



DataXpertIA

Dados & Inteligência Artificial

APRESENTAÇÃO ESTRATÉGICA

OpenSearch

Busca, Observabilidade e IA

Busca de classe enterprise, análise de dados em tempo real e IA generativa

SOBERANIA TECNOLÓGICA

CÓDIGO ABERTO

ZERO VENDOR LOCK-IN



O PROBLEMA

A Instituição Opera no Escuro



A Dor

- Volume de dados não estruturados cresce 40% ao ano (IDC, 2024); relatórios defasados e silos impedem decisões ágeis
- Logs, métricas e traces dispersos → MTTR em dias, não minutos (Gartner, 2024)
- 80% do dado corporativo não gera valor por falta de integração e governança (Forrester Research)



O Risco

- LGPD: multas de até 2% do faturamento, limitadas a R\$ 50 milhões (Lei 13.709/2018, art. 52)
- Cibersegurança: incidentes levam, em média, 258 dias para serem identificados e contidos (IBM Cost of a Data Breach Report 2024)
- Oportunidade perdida: decisões baseadas em dados geram economia de 15–30% (TCU, Estudos Setoriais)

80%

dados ociosos (Forrester)

40%

crescimento anual (IDC)

258

dias para identificar e conter incidentes (IBM)

R\$ 50M

limite de multa (Lei 13.709)



A SOLUÇÃO

OpenSearch: Busca e Observabilidade para a Instituição Moderna

01



CONFORMIDADE

LGPD, INDACT e auditoria by design. Apache 2.0 garante transparência total.

02



ECONOMIA

TCO previsível, zero licenças e redução de até 90% no custo de armazenamento com tiered storage.

03



GOVERNANÇA

Projeto independente na Linux Foundation, com governança comunitária e auditoria aberta.

04



SEGURANÇA

RBAC, field-level security, SIEM open-source, TLS nativo — sem custo extra.

05



IA GENERATIVA

RAG, busca vetorial, IA agêntica e embeddings multimodais na plataforma de dados da instituição.

CAPACIDADES

Uma Plataforma Completa — Busca, Observabilidade e IA Nativas

CAPACIDADE	DESCRIÇÃO	DIFERENCIAL
Full-text Search	Busca textual avançada com relevance tuning (Lucene nativo)	Classe enterprise
Vector Search	Embeddings de até 16.000 dimensões (FAISS / NMSLIB)	Ideal para RAG e IA
SQL / PPL	Linguagens declarativas unificadas no Discover	SQL familiar para o time
Observabilidade	OpenTelemetry nativo: logs, métricas, traces em um só lugar	Vendor-neutral
Security Analytics	SIEM open-source, correlation zones, trace analytics	Sem custo extra
IA Generativa	RAG, IA agêntica e embeddings multimodais (OpenSearch 3.2)	Nativo e gratuito



A TECNOLOGIA

Suite Open-Source de Busca, Analytics e Observabilidade

- Nascido do fork do Elasticsearch 7.10.2; hoje projeto independente da Linux Foundation (setembro de 2024)
- Licença Apache 2.0 — use, modifique, audite sem pagar licenças ou royalties
- Comunidade global, ativa e crescente — uma das maiores do ecossistema open-source
- Suite completa: Search Engine + OpenSearch Dashboards + Data Prepper + ML Commons + Security Analytics

SOBERANIA DIGITAL

Código aberto, governança comunitária e auditoria total. Zero risco de mudança unilateral de licenciamento.



PROVA & ROI

Comunidade Forte e TCO Previsível



Prova Social

- #17 entre todos os projetos da Linux Foundation em atividade de contribuidores
- 3.300+ contribuidores e 400+ organizações ativas; 1.400+ contribuidores únicos no acumulado do projeto
- AWS, Uber, instituições acadêmicas e órgãos de transparência pública adotam e contribuem



ROI / TCO

- Licenciamento zero: Apache 2.0 elimina royalties, assinaturas e surpresas de faturamento
- Tiered storage: redução de até 90% no custo de armazenamento, mantendo a performance
- Economia potencial de R\$ 700 mil a R\$ 3 milhões/ano em licenças proprietárias equivalentes — a dimensionar em diagnóstico prévio

PROVA

Caso de Sucesso: Uber — Busca Vetorial em Escala de Bilhões

- Uber executou prova de escala de busca vetorial sobre OpenSearch com 1,5 bilhão de itens indexados (~400 dimensões, engine FAISS)
- Consultas de similaridade em larga escala com latência de milissegundos, viabilizadas pelos engines FAISS/NMSLIB
- Benchmark Trail of Bits (março de 2025): OpenSearch 1.6x mais rápido que o Elasticsearch 8.15 em cargas de trabalho reais de logs
- No mesmo benchmark, os engines FAISS/NMSLIB foram 11–14% mais rápidos que o Lucene do Elasticsearch em busca vetorial de alta dimensionalidade
- Referências públicas: uber.com/blog (Powering Billion-Scale Vector Search with OpenSearch) · blog.trailofbits.com (março de 2025)

1.6x

vs. Elasticsearch em logs reais

11–14%

mais rápido em busca vetorial

1,5 bilhão

de itens indexados (Uber)



PROVA

Casos de Uso de Sucesso do OpenSearch

Grandes empresas utilizam o OpenSearch para busca em tempo real, análise de logs (observabilidade) e busca vetorial/IA — em nuvem gerenciada ou infraestrutura própria, sempre com o mesmo código 100% aberto



Adobe

Acrobat AI Assistant

CASO DE USO

Escalar o assistente de IA do Acrobat para atender centenas de milhões de usuários lidando com trilhões de arquivos PDF.

SOLUÇÃO

Amazon OpenSearch Service com busca vetorial (vector search) para recuperação contextual e citação de fontes (attribution citations) combinadas a modelos de linguagem.

Fonte: AWS Customer Case Studies — Adobe



DoorDash

Atendimento Automatizado

CASO DE USO

Atendimento automatizado em escala para centenas de milhares de entregadores e clientes por dia.

SOLUÇÃO

Amazon OpenSearch Serverless como banco de dados vetorial integrado a LLMs, alimentando a base de conhecimento e respondendo chamadas via IA generativa em tempo real.

Fonte: AWS Customer Case Studies — DoorDash



Prime Video

Amazon — Streaming

CASO DE USO

Descoberta de conteúdo e relevância de busca para centenas de milhões de assinantes durante grandes eventos de transmissão ao vivo.

SOLUÇÃO

Busca semântica orientada por modelos de embedding customizados com OpenSearch, melhorando a precisão da busca por catálogo e esportes.

Fonte: AWS Customer Case Studies — Prime Video



OBJEÇÕES

“E se...?” — Antecipando as Principais Dúvidas



Objecção: Performance

“OpenSearch é tão rápido quanto soluções proprietárias?”

- Otimizações de segment replication, query caching e remote storage reduziram latência em ~90% nas versões recentes
- O próprio Uber validou performance em escala de bilhões de itens; tiered storage mantém a performance reduzindo o custo de armazenamento



Objecção: Suporte Enterprise

“Quem garante o suporte se algo der errado?”

- Múltiplas camadas: AWS, Aiven, Instaclustr para hospedagem gerenciada; comunidade global ativa; consultorias especializadas em software livre enterprise
- A instituição nunca fica refém de um único fornecedor



QUEM SOMOS

DataXpertIA — Dados, eXpertise e IA



Nossa Proposta de Valor

- Dados como Ativo Estratégico — Engenharia de dados, DataOps e pipelines modernos (Airflow, Spark, Iceberg)
- Soberania Tecnológica — Stack 100% open-source enterprise, livre de vendor lock-in, com total auditoria
- IA Nativa — Assistentes de dados, análise preditiva, detecção de anomalias, geração automática de relatórios e RAG
- Expertise Governamental — Experiência com órgãos públicos, licitações, compliance (LGPD, INDACT)
- Custo Previsível + Treinamento — Modelo transparente sem surpresas de licenciamento; mentoria para autonomia do time

NOSSA VISÃO

Toda organização — pública ou privada — pode ter uma plataforma de dados moderna, governada e pronta para IA, sem abrir mão do controle sobre arquitetura e custos.

NOSSO COMPROMISSO

Entregar transformação inteligente de dados, com soberania tecnológica, custo previsível e decisões aceleradas por IA.

PRÓXIMO PASSO

Agende uma avaliação técnica gratuita — mapeamos o cenário atual e propomos uma arquitetura inicial em poucos dias.
contato@dataxpertia.com · +55 61 98657-1054 · dataxpertia.com