Leonardo Rodrigues

LinkedIn | +55 11 97773-3418 | lmorenogr@gmail.com | GitHub

- Senior Data Engineer with extensive experience designing and implementing robust data solutions for global organizations in consulting, banking, retail, and logistics.w``
- Skilled in data architecture, cloud migrations, process optimization, and cross-functional collaboration. Proven track record of delivering scalable pipelines, integrating enterprise systems, and enhancing performance and data reliability using technologies like Hadoop, Spark, Python, Scala, AWS, and Azure.
- Certified in DP-900 and DP-203, with a passion for innovation and translating data into strategic business value.

Skills

- **Programming Languages:** SQL, Go, Scala, Rust, Python, Bash
- **Databases & Storage:** HDFS, MS SQLServer, Hive, MySQL Server, AWS S3, Redshift
- **Data Tools:** DataFlow, Spark, Databricks, Azure Data Factory, AWS Athena, AWS Glue, Azure Synapse, Hadoop, PySpark, Parquet, CSV, Blob Storage, EMR, Step Function
- **Cloud:** AWS, Azure
- **Others:** Gitlab, Linux, Git, Shell Scripts, Lambda

Experience

Senior Data Engineer — Keyrus @ Pluxee — *São Paulo, SP. Brazil* — **10/2024 - Current**

- Migrated business-critical data pipelines from AWS to Azure for a global employee benefits platform using PySpark, Scala, Azure Data Factory, OOP, and SQL, ensuring zero downtime and enabling continuity for data science and BI teams.
- Optimized batch ETL pipelines handling financial datasets using PySpark and Bash, reducing execution time by 40% and ensuring SLA compliance during the cloud transition.
- Used SQL and Azure Blob Storage to redesign data models and workflows for financial reporting, improving data retrieval performance by 30%.
- Migrated datasets originally stored in AWS S3 and cataloged in AWS Glue, translating configurations and workflows to native Azure services.
- Delivered consistent cost and performance improvements in Azure, reducing infrastructure spend by 20% post-migration through optimization of data pipeline runtimes.

Senior Data Engineer — Keyrus @ Visa Inc. — *São Paulo, SP. Brazil* — **10/2023 - 10/2024**

- Engineered cross-region ETL pipelines using PySpark, Hive, and Azure Data Factory to process transactional and behavioral payment data, improving time-to-insight for fraud analysis in a financial services company.
- Spearheaded the creation of efficient, scalable data algorithms using PySpark to handle high-volume datasets, driving seamless data processing and transformation across global systems.
- Deployed CI/CD for data applications using GitLab, Python, and automated testing, shortening deployment cycles from weekly to daily.
- Led multi-region coordination with teams in Miami, India, and Mexico, supporting agile delivery and business continuity for critical finance workflows.
- Created SQL-based dashboards and analytics queries that surfaced user behavior metrics, helping stakeholders reduce payment anomalies by 18% and reducing delay in fraud detection by 25%.
- Assisted in pilot usage of AWS EMR and S3 for scalable data replication and backup workflows, providing technical input on architectural limitations and resource configurations.

Senior Data Engineer, Founder — Black Raven — *São Paulo, SP. Brazil* — **10/2023 - Current**

- Designed and implemented data lake architectures using Azure Data Lake, Spark, OOP, and Python for retail and logistics clients, improving data availability and reporting latency by 60%.
- Built scalable, secure ETL pipelines with Databricks, Hive, Python and Scala, reducing total cost of ownership (TCO) by eliminating manual batch processing.
- Integrated disparate systems via custom connectors and API ingestion in Python and Bash, streamlining client data flows and enabling real-time analytics capabilities.

- Applied data governance practices, including encryption, access control, and metadata documentation to ensure compliance with GDPR and industry standards in finance and healthcare sectors.
- Created PoCs for clients using AWS Redshift and S3-based staging areas, with ingestion pipelines leveraging AWS Glue and orchestration via Step Functions and Lambda (basic usage).

Mid level & Senior Data Engineer **Santander** *São Paulo, SP. Brazil* **12/2021 - 10/2023**

- Built and optimized ETL processes for internal private investment reporting using Azure Data Factory, PySpark, OOP and SQL Server, cutting report generation time by 50%.
- Led data model refactoring efforts using Parquet, Azure Blob Storage, and Python, reducing data duplication and improving storage efficiency.
- Promoted to Senior Engineer for driving architectural improvements and mentoring junior engineers within the investment reporting team.
- Enhanced CI/CD pipelines using GitLab CI, Bash, and Docker, increasing the reliability of production releases in a regulated financial environment.
- Maintained and monitored production workflows in Spark and Shell Script, ensuring consistent delivery of accurate investment performance reports.

Junior Data Engineer **Accenture** *São Paulo, SP. Brazil* **04/2021 - 12/2021**

- Designed, developed, and maintained ETL processes, automation workflows, and reporting systems for the financial operations of a telecommunications company, working extensively with Scala, Python, and SQL Server to build and optimize efficient data pipelines and ensure the smooth flow of financial data. Acquiring hands-on experience managing complex data workflows and delivering reliable and actionable insights for finance teams.
- Leveraged Spark and Hive for processing and querying large datasets to support analytical reporting and decision-making.
- Developed and maintained automation tools to streamline operations and improve data accuracy and reporting efficiency.
- Utilized Visual Studio Code as a primary development environment for building and deploying scalable solutions.

Education

BSc, Data Science **Anhembi Morumbi University** *São Paulo, SP. Brazil* **01/2020 - 12/2023**

- Gained solid foundation in data analysis, statistical modeling, and machine learning, with practical experience using Python, Java, Scala, OOP, SQL and cloud platforms to solve real-world business problems

Certifications: Microsoft Azure Data Fundamentals - DP-900 (Microsoft), Microsoft Certified: Azure Data Engineer Associate DP-203 (Microsoft) AWS Certified Data Analytics – Preparing for Associate Level (In Progress)

Projects

M3 Engine **Santander** *São Paulo, SP. Brazil* **2022**

- Designed and implemented a custom orchestration engine for on-premise data pipelines in a cost-constrained banking environment, eliminating the need for commercial solutions like Airflow or dbt.
- Built a modular orchestration layer in Python, triggered via Control-M and a Bash entry script, which dynamically ingested execution parameters (e.g., date ranges, pipeline names, environment flags) to manage complex workloads.
- Enabled parallel execution of Spark pipelines with real-time log monitoring through YARN, significantly reducing total runtime and improving operational observability.
- Established a consistent repository structure and development framework in GitLab, ensuring compatibility with the orchestrator and standardization across teams.
- Developed an internal Python library to streamline I/O operations, embed business logic, and serve as a foundation for future automation, reducing dev time for new pipelines by over 40%.

Open Brewery DB **Personal project** *São Paulo, SP. Brazil* **2025**

- Developed a fully automated data pipeline using Apache Airflow to extract data from the Open Brewery DB API and store it in a Snowflake data lake, structured according to the medallion architecture (bronze, silver, gold).
- Leveraged AWS S3 as the storage layer for raw JSON data and SQL query assets, implementing secure access via IAM credentials and dedicated bucket paths.
- Orchestrated a multi-stage ETL workflow in Airflow with modular task groups for ingestion, transformation, validation, and aggregation of brewery data. Gained hands-on experience with AWS service setup, including S3, low-level use of IAM, and Airflow's native support for AWS connections via environment variables.
- Implemented automated validation checks and retry logic to ensure data quality, and integrated Slack notifications for real-time alerting in case of failures.
- Containerized the full environment using Docker and exposed the Airflow UI for local execution, simulating production-like orchestration.