

Best Technology Talent For Companies Around The World



Robson

Senior Data Engineer

Summary

Senior Data Engineer focused on designing, building, and optimizing highly scalable and maintainable ETL/ELT solutions. I transform complex, high-volume data into reliable, business-critical assets. My expertise is in delivering end-to-end data systems that can handle millions of users and require zero-downtime performance.

Specializes in Cloud Data Warehousing and the full lifecycle of data platform modernization and migration (AWS, Azure, GCP).

TECHNICAL SKILLS:

- Clouds: AWS, Azure, Google Cloud
- Programming languages: Python, Golang
- DevOps: GitHub Actions, Git, Terraform, Bash Scripting, Docker
- Database: SQL, PostgreSQL, RedShift, Google Big Query
- Data Ingestion: Apache Nifi, Apache Sqoop, Apache Flume, Streamsets, Logstash, Data Factory
- Storage: Hdfs, Postgres, MySQL, Oracle, Elastic Search, SnowFlake, S3, Cloud Storage
- Orchestration: Apache Airflow
- Middleware: Apache Kafka
- PowerBI, Apache Superset, Metabase, Data Studio, Kibana, Matplotlib
- Pandas, numpy, PySpark
- Kanban, Sprints, CI/CD, TDD EDUCATION
- Scikit-learn, Pyspark MLlib

Professional Career

Senior Data Engineer

January 2022 – Present | Santa Catarina, Brazil

Project Info:

Working across US clients delivering data engineering for SaaS applications

Responsibilities:

- Creation of Data Quality Checks with both Python code and SQL Procedures, implementation of an alerting system with Slack, running on Airflow DAGs call-back, usage of audit tables to control data flow from origin to target, usage of built in DBT data quality checks
- Architecture and implementation of a SaaS data analytics project using Databricks, Airbyte and Apache Superset
- Design and implementation of an ETL project with Airflow and Golang for a SaaS client in the Financial industry, including IaC with Terraform, Ci/CD with Github Actions that boosted processing speed by 25% deployed at AWS
- Refactored Python backend to PySpark, achieving a remarkable 50% enhancement in processing speed with Google DataProc and SnowFlake
- Maintenance in Azure Data Factory pipelines across different clients
- Maintenance in Airflow ETL Python scripts deployed in GCP
- Requirements gathering, definition of priorities within stakeholders and design of data architectures based on business problems

Software and technologies:

DBT, GCP, Big Query, Azure, Data Factory, AWS, Airflow, Docker, Terraform, PostgreSQL, Python, PySpark, Golang, Linux, Git, Databricks, Docker, SnowFlake

Senior Data Engineer

June 2021 – December 2021 | Santa Catarina, Brazil

Project Info:

Working in the Big Data project in the State Income Tax Department of the State of Santa Catarina.

Responsibilities:

- Implementation of Data Quality and Audit mechanisms on PySpark jobs
- Fostered data engineering efforts, achieving improvements in data processing efficiency for the Big Data project in the Income Tax Department of the State of Santa Catarina
- Managed Oracle BDA Cluster and Cloudera Manager, resulting in reduction in system maintenance downtime

- Implemented data ingestion processes with Python and Apache Airflow, optimizing workflows and reducing processing time
- Conducted data processing operations with PySpark, contributing to around 30% increase in analytical capabilities
- Played a key role in various Cloud Azure and Databricks mini-projects, maintaining Azure Data Factory pipelines and Databricks spark jobs to process thousands of records

Software and technologies:

Airflow, Oracle, Linux, Cloudera, Azure, Hadoop, Python, PySpark, Git.

Senior Data Engineer

April 2021 – June 2021 | Santa Catarina, Brazil

Project Info:

An omnichannel platform to support and manage all venues of conversation in a single environment, seamlessly

Responsibilities:

- Developed a robust data pipeline utilizing Apache Airflow, AWS S3, DBT, and AWS Redshift, enhancing data processing efficiency
- Executed seamless data ingestion from REST APIs using Python, ensuring reliable and automated workflows
- Applied advanced data modelling techniques with DBT and SQL to optimize data structures and enhance analytical capabilities

Software and technologies:

AWS, Airflow, Python, DBT, SQL, RedShift, S3

Senior Data Engineer, Architect

September 2020 – April 2021 | Santa Catarina, Brazil

Project Info:

A largest communications company in the state of Santa Catarina. It is live, multichannel content that reaches people wherever they are

Responsibilities:

- Developed a Metrics Mapping Framework using CRISP-DM, contributing to a 20% increase in data-driven decision-making efficiency
- Established a real-time data pipeline with Apache Nifi , Kafka, Postgres, and Superset for tweet analysis, resulting in improvements in real-time analytics capabilities
- Designed the architecture for efficient data processing, ensuring scalability and optimal performance

- Held the role of Product Owner/Leader for the Big Data project at the NSC Innovation LAB

Software and technologies:

Apache Nifi , Kafka, Python, Hadoop, SQL, AWS, Github Actions

Data Analyst

February 2019 – May 2020 | Santa Catarina, Brazil

Project Info:

B2B sales and e-commerce system for industries, distributors and sales representatives.

Responsibilities:

- Implemented an efficient ETL pipeline with a low code paid tool, GCP Bigquery, and Metabase, improving data processing by 100%
- Documented processes for the Partners Program, enhancing program efficiency
- Designed metrics, contributing to an increase in the effectiveness of key performance indicators and management decision making

Software and technologies:

GCP, Python, Metabase, SQL

Education

2018-2020

Sustentare Business School

MBA Business Intelligence Data analysis, Strategic marketing management, market research, and digital transformation

2015 - 2017

State University of Santa Catarina

Bachelor's degree, Information Technology

2009 – 2013

Univille University

Bachelor's Degree in Business Administration

Certificates:

- A.I Engineering (400hrs) - Data Science Academy - on going
- Machine Learning Specialization - DeepLearning.ai - 2024
- Techlead Mentorship - IFTL - 2024

- Datalake Design and Implementation - Data Science Academy - 2022
- Cloud Practitioner Essentials - Amazon Web Services (AWS) - 2023
- Data Engineering Certificate - Data Science Academy (DSA20225199E891000000076) - 2022
- Data Engineering with Python - DataCamp – 2021

Foreign Languages Skills

English: Upper-Intermediate

Portuguese: Native

Spanish: Advanced

Italian: Intermediate/Advanced