

Marcos Aurelio Wozhiak Junior de Oliveira

<https://www.linkedin.com/in/aureliowozhiak/>

aureliowozhiak@gmail.com

(41) 99910-9170

Senior Data Engineer with over 5 years of experience designing, building, and optimizing large-scale data pipelines and cloud-based architectures, building of end-to-end data processes, from ingestion and transformation to modeling, orchestration, and governance, across retail, SaaS, and consulting industries. Recent projects include developing a dbt and Snowflake based data warehouse, enabling consistent analytical layers for ML workloads, and automating data workflows with Airflow and Spark on AWS to reduce processing time by 30%. Beyond delivery, created a data engineering platform for aspiring data professionals, fostering technical growth in the data community.

SKILLS

Programming Languages: Scala, Python, Java, SQL

Databases & Storage: MongoDB, AWS Redshift, PostgreSQL, AWS S3, Google BigQuery, Oracle, Snowflake

Data Tools: Spark, Airflow, dbt, DataFlow, EMR, Glue, Athena

Cloud: AWS, GCP

Others: AWS Lambda, Terraform, Git, GitHub

EXPERIENCE

Senior Data Engineer, C&A Brasil

April 2025- Present

- Act as a Senior Data Engineer in a multi-cloud environment (AWS and GCP), designing, maintaining, and optimizing large-scale data pipelines and cloud-based data infrastructure for a global retail organization.
 - Developed and maintained integrations with external platforms such as Salesforce, Meta, and Oracle, ensuring reliable ingestion, data consistency, and traceability across analytical and operational use cases.
- Worked extensively with Teradata, Google BigQuery, and AWS services to support analytical workloads, data modeling, and reporting layers aligned with business requirements.
 - Led cost optimization initiatives on BigQuery and GCP, applying partitioning strategies, query optimization, and resource governance practices, resulting in reduced cloud spend and improved performance.
- Performed ongoing maintenance and optimization of data processes and infrastructure on AWS, ensuring high availability, reliability, and scalability of batch and near real-time pipelines.

- Collaborated closely with cross-functional teams in Agile environments, supporting data governance, access control, monitoring, and continuous improvement of the data platform.

Senior Data Engineer, Grupo Boticário

March 2024 – April 2025

- Designed and implemented transformation layers and data models using dbt and Snowflake, ensuring consistent and high-quality data for analytics and ML workloads for a beauty and cosmetics retailer.
 - Built and orchestrated scalable ETL pipelines for batch and near real-time processing using Apache Spark, Airflow, and AWS services such as S3, Glue, Lambda, EMR, and Athena, following Agile delivery practices.
- Led technical delivery initiatives and performed code reviews, establishing CI/CD workflows that standardized engineering practices and improved deployment reliability.
 - Defined and enforced best practices for data observability, access control, and pipeline monitoring, enhancing platform governance and maintainability.
- Guided the integration of new API-based data sources into the Data Lake, ensuring backward compatibility and traceability across the entire data lifecycle. Optimized Spark jobs through partitioning, caching, and resource tuning strategies, reducing runtime by 25%.
 - Mentored junior and mid-level engineers, promoting technical excellence and a culture of ownership within the team.

Senior Data Engineer, Rock Content

Jan 2023 – May 2024

- Automated ETL and data validation workflows using dbt and AWS Services, including Redshift, Glue and S3, improving operational efficiency and consistency across analytical layers for a No-code platform that creates high-performing interactive content.
 - Led technical delivery of scalable data models supporting marketing, finance, and product teams, ensuring full alignment with business KPIs.
- Refactored legacy data pipelines to modern CI/CD-enabled workflows, reducing maintenance overhead and improving job reliability. Designed and developed end-to-end ingestion pipelines within a Data Lake architecture on AWS, leveraging APIs, process scheduling, and partitioning strategies to ensure data quality and performance.
- Implemented access control and comprehensive data lineage documentation to strengthen governance, security, and auditability across all data assets. Collaborated in architectural decisions around schema evolution and storage formats, ensuring backward compatibility and platform scalability within Agile delivery cycles.

Mid-Level Data Engineer, Falconi

July 2022 – Jan 2023

- Designed and deployed cloud-native ETL solutions for ingesting and processing large-scale datasets using AWS and GCP services, supporting enterprise-grade analytics and reporting for a major retail company.
 - Developed real-time streaming pipelines with BigQuery and Apache Dataflow, enabling low-latency dashboards for operational decision-making.
- Implemented process scheduling, retry mechanisms, and CI/CD workflows to enhance pipeline reliability and ensure consistent delivery of curated datasets for data science teams. Enforced access control and compliance with privacy and security regulations for handling sensitive operational and financial data.
- Maintained metadata repositories and contributed to Data Lake governance for improved discoverability and lineage tracking. Collaborated in Agile sprints, delivering incremental improvements aligned with OKRs, and supported proactive monitoring and alerting dashboards to ensure high availability and resilience in data flows.

Mid-Level Data Engineer, Aegro

Sep 2021 – July 2022

- Developed and maintained ETL pipelines to ingest agricultural data into cloud-based Data Lakes, ensuring scalability, performance, and data consistency across multiple environments. Built API connectors and web scrapers for both real-time and batch ingestion from diverse providers, enhancing the company's data coverage and automation capabilities.
- Implemented process scheduling and logging using Apache Airflow, improving pipeline monitoring and reliability. Optimized job performance through efficient partitioning strategies and schema tuning, reducing data processing time and resource consumption.
- Integrated and synchronized data across relational and NoSQL databases, maintaining backward compatibility and alignment with evolving business logic. Supported cross-functional analytics teams with curated datasets, following Agile delivery principles and standardized technical delivery documentation to improve internal collaboration and knowledge sharing.

Junior Data Analyst, Ebanx

Oct 2020 – Sep 2021

- Built and maintained BI dashboards to support strategic and operational decision-making across customer operations. Processed and transformed large datasets to align with business logic, applying ETL best practices to ensure data accuracy and consistency for a payments solutions company.
- Collaborated with stakeholders to define KPIs, reporting requirements, and data integration flows from internal systems and CRM platforms. Contributed to the rollout of data governance and access control policies, improving data security and compliance.

- Delivered actionable insights through ad hoc analyses, helping optimize internal workflows and reduce response times. Operated within Agile delivery cycles, ensuring high-quality technical delivery of dashboards and analytical solutions to business teams.

Junior Operations Analyst, Ebanx

Jan 2020 – Oct 2020

- Automated internal workflows using Zendesk triggers, macros, and API integrations, reducing manual effort and response times across customer support operations. Developed and maintained a chatbot platform integrated with CRM and service tools, improving efficiency and client satisfaction.
- Configured and monitored service channels across multiple digital platforms, ensuring reliability and scalability of communication flows. Structured and cleaned CRM datasets to enhance reporting accuracy and decision-making.
- Collaborated cross-functionally with operations and product teams to deliver high-quality technical solutions and documentation, applying Agile methodologies to continuously improve internal processes and knowledge sharing.

Customer Operations Analyst, Ebanx

Jan 2019 – Jan 2020

- Escalated technical and financial issues across levels 2 and 3. Extracted and processed data to support team performance tracking. Handled real-time customer support inquiries across digital channels. Participated in an internal development program focused on operational metrics and dashboard creation.

EDUCATION & TRAINING

MBA, Data Analysis, Descomplica (2021)

- Coursework included Data Visualization, Statistics for Data Science, Predictive Modeling, Machine Learning, and Business Intelligence.
- Developed a final capstone project focused on optimizing ETL workflows and improving dashboard performance through automated SQL-based analytics.

BSc, Mathematics, Camoes (2018)

- Coursework included Linear Algebra, Calculus, Probability, Statistics, and Computational Modeling.
- The graduation project focused on applying mathematical modeling to real-world data analysis and trend prediction problems.