

# Pedro Henrique

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

## Introduction

---

With practical experience in cloud computing environments (GCP and AWS), focused on developing, orchestrating and maintaining scalable and efficient data pipelines. I am proficient in tools such as Apache Beam, BigQuery, Dataflow, Spark, Airflow and Docker, as well as having a strong background in Python and SQL (PostgreSQL, MySQL, SQL Server). My work ranges from ingesting and transforming data in ETL/ELT processes, to modeling and making it available in Data Lakes and Data Warehouses with governance and security (LGPD). I work with orchestration via Airflow, versioning with Git and deploying performance and scalability-oriented solutions. I have experience with Dataplex, Looker Studio, Athena, EMR, S3 and DataOps practices, as well as my own projects aimed at automating tasks and extracting value from data.

## Skills

---

- **Programming:** Python (Pandas, NumPy, Matplotlib), SQL (MySQL, PostgreSQL, SQL Server)
- **Database Management:** Modeling, optimized queries, and database maintenance in productive environments
- **Data Engineering:** ETL/ELT, Spark, Airflow, Docker, Apache Beam
- **Cloud Computing:** AWS (S3, EMR, Athena, EKS), infrastructure-as-code practices and cloud data management. GCP (BigQuery, Dataflow, Dataplex, Storage)
- **Visualization and Analysis:** Power BI, Excel (Dashboards, PivotTables)

## Experience

---

### Junior Data Engineer

ONR • Electronic  
Property Registry

Remote, BR 01/2025 - Present

I work in the development and monitoring of data pipelines in a GCP environment, focusing on performance, governance, and reliability of information. I use Apache Beam and Dataflow for data orchestration in BigQuery, applying best practices of modularization and asynchronous processing, reducing costs in a cloud environment and having a similar or improved impact.

- Creation of pipelines with Apache Beam (Python) in Dataflow, processing data in batches and streaming.
- Automatic validation of tables and partitions in BigQuery, with generation of dynamic models from JSON files.
- Monitoring of jobs via API (Dataflow and BigQuery), focusing on status, metrics, and data integrity.
- Data governance and classification with Dataplex, aligned with LGPD (Brazilian General Data Protection Law).
- Optimization and refactoring of legacy pipelines for greater efficiency and reusability.

### Data Engineer (Internship)

A3Data

Remote, BR 07/2024 - current

I worked as a Data Engineer on projects in an AWS environment, focusing on building and optimizing large-scale data pipelines. I worked on data ingestion, transformation and delivery processes, contributing directly to the efficiency and governance of the solutions

- Pipeline development with Spark, SQL and Python in an AWS EMR environment, with Airflow orchestration.
- Optimization of queries and data flows using Athena, with data stored in S3, ensuring performance and cost-effectiveness.
- Use of Docker and Kubernetes (EKS) for packaging and deploying data applications.
- Support in implementing good versioning practices with Git/GitHub and code modularization for reuse and maintenance.
- Active participation in data requirements analysis, contributing technical insights to strategic decisions.
- Initial contact with data governance practices, exception handling and automated validations.

# Pedro Henrique

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

## Technology Manager

Cartório Registro de Imóveis

*In-person, BR* 02/2022 - 07/2024

I acted as database and server manager, responsible for the technical quality of the technology in the service where I implemented innovations for the security of people's data, acting in accordance with the LGPD rules. Answering questions and resolving problems, as well as developing software for smooth operation. SQL, Excel, Information Security and problem solving.

## Technology Assistant

ORTECONTE CONTABILIDADE

*In-person, BR* 10/2019 - 02/2022

Server monitoring at the branch office, solutions manager for mass task development, system analyst for task management and execution. Using Excel to monitor activities and analyze tasks based on dashboards and structure qualification.

## Projects

---

- **To Text:** Responsible for converting typewritten images into text files, helping to meet deadlines and automation. Using the tesseract library to carry out the OCR process and other complementary libraries such as Matplotlib, Python and following the LGPD rules.

## Education

---

- Degree in **Data Science**, UNOPAR Institution, completion in 2026
- Technical course in Computer Science

## Certifications

---

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