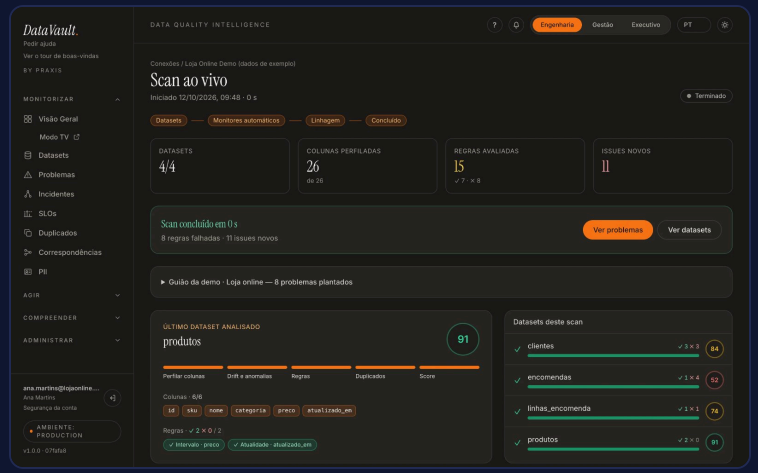
07

Provar

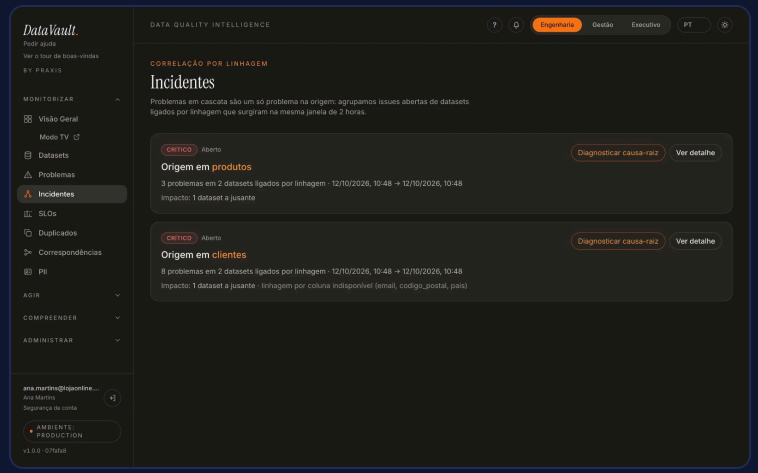
O relatório executivo resume o risco para o negócio, a exposição financeira estimada, a tendência e os incidentes críticos, com a ação recomendada.

Reproduzível: a conta gratuita traz a demo de um clique com este pack e com mais três (ERP, Neobanco e Hospital).

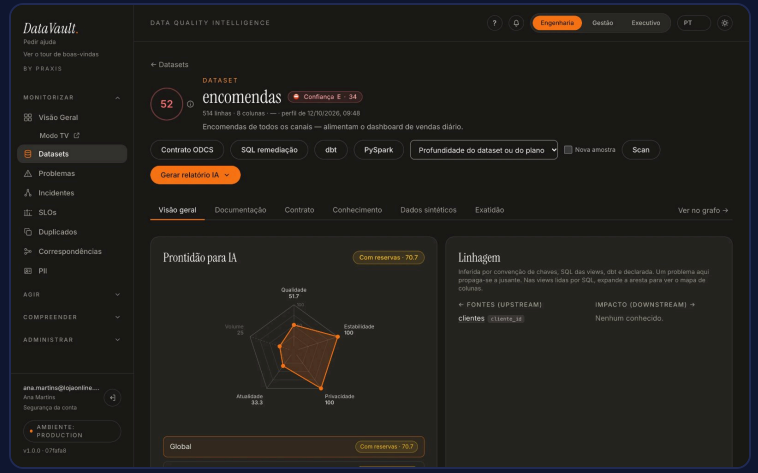
O que a equipa vê



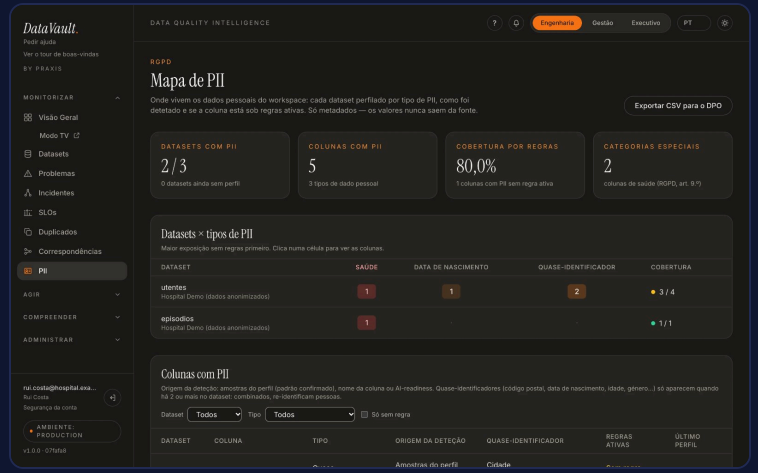
Scan ao vivo: regras falhadas, problemas novos e score por dataset



Incidentes: problemas agrupados pela linhagem, com a origem e o impacto



Detalhe do dataset: score por dimensão, prontidão para IA e linhagem



Mapa de dados pessoais do workspace, com categorias especiais (saúde)

Capturas da app em execução com os packs de demonstração (dados sintéticos).

Arquitetura de referência

PESSOAS E FERRAMENTAS

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

BERYX · GOOGLE CLOUD, EUROPE-WEST1

- API
- Workers assíncronos para scans e relatórios, com filas e DLQ
- Postgres com os metadados, nunca as tabelas
- Fachada de IA com router por tarefa
- Servidor MCP com 21 ferramentas só de leitura

AS TUAS FONTES

- Ligações só de leitura pelos métodos oficiais de cada fornecedor

> Das fontes saem agregados, até 20 valores de exemplo por coluna e, para as regras de linha, uma amostra de até 5 000 linhas por tabela.

- SaaS partilhado na UE (europe-west1) — EXISTE HOJE
- Instância dedicada por cliente, com IP de saída próprio e CMEK opcional — POR ATIVAR
- Execução no perímetro do cliente — VISÃO

Subcontratantes e regiões publicados na página de segurança, com 30 dias de aviso antes de qualquer mudança.

IA com regras: onde entra, onde não entra e quem decide

Onde a IA entra

- Sugestões de regras, causa provável de incidentes e relatórios por perfil — EXISTE HOJE
- Assistente em linguagem natural sobre o catálogo e os problemas — EXISTE HOJE
- Glossário e documentação de datasets, sempre para aprovação humana — EXISTE HOJE

Onde não entra

- Nenhum número de qualidade, de anomalia ou de custo vem de um LLM: os métodos são determinísticos e publicados — EXISTE HOJE
- Propor nunca executa: cada ação pede aprovação humana e fica auditada — EXISTE HOJE

Controlos do cliente

- IA desligada em qualquer plano, com relatórios em modo offline — EXISTE HOJE
- Amostras em bruto ou sintéticas, à escolha do workspace — EXISTE HOJE
- Chave Anthropic própria (Business) ou Claude no Amazon Bedrock ou no Vertex AI, em regiões europeias (Enterprise) — EXISTE HOJE

Guardas

- Teto diário de custo de IA e quota mensal de tokens por plano — EXISTE HOJE
- Router por tarefa: por omissão, só a Anthropic recebe dados de clientes — EXISTE HOJE
- Guarda de saída que mascara segredos, emails, IBAN, cartões e NIF antes de gravar (ainda não cobre o assistente em stream) — PARCIAL
- Defesa contra injeção indireta de instruções, com testes adversariais no CI — EXISTE HOJE
- Marcação do conteúdo gerado por IA, como pede o art. 50.º do Regulamento da IA — EXISTE HOJE

POR ATIVAR — Por omissão, amostras limitadas seguem para a Anthropic, nos EUA, ao abrigo de cláusulas contratuais-tipo. O anonimizador (Presidio) está feito, mas ainda não está na imagem de produção.

RGPD: do inventário manual ao mapa que se atualiza a cada scan

Saber onde estão os dados pessoais

- Mapa de dados pessoais: que datasets guardam que tipos de dados, como foram detetados e se têm regras ativas — EXISTE HOJE
- Sensibilidade por coluna em quatro níveis, com herança pela linhagem (sugerida) e curadoria com autor e data — EXISTE HOJE
- Risco de reidentificação por quase-identificadores (k-anonimato); os exemplos enviados à IA saem generalizados — EXISTE HOJE

Mantê-los exatos

- Dígitos de controlo de NIF, IBAN e cartões, e listas de referência — EXISTE HOJE
- Exatidão contra VIES, GeoAPI.pt, planos de numeração e registos MX, por opção — EXISTE HOJE
- Auditoria de exatidão por amostragem humana, com intervalo de confiança — EXISTE HOJE

Responder e provar

- Registo de pedidos de titulares com o prazo legal de cada jurisdição: UE, Reino Unido, Suíça, Brasil, Califórnia e Angola — POR ATIVAR
- Procurar e apagar o valor de um titular nas amostras guardadas pelo BeryX — POR ATIVAR
- Retenção das amostras configurável por workspace e trilho de auditoria exportável para SIEM — EXISTE HOJE
- Varredura de credenciais esquecidas em código e notebooks — EXISTE HOJE

Os casos da CNPD, em perguntas que os dados respondem

Onde estão os dados de saúde, de religião ou financeiros?

Mapa de dados pessoais e sensibilidade por coluna. EXISTE HOJE

Quem tem acesso, e bate com quem devia ter?

Uma regra de integridade referencial entre o diretório de contas e o registo de pessoal. EXISTE HOJE

Para onde vão os dados?

Linhagem entre as fontes ligadas. Envios por email ou para terceiros ficam fora. PARCIAL

FinOps de dados: o custo lido da própria plataforma e cruzado com a qualidade

Custo e qualidade lado a lado (Business)

- Telemetria nativa: BigQuery (INFORMATION_SCHEMA.JOBS), Snowflake (WAREHOUSE_METERING_HISTORY) e Databricks (system.billing.usage) — EXISTE HOJE
- Quanto do compute do mês correu sobre dados que falham a medição — EXISTE HOJE
- Custo por tabela no BigQuery; na Snowflake e no Databricks o custo é da conta, e a repartição aparece rotulada como atribuição. Sem telemetria, não aparece número nenhum — EXISTE HOJE

Picos de custo

- Um pico diário abre um problema e alerta nos canais; no BigQuery, com as 5 queries mais caras do dia e os literais mascarados — EXISTE HOJE
- Critério publicado: z ≥ 3, mais 50% e mais 1 USD sobre a mediana, com 7 dias de histórico — EXISTE HOJE

Uso real

- Tabelas quentes (20 ou mais queries por dia), ativas, frias e sem uso (0 queries em 90 dias, só com telemetria que cubra pelo menos 80 dias) — EXISTE HOJE
- As tabelas sem uso ficam candidatas a descomissionamento — EXISTE HOJE
- Scans guiados pelo uso: as quentes em profundidade a cada scan, as sem uso só no perfil mensal, por opção e com aprovação do steward — EXISTE HOJE

O custo do próprio BeryX

- Teto de bytes antes de ler no BigQuery — EXISTE HOJE
- Custo de IA por relatório, por tarefa e por fornecedor; relatórios agendados pela API de lotes a metade do preço — EXISTE HOJE
- Chargeback do uso por workspace, domínio e equipa, com CSV mensal e feed para o Power BI — EXISTE HOJE

PARCIAL — A telemetria de custo está feita e testada, mas ainda não foi provada numa conta real de BigQuery, Snowflake ou Databricks. Mede o custo de compute, e não a eficiência do formato dos dados.

Fontes: Métodos do BeryX, anomalias de custo

Segurança: só leitura imposto no código, e as lacunas ditas por escrito

Acesso às fontes

- Só leitura no código, testado contra servidores reais — EXISTE HOJE
- Métodos oficiais de cada fornecedor e credenciais cifradas (Fernet) com rotação de chave — EXISTE HOJE
- Anti-SSRF com DNS fixado, também nos webhooks e nos tickets — EXISTE HOJE

Isolamento e identidade

- Um teste percorre todas as rotas da API à procura de fugas entre clientes — EXISTE HOJE
- SSO por OIDC e SAML 2.0 em todos os planos; MFA obrigatória para administradores da organização — EXISTE HOJE
- Lista de IPs permitidos por organização, também para tokens de API — EXISTE HOJE

Rasto e controlo

- Auditoria gravada na mesma transação da mudança, com exportação para SIEM — EXISTE HOJE
- Acesso de suporte da BeryxTech só com consentimento, só de leitura e por 72 horas no máximo — EXISTE HOJE
- Exportação dos dados do workspace e apagamento em cascata com certificado assinado — EXISTE HOJE

O que ainda não temos

- × Ainda não temos ISO 27001 nem SOC 2: estão em programa, sem datas.
- × Ainda não fizemos um pentest externo.
- × WAF e Cloud Armor estão escritos, mas ainda não ativos em produção.
- × Ainda sem SLA contratual de disponibilidade e de suporte.
- × Os modelos de contrato (DPA e MSA) ainda não foram revistos por um advogado.

Ecossistema: 30 tipos de fonte e as ferramentas onde a equipa já trabalha

BASES DE DADOS 7

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

WAREHOUSES 4

BigQuery · Snowflake · Databricks · Microsoft Fabric *

FICHEIROS E OBJECT STORAGE 6

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

APLICAÇÕES DE NEGÓCIO 5

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

DOCUMENTOS 4

Confluence · Jira · SharePoint · Google Drive

STREAMING 2

Kafka · Pub/Sub

COLABORAÇÃO 2

Slack · Teams

* Código feito; o driver ainda não está na imagem do SaaS.

- **Catálogos** · OpenMetadata e DataHub; Collibra e Purview validados só com transporte simulado — PARCIAL
- **BI** · Power BI, Tableau e Looker — EXISTE HOJE
- **Tickets** · Jira, Linear, ServiceNow, Azure DevOps, GitHub e GitLab — EXISTE HOJE
- **Orquestração** · dbt Cloud, dbt Core, Airflow, Fivetran e Airbyte — EXISTE HOJE
- **Agentes de IA** · Servidor MCP publicado, com 21 ferramentas só de leitura — EXISTE HOJE
- **Pacotes para programadores** · SDK Python, provider Terraform, extensão VS Code, Airflow e plugin Claude Code, por publicar — POR ATIVAR
- **SIEM** · Splunk, Microsoft Sentinel e Google SecOps — EXISTE HOJE

Prova por conector (matriz de 10/10/2026): 7 contra fontes públicas reais, 2 contra serviços reais em Docker, 13 com transporte simulado, 6 sem driver no ambiente de prova e 1 por verificar.

Cada número diz como foi calculado

Três tipos de número

exato, estimado ou por amostra, marcados na app

EXISTE HOJE

Intervalos de confiança

a 95%, de Wilson com correção para população finita; z de Fisher nas correlações

EXISTE HOJE

Score com pesos fixos

completude 25%, unicidade 25%, validade 20%, atualidade 15% e consistência 15%

EXISTE HOJE

Anomalias

z-score modificado (mediana e MAD) com controlo de falsos positivos de Benjamini-Hochberg

EXISTE HOJE

Impacto em euros

heurística declarada, sempre estimada, que tende a subestimar

EXISTE HOJE

Métodos públicos

o documento está publicado e um teste falha se uma constante documentada mudar no código

EXISTE HOJE

> Nenhum método de cálculo usa um LLM.

Fontes: Métodos do BeryX

Seis perfis, cada um com o seu guia e o seu tour

Administrador do workspace

Liga fontes, gere a equipa, os canais de alerta e a política de IA.

● EXISTE HOJE

Guia do perfil →

Membro

Trabalha problemas, regras, sugestões e incidentes.

● EXISTE HOJE

Guia do perfil →

Leitor

Consulta scores, relatórios e catálogo sem alterar nada. Ainda não se atribui na página Equipa.

◐ PARCIAL

Guia do perfil →

Responsável de domínio

Decide sobre os datasets, os incidentes e as regras do seu domínio.

● EXISTE HOJE

Guia do perfil →

Administrador da organização

Governação de todos os workspaces: SSO, IPs permitidos, política de dados e chargeback.

● EXISTE HOJE

Guia do perfil →

Faturação

Plano, faturas e relatório de custos por equipa.

● EXISTE HOJE

Guia do perfil →

No primeiro início de sessão, um tour por perfil mostra a app. Os guias estão publicados nos docs do BeryX.

O que muda para cada função

FUNÇÃO	HOJE	COM O BERYX
Engenharia de dados	Descobre as falhas pelo negócio; testes espalhados por repositórios	Monitores automáticos, incidentes com linhagem, regras versionadas e gate no CI
DPO e jurídico	Inventário manual, revisto uma vez por ano	Mapa de dados pessoais a cada scan, sensibilidade com autor e evidência exportável
FinOps e direção financeira	A fatura no fim do mês, sem ligação à qualidade	Custo diário por dataset ou por conta, picos de custo com as queries mais caras e tabelas sem uso
Direção e CDO	Relatórios sem fonte nem tendência	Painel com metas e semáforos e relatórios por perfil
Equipas de IA	Dados de treino sem medição	Prontidão por finalidade, dossiê do art. 10.º e dados sintéticos com avaliação de privacidade

Panorama: onde cada abordagem é mais forte

Suites enterprise · Collibra, Informatica

- Catálogo e glossário como sistema de registo da empresa
- ISO 27001 e SOC 2, com trust center público
- Limpeza e MDM que escrevem nos teus sistemas, com credenciais de escrita

CUSTO Preço público no UK G-Cloud: 74 999 £ a 122 919 £ por ano, mais implementação

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

Open source · Great Expectations, Soda

- Grátis e a correr no teu perímetro
- SQL e Python sem limites nos testes
- Cada teste é escrito e mantido pela tua equipa

CUSTO Licença 0; Soda Team a 750 $ por mês, mais compute e horas de engenharia

Great Expectations — planos e preços · Soda — planos e preços

Observabilidade · Monte Carlo, Sifflet

- Monitores por metadados maduros em warehouses cloud
- SOC 2 Type II e ISO 27001
- Detetam e alertam; a correção fica com a equipa

CUSTO Contratos anuais; Sifflet a cerca de 48 k$ por ano no preço publicado

Monte Carlo — documentação: segurança e conformidade · Monte Carlo — Trust Center · Sifflet — planos e preços

BeryX

- Qualidade, dados pessoais e custo medidos no mesmo catálogo
- Propõe a correção em SQL ou como pull request, sem credenciais de escrita
- Ainda sem certificações nem SLA contratual

CUSTO Gratuito · desde 490 €/mês

Matriz completa, com fonte datada em cada célula →

Resumo da matriz comparativa do site, com uma fonte pública datada em cada célula (verificadas a 28/09/2026).

Onde está a diferença, e o que não afirmamos

O que se pode testar hoje

- × Qualidade, dados pessoais e custo medidos sobre o mesmo catálogo e o mesmo histórico
- × Só leitura imposto no código, com correções como propostas e pull requests
- × Cada número com método publicado e intervalo de confiança
- × IA com regras de saída, modo offline e a opção de a desligar
- × Preço publicado e conta gratuita sem chamada comercial

O que não afirmamos

- ✓ Que o BeryX evitaria uma coima concreta
- ✓ Certificações: ainda não temos ISO 27001 nem SOC 2
- ✓ Um SLA contratual
- ✓ Que todos os conectores foram provados numa conta real
- ✓ Que substitui um MDM ou um catálogo enterprise
- ✓ Que corrige os dados sozinho

A prova que propomos é um piloto, com métricas medidas antes e depois.

O que passa a ser mensurável

MÉTRICA	COMO O BERYX A MEDE	ESTADO
Score de qualidade por dataset e por domínio, com tendência	Cinco dimensões com pesos fixos, a cada scan	EXISTE HOJE
Cumprimento do SLO por dataset	Objetivo de serviço por dataset e histórico de cumprimento	EXISTE HOJE
Problemas abertos e resolvidos por severidade	Problemas e incidentes com datas, dono e prazo de resposta	EXISTE HOJE
Colunas com dados pessoais sem regras ou sem dono	Mapa de dados pessoais cruzado com as regras e os donos	EXISTE HOJE
Prazo de cada pedido de titular	Registo de pedidos com o prazo da jurisdição	POR ATIVAR
Compute gasto em dados que falham a medição	Telemetria de custo da plataforma cruzada com o score	EXISTE HOJE
Tabelas sem uso há 90 dias	Contagem de queries por tabela nas vistas de sistema	EXISTE HOJE
Custo de IA por relatório e por equipa	Custo por chamada, por tarefa e chargeback por equipa	EXISTE HOJE

Os valores de partida saem do primeiro mês de medição; não usamos médias de mercado.

O livro honesto

● EXISTE HOJE

- Perfil, 16 tipos de regra, monitores automáticos, anomalias e SLOs
- Incidentes com linhagem, causa provável e tickets
- Correção proposta, ensaiada e por pull request (Business)
- Mapa de dados pessoais, sensibilidade e risco de reidentificação
- Custo e qualidade lado a lado, picos de custo, uso real e chargeback
- Relatórios por perfil, assistente e servidor MCP
- SSO por OIDC e SAML, MFA, IPs permitidos, auditoria e SIEM
- Tour e guia escrito por perfil

◐ PARCIAL

- Conectores ainda sem prova numa conta real, entre eles BigQuery, Snowflake, Databricks e Oracle
- Collibra e Purview validados só com transporte simulado
- Páginas de metadados e dossiê de elementos críticos (BCBS 239) só por API
- Correspondência de entidades sobre amostras até 50 000 linhas
- Guarda de saída da IA ainda fora do assistente em stream
- Papel Leitor ainda por atribuir na página Equipa

◐ POR ATIVAR

- Pedidos de titulares e exportação para SodaCL e Great Expectations: feitos, no próximo deploy
- Sete conectores fora da imagem do SaaS
- Anonimizador na imagem de produção
- Instância dedicada, alta disponibilidade, WAF e IP de saída fixo
- SDK, Terraform, VS Code e Airflow por publicar
- Pagamento online

○ VISÃO

- Certificações ISO 27001, SOC 2, ISO 27701 e ISO 42001: ainda não temos nenhuma
- Pentest externo
- SCIM e chaves de segurança (WebAuthn)
- Conectividade privada e execução no perímetro do cliente
- Sincronização com o ServiceNow nos dois sentidos

Planos

Gratuito

0 € /mês

- 1 fonte
- 25 datasets
- 5 relatórios de IA por mês
- 30 dias de histórico

Team

490 € /mês

- 5 fontes
- 500 datasets
- 200 relatórios de IA por mês
- 365 dias de histórico
- ✓ Tokens de API
- ✓ Webhooks de alertas
- ✓ Análise de documentos para RAG
- ✓ Relatórios agendados por email
- ✓ Dados sintéticos a partir do perfil

Business

1190 € /mês

- 15 fontes
- 2000 datasets
- 1000 relatórios de IA por mês
- 730 dias de histórico
- ✓ Dossiê de conformidade do AI Act
- ✓ Telemetria de custo das plataformas
- ✓ Remediação como Pull Request (GitHub/GitLab)
- ✓ Chave Anthropic própria (BYOK)
- ✓ Escolha de modelos de IA por tarefa

Enterprise

Sob consulta

- sem limite
- ✓ Claude na conta cloud da organização (Amazon Bedrock ou Google Vertex AI)

Em todos os planos: motor de qualidade, sugestões, modo offline, IA desligável e SSO.

Preços sem IVA, com faturação anual; a faturação mensal custa mais 20%. Enquanto o pagamento online não está ativo, a subscrição faz-se com a BeryxTech.

Como começar

01

Experimentar sozinho

Conta gratuita com uma fonte e a demo de um clique, com quatro packs de dados sintéticos: Loja online, ERP, Neobanco e Hospital.

Criar conta gratuita →

02

Conversa com a BeryxTech.

Trinta minutos para preencher o quadro de custos com os teus números e ver a demo sobre o teu caso.

Falar connosco →

03

Piloto num domínio

Um domínio de dados, as métricas do slide anterior medidas no início e no fim, e uma decisão tomada com números.

Propor um piloto →

Dados em que a engenharia, o DPO e a direção financeira confiam, com a mesma prova.

BeryX, um produto da BeryxTech.

beryxtech.com/apresentacao