

Pedro Henrique

 [LinkedIn](#) |  pedrohentec@gmail.com |  [GitHub](#)

Introdução

Com experiência prática em ambientes de computação em nuvem (GCP e AWS), focado no desenvolvimento, orquestração e manutenção de pipelines de dados escaláveis e eficientes. Tenho domínio em ferramentas como Apache Beam, BigQuery, Dataflow, Spark, Airflow e Docker, além de forte base em Python e SQL (PostgreSQL, MySQL, SQL Server). Minha atuação inclui desde a ingestão e transformação de dados em processos ETL/ELT, até a modelagem e disponibilização em Data Lakes e Data Warehouses com governança e segurança (LGPD). Trabalho com orquestração via Airflow, versionamento com Git e deploy de soluções orientadas à performance e escalabilidade. Tenho experiência com Dataplex, Looker Studio, Athena, EMR, S3 e práticas de DataOps, além de projetos próprios voltados à automação de tarefas e extração de valor a partir de dados.

Habilidades

- **Programação:** Python (Pandas, NumPy, Matplotlib), SQL (MySQL, PostgreSQL, SQL Server)
- **Gestão de Banco de Dados:** Modelagem, consultas otimizadas, e manutenção de bases em ambientes produtivos
- **Engenharia de Dados:** ETL/ELT, Spark, Airflow, Docker, Apache Beam
- **Computação em Nuvem:** AWS (S3, EMR, Athena, EKS), práticas de infraestrutura como código e gestão de dados em nuvem. GCP (BigQuery, Dataflow, Dataplex, Storage)
- **Visualização e Análise:** Power BI, Excel (Dashboards, PivotTables)

Experiência

Engenheiro de Dados Júnior

ONR • Registro de Imóveis Eletrônico

Remote, BR 01/2025 - Atual

Atuo no desenvolvimento e monitoramento de pipelines de dados em ambiente GCP, com foco em performance, governança e confiabilidade das informações. Utilizo Apache Beam e Dataflow para orquestração de dados em BigQuery, aplicando boas práticas de modularização e tratamento assíncrono, reduzindo gastos em ambiente cloud e tendo impacto semelhante ou melhorado.

- Criação de pipelines com Apache Beam (Python) em Dataflow, processando dados em lote e streaming.
- Validação automática de tabelas e partições no BigQuery, com geração de modelos dinâmicos a partir de arquivos JSON.
- Monitoramento de jobs via API (Dataflow e BigQuery), com foco em status, métricas e integridade dos dados.
- Governança e classificação de dados com Dataplex, alinhado à LGPD.
- Otimização e refatoração de pipelines legadas para maior eficiência e reusabilidade.

Engenheiro de Dados (Estágio)

A3Data

Remote, BR 07/2024 - 01/2025

Atuei como Engenheiro de Dados em projetos, em um ambiente AWS, com foco na construção e otimização de pipelines de dados em larga escala. Trabalhei em processos de ingestão, transformação e entrega de dados, contribuindo diretamente para a eficiência e governança das soluções.

- Desenvolvimento de pipelines com Spark, SQL e Python em ambiente AWS EMR, com orquestração por Airflow.
- Otimização de consultas e fluxos de dados utilizando Athena, com dados armazenados em S3, garantindo performance e custo-benefício.
- Uso de Docker e Kubernetes (EKS) para empacotamento e deploy de aplicações de dados.
- Suporte na implementação de boas práticas de versionamento com Git/GitHub e modularização de código para reuso e manutenção.
- Participação ativa na análise de requisitos de dados, contribuindo com insights técnicos para decisões estratégicas.
- Contato inicial com práticas de governança de dados, tratamento de exceções e validações automatizadas.

Pedro Henrique

 [LinkedIn](#) |  pedrohentec@gmail.com |  [GitHub](#)

Gestor de Tecnologia

Cartório Registro de Imóveis

Presencial, BR 02/2022 - 07/2024

Atuei como gestor do banco de dados e servidor, responsável pela qualidade técnica da tecnologia na serventia onde implementei inovações para segurança dos dados pessoas agindo de acordo com as regras da LGPD. Atendimento a dúvidas e resolução de problemas bem como o desenvolvimento de softwares para o bom funcionamento. SQL, Excel, Segurança da Informação e resolução de problemas.

Auxiliar de Tecnologia

ORTECONTE CONTABILIDADE

Presencial, BR 10/2019 - 02/2022

Monitoramento de servidor na filial, gestor em soluções para o desenvolvimento em tarefas em massa, analista de sistema do gestor de tarefas e execução. Utilização do Excel para monitoramento de atividades e análise das tarefas com base em Dashboards e qualificação de estrutura.

Projetos

- **To Text:** Responsável pela conversão de imagens datilografadas para arquivos de texto, contribuindo para seguir o prazo e automação. Utilizando a biblioteca tesseract para fazer o processo de OCR e outras complementares para o funcionamento como Matplotlib, Python e seguindo as regras da LGPD.

Formações

- Graduação em **Ciência de Dados**, Instituição UNOPAR, conclusão em 2026
- Curso Técnico em Informática

Certificações

- [Formação Python para Data Science](#): Python, NumPy, Pandas, Matplotlib, Seaborn, Pandas I/O.