

Davi Specia

📍 Toledo, PR, Brasil ✉️ davi.specia@gmail.com ☎️ (45) 99128-5227 🌐 davispecia.com.br 📱 davispecia
🔗 davidogral

Resumo

Construo pipelines de dados que alimentam decisões de negócio. Sou sócio-fundador da Alugaix, onde integrei por API os sistemas de gestão, precificação e venda de uma operação com 60 imóveis e montei a camada analítica usada no fechamento mensal com os proprietários. Em projeto próprio, coloquei no Databricks um pipeline de análise de crédito com Delta Lake, Unity Catalog, carga incremental com MERGE e quality gate. Procuro uma vaga de engenharia de dados em uma equipe onde eu seja liderado e ganhe senioridade com gente mais experiente do que eu. Python e SQL.

Experiência

Alugaix, Sócio-fundador e analista de dados

Toledo, PR

Mar 2024 – presente

- Responsável pelos dados de uma operação com 60 imóveis sob gestão e faturamento anual acima de R\$ 1,5 milhão. Integro por API o PMS (Hostify), a precificação (Price-Labs) e os canais de venda (Airbnb, Booking).
- Construí a camada analítica da operação, com ingestão via API, padronização e modelagem dos dados e cálculo de ocupação, ADR e RevPAR. Esses números são a base do fechamento mensal com os proprietários.
- Automatizei a criação de anúncios com o Compacta. Um modelo de visão classifica as fotos por cômodo, LLMs extraem os dados dos documentos e escrevem os textos, e o anúncio é publicado direto na Hostify pela API. Cada anúncio leva pelo menos 5 horas a menos para ficar pronto. Em produção com FastAPI e Docker. (sistema interno)
- Como sócio-fundador, trabalho com a operação para transformar necessidades do negócio em regras de cálculo, arquitetura técnica e roadmap.

Projetos

Pipeline de análise de crédito (Ver projeto)

- Arquitetura medalhão rodando no Databricks como job serverless de 6 tarefas, com tabelas Delta no Unity Catalog e infraestrutura como código (Databricks Asset Bundles).
- Carga incremental. A Bronze guarda o histórico de ingestões e a Silver é atualizada com MERGE INTO por cliente. Um quality gate com 20 regras bloqueia a publicação, e cada execução fica registrada numa tabela de auditoria.
- Otimização em PySpark. As 20 regras rodam em uma única passada sobre os dados, a Gold é cacheada porque alimenta três agregações e as partições de shuffle foram ajustadas ao volume.
- A mesma lógica roda localmente com Parquet, PostgreSQL (carga transacional com chaves primárias e estrangeiras) e Airflow, com mais de 40 testes automatizados e CI no GitHub Actions.

Cinerd, busca de filmes por descrição (Ver projeto)

- Pipeline de ingestão que junta a API do TMDB, a base pública do IMDb e o enredo da Wikipédia em um catálogo de 22.029 títulos, com dados e índices versionados em DVC.
- Deploy próprio na Oracle Cloud, com CI (ruff, mypy e pytest), avaliação automatizada que barra regressão de qualidade e métricas por etapa no formato Prometheus.
- Decisões de custo com dado medido. Testei a quantização INT8 com ONNX Runtime, que reduziu a memória de 2,3 para 1,7 GB quase sem perda de qualidade, e tirei um cross-encoder da produção porque o ganho ficou dentro do ruído e a latência aumentou cerca de 40 vezes.

Rede neural guiada por física para Itaipu (Ver projeto)

- Série diária do ONS de 2015 a 2024, com 2023 e 2024 reservados para teste e validação móvel de 2020 a 2024. Rede implementada do zero em Rust, sem bibliotecas de machine learning.
- O cenário operacional chegou a RMSE de 508 MWmed (R^2 de 0,91), contra 590 da persistência, e venceu a persistência em todos os anos da validação.
- Os resultados se reproduzem a partir de um clone limpo, com CI rodando clippy, testes e checagem SHA-256 dos dados. Artigo aprovado no Latin.Science 2026 (Latinoware).

Publicações

Estimativa de peso de suínos a partir de imagens RGB utilizando redes neurais convolucionais e o dataset PigRGB-Weight Ago 2026

Vinicius Tessele, Jhonatan da Silva Margraf, Vinícius Hillebrand Carraro, Marco Antonio Wandraski, Luiz Henrique Zavatini Feltrin, Davi Specia
RECIMA21, v. 7, n. 8, DOI [10.47820/recima21.v7i8.8496](https://doi.org/10.47820/recima21.v7i8.8496)
recima21.com.br/recima21/pt_BR/article/view/8496

Modelagem da Geração Hidrelétrica de Itaipu via Redes Neurais Guiadas por Física em Rust: Validação Temporal, Persistência e Previsão Operacional 2026

Davi Specia, Luiz Henrique Zavatini Feltrin, Tainá Dreissig, Miguel Mantoan Castellani, Vinícius Carraro
Latin.Science 2026, Latinoware, artigo aprovado · [Ler artigo no Drive](#)

Formação

Faculdade Donaduzzi (Biopark), Bacharelado em Inteligência Artificial Toledo, PR
• Os primeiros 2,5 anos seguem a mesma grade do curso de Ciência de Dados, o que Fev 2025 – Dez 2029
permite obter também esse diploma.

IFPR, Campus Toledo, Técnico em Informática para Web integrado ao Ensino Médio Toledo, PR
Fev 2022 – Dez 2025

Competências

- Linguagens:** Python, SQL, Rust
- Engenharia de dados:** Databricks (jobs serverless, Asset Bundles, Unity Catalog), Delta Lake, PySpark, Spark SQL, Airflow, PostgreSQL, arquitetura medalhão, ETL/ELT, modelagem de dados, integração via API
- Qualidade e versionamento:** quality gate, pytest, DVC, checagem de integridade (SHA-256), avaliação automatizada em CI
- Infraestrutura:** Docker, GitHub Actions, Oracle Cloud, Prometheus, Grafana
- IA:** PyTorch, scikit-learn, embeddings, LLMs
- Visualização:** Power BI, Streamlit

Idiomas

- Português:** nativo
- Inglês:** básico, leitura de documentação técnica